JKMS

Review Article
Biomedical Engineering

Check for updates

Polygenic Risk Score for Cardiovascular Diseases in Artificial Intelligence Paradigm: A Review

Narendra N Khanna [ID],[1,2] Manasvi Singh [ID],[3,4] Mahesh Maindarkar [ID],[2,3,5]
Ashish Kumar [ID],[4] Amer M. Johri [ID],[6] Laura Mentella [ID],[7] John R Laird [ID],[8]
Kosmas I. Paraskevas [ID],[9] Zoltan Ruzsa [ID],[10] Narpinder Singh [ID],[11]
Mannudeep K. Kalra [ID],[12] Jose Fernandes E. Fernandes [ID],[13] Seemant Chaturvedi [ID],[14]
Andrew Nicolaides [ID],[15] Vijay Rathore [ID],[16] Inder Singh [ID],[3] Jagjit S. Teji [ID],[17]
Mostafa Al-Maini [ID],[18] Esma R. Isenovic [ID],[19] Vijay Viswanathan [ID],[20]
Puneet Khanna [ID],[21] Mostafa M. Fouda [ID],[22] Luca Saba [ID],[23] and Jasjit S. Suri [ID] [2,3,24]

[1]Department of Cardiology, Indraprastha APOLLO Hospitals, New Delhi, India
[2]Asia Pacific Vascular Society, New Delhi, India
[3]Stroke Monitoring and Diagnostic Division, AtheroPoint™, Roseville, CA, USA
[4]Bennett University, Greater Noida, India
[5]School of Bioengineering Sciences and Research, Maharashtra Institute of Technology's Art, Design and Technology University, Pune, India
[6]Department of Medicine, Division of Cardiology, Queen's University, Kingston, Canada
[7]Department of Medicine, Division of Cardiology, University of Toronto, Toronto, Canada
[8]Heart and Vascular Institute, Adventist Health St. Helena, St. Helena, CA, USA
[9]Department of Vascular Surgery, Central Clinic of Athens, Athens, Greece
[10]Invasive Cardiology Division, University of Szeged, Szeged, Hungary
[11]Department of Food Science and Technology, Graphic Era Deemed to be University, Dehradun, Uttarakhand, India
[12]Department of Radiology, Harvard Medical School, Boston, MA, USA
[13]Department of Vascular Surgery, University of Lisbon, Lisbon, Portugal
[14]Department of Neurology & Stroke Program, University of Maryland, Baltimore, MD, USA
[15]Vascular Screening and Diagnostic Centre and University of Nicosia Medical School, Cyprus
[16]Nephrology Department, Kaiser Permanente, Sacramento, CA, USA
[17]Ann and Robert H. Lurie Children's Hospital of Chicago, Chicago, IL, USA
[18]Allergy, Clinical Immunology and Rheumatology Institute, Toronto, ON, Canada
[19]Department of Radiobiology and Molecular Genetics, National Institute of The Republic of Serbia, University of Belgrade, Beograd, Serbia
[20]MV Diabetes Centre, Royapuram, Chennai, Tamil Nadu, India
[21]Department of Anaesthesiology, AIIMS, New Delhi, India
[22]Department of Electrical and Computer Engineering, Idaho State University, Pocatello, ID, USA
[23]Department of Radiology, Azienda Ospedaliero Universitaria, Cagliari, Italy
[24]Department of Computer Engineering, Graphic Era Deemed to be University, Dehradun, India

OPEN ACCESS

ORCID iDs
Narendra N Khanna [ID]
https://orcid.org/0000-0003-2957-3825
Manasvi Singh [ID]
https://orcid.org/0000-0001-8551-784X
Mahesh Maindarkar [ID]
https://orcid.org/0000-0002-0813-4906
Ashish Kumar [ID]
https://orcid.org/0000-0003-3466-8845
Amer M. Johri [ID]
https://orcid.org/0000-0001-7044-8212
Laura Mentella [ID]
https://orcid.org/0000-0002-6527-426X

ABSTRACT

Cardiovascular disease (CVD) related mortality and morbidity heavily strain society. The relationship between external risk factors and our genetics have not been well established. It is widely acknowledged that environmental influence and individual behaviours play a significant role in CVD vulnerability, leading to the development of polygenic risk scores (PRS). We employed the PRISMA search method to locate pertinent research and literature to extensively review artificial intelligence (AI)-based PRS models for CVD risk prediction. Furthermore, we analyzed and compared conventional vs. AI-based solutions for PRS. We summarized the recent advances in our understanding of the use of AI-based PRS for risk prediction of CVD. Our study proposes three hypotheses: i) Multiple genetic variations

John R Laird
https://orcid.org/0000-0002-7454-4145

Kosmas I. Paraskevas
https://orcid.org/0000-0001-6865-2919

Zoltan Ruzsa
https://orcid.org/0000-0002-2474-5723

Narpinder Singh
https://orcid.org/0000-0002-9683-7846

Mannudeep K. Kalra
https://orcid.org/0000-0001-9938-7476

Jose Fernandes E. Fernandes
https://orcid.org/0000-0003-4677-6077

Seemant Chaturvedi
https://orcid.org/0000-0001-8660-7644

Andrew Nicolaides
https://orcid.org/0000-0003-3912-7394

Vijay Rathore
https://orcid.org/0000-0003-4315-5072

Inder Singh
https://orcid.org/0000-0002-2844-6050

Jagjit S. Teji
https://orcid.org/0000-0001-7773-2144

Mostafa Al-Maini
https://orcid.org/0000-0003-2553-591X

Esma R. Isenovic
https://orcid.org/0000-0002-0012-2636

Vijay Viswanathan
https://orcid.org/0000-0001-9938-7476

Puneet Khanna
https://orcid.org/0000-0002-9243-9963

Mostafa M. Fouda
https://orcid.org/0000-0003-1790-8640

Luca Saba
https://orcid.org/0000-0003-3610-8526

Jasjit S. Suri
https://orcid.org/0000-0002-0813-4906

and risk factors can be incorporated into AI-based PRS to improve the accuracy of CVD risk predicting. ii) AI-based PRS for CVD circumvents the drawbacks of conventional PRS calculators by incorporating a larger variety of genetic and non-genetic components, allowing for more precise and individualised risk estimations. iii) Using AI approaches, it is possible to significantly reduce the dimensionality of huge genomic datasets, resulting in more accurate and effective disease risk prediction models. Our study highlighted that the AI-PRS model outperformed traditional PRS calculators in predicting CVD risk. Furthermore, using AI-based methods to calculate PRS may increase the precision of risk predictions for CVD and have significant ramifications for individualized prevention and treatment plans.

Keywords: Cardiovascular Disease; Genomics; Polygenic Risk Score; Artificial Intelligence; Precision Medicine

INTRODUCTION

Cardiovascular disease (CVD) is the most prevalent cause of fatalities worldwide and is a global medical concern. According to a recent study by Benjamin et al.[1], just in the United States, in 2017, there were 69,255 more resident deaths (2,813,503) than there were in 2016. Additionally, 17.8 million deaths worldwide (an increase of 21.1% from 2007) were due to CVD in 2017. Moreover, the global burden of CVD is predicted to rise significantly over the next few decades, driven by aging populations and unhealthy lifestyle habits.[2] While behavioral and environmental factors are known to contribute to CVD development, genetic factors also play a substantial role.[3-5] As the science develops in high-throughput genotyping technologies,[6,7] particularly genome-wide association studies (GWAS),[8-10] have identified numerous genetic loci that are robustly associated with various CVD subtypes,[11,12] including coronary artery disease (CAD),[13-15] coronary heart disease,[16-18] heart failure,[19] and atrial fibrillation.[20] These discoveries present opportunities for precision medicine and personalized interventions, but pose significant computational and analytical challenges. This is due to the vast amounts of genomic and clinical data generated by these studies, which necessitate the application of advanced machine learning (ML) and artificial intelligence (AI) algorithms for accurate and efficient analysis.[21-24]

AI comprises a group of methods that draw on concepts like ML and deep learning (DL)[25-36] which can be applied to integrate and decipher complicated and vast amounts of genetic and medical data in circumstances where conventional statistical methods may be inadequate.[37-39] It has revolutionized the healthcare field, including CVD risk prediction, especially with the incorporation of polygenic risk scores (PRS) derived from genomics data.[40,41]

In the field of CVD, the adoption of AI-based PRS (aiPRS) models can dramatically improve our comprehension of CVD pathophysiology and provide more efficient prevention and therapy methods.[42] The aiPRS (originally coined by AtheroPoint™, Roseville, CA, USA) basically refers to the use of AI in the context of PRS models for CVD risk prediction. It combines the complete PRS approach to measuring CVD risk with the power of AI technologies, such as ML or DL techniques. It can analyze enormous volumes of data, spot trends, and offer more accurate risk evaluations by incorporating AI techniques. As a result, "aiPRS" does in this case refer to a classification algorithm based on DL[43,44] or ML[37,40,45] techniques that are especially employed in the field of CVD risk prediction. While AI can

analyze vast amounts of data and identify patterns that may otherwise go unnoticed, PRS provide a more comprehensive and personalized approach to risk assessment by considering the complex interplay between various genetic and environmental factors.[46] Incorporating AI in PRS has improved patient outcomes by enabling more precise and accurate risk assessment. AI-based algorithms can also stratify patients based on their risk profiles and forecast and uncover novel biomarkers of CVD, enabling the development of customized therapy regimes.[47,48]

The promise of aiPRS for CVD research has recently been shown by studies. **Fig. 1** shows the CVD PRS system and treatment planning using the aiP³ model. A test patient's personal information, lifestyle biomarkers, omics-based biomarkers, laboratory-based biomarkers, and radiomics-based biomarkers are all collected first. The PRS system, an offline model that creates risk predictions based on the gathered data, is then trained using this data. Real-time risk prediction and customized treatment planning are made possible using this offline trained system. This is via the integration of the PRS system into an online prediction system. The aiP³model is then used to create a precise, individualized, and preventative treatment plan that considers the individual's unique CVD risks. To optimise CVD preventive and management techniques, this holistic process integrates data-driven risk assessment, online prediction capabilities, and personalised interventions. As an illustration, a study published in Circulation demonstrated that a genetic risk score based on 6.6 million variations could identify those at higher risk of getting CVD, even without conventional risk indicators like high blood pressure and cholesterol levels.[49] Another study published in Medical Oncology emphasizes the crucial roles that ML algorithms play in precision and genomic medicine, highlighting their importance in enhancing individualized healthcare. These algorithms help the healthcare professionals to improve patient care, customize medicines to each patient's needs, and use prediction models to promote precision medicine by harnessing the power of AI and integrating it with genomics.[50] New opportunities for early disease identification, individualized treatment, and disease prevention have been made possible by developing PRS.[51] Additionally, applying AI algorithms to PRS analysis has shown to be a viable method for locating previously unknown risk factors and enhancing prediction precision.[38] Better patient outcomes and more successful interventions may follow as a result.
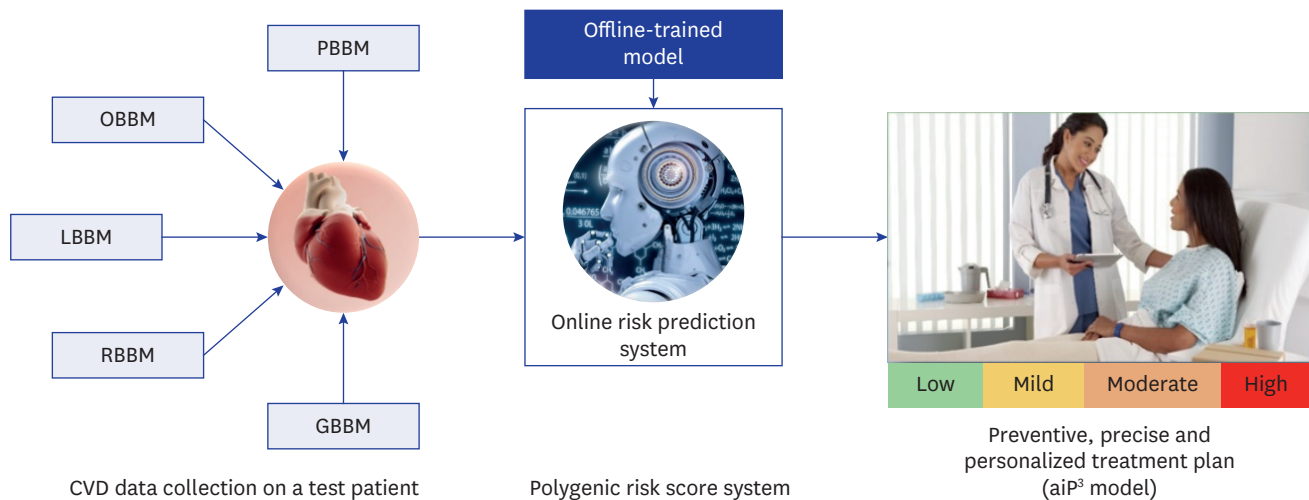
Fig. 1. CVD PRS system and treatment planning using aiP³ model.
OBBM = office-based biomarker, LBBM = laboratory-based biomarker, RBBM = radiomics-based biomarker, GBBM = genomics-based biomarker, PBBM = proteomics-based biomarker, PRS = Polygenic Risk Score, aiP³ = AI for preventive, precise, and personalized system, CVD = cardiovascular disease.

Despite these encouraging advantages, implementing aiPRS models in clinical practice poses many difficulties. These difficulties include addressing data privacy and ethical issues, ensuring the model's interpretability and transparency, and assessing the model's effectiveness across a range of patient populations and clinical settings.[52] However, it is impossible to overlook the potential of aiPRS models to revolutionize preventive measures by enabling early detection of high-risk individuals and focused interventions to lower the burden of CVD globally. Given the rising prevalence of CVD globally, AI offers a potent method to meet the urgent demand for appropriate risk assessment and management strategies.

In this comprehensive study, we further discuss the utility of PRS obtained from genetic data as an effective tool for predicting CVD.[53] We offer a critical evaluation of recent developments in aiPRS for CVD, a ground-breaking framework that combines the strengths of genetic data and AI to improve the precision and accuracy of CVD risk prediction.[54] We dig into the severe obstacles and bright prospects that the creation of aiPRS presents for CVD. The most promising fields for an upcoming study and clinical application were also determined. We hope to encourage new methodologies for CVD risk prediction and allow more efficient preventive interventions by illuminating the strength and promise of aiPRS-CVD.

SEARCH STRATEGY AND STATISTICAL DISTRIBUTIONS

The PRISMA model

Fig. 2 shows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) search technique. The keywords used for the search are "Polygenic Risk Score and cardiovascular disease," "Polygenic Risk Score and AI," "Polygenic Risk Score and genomics," "Cardiovascular diseases and AI," "CVD, PRS and AI," "Polygenic Risk Score and Deep Learning," "Genomics and AI," "Cardiovascular Diseases and PubMed and Google Scholar
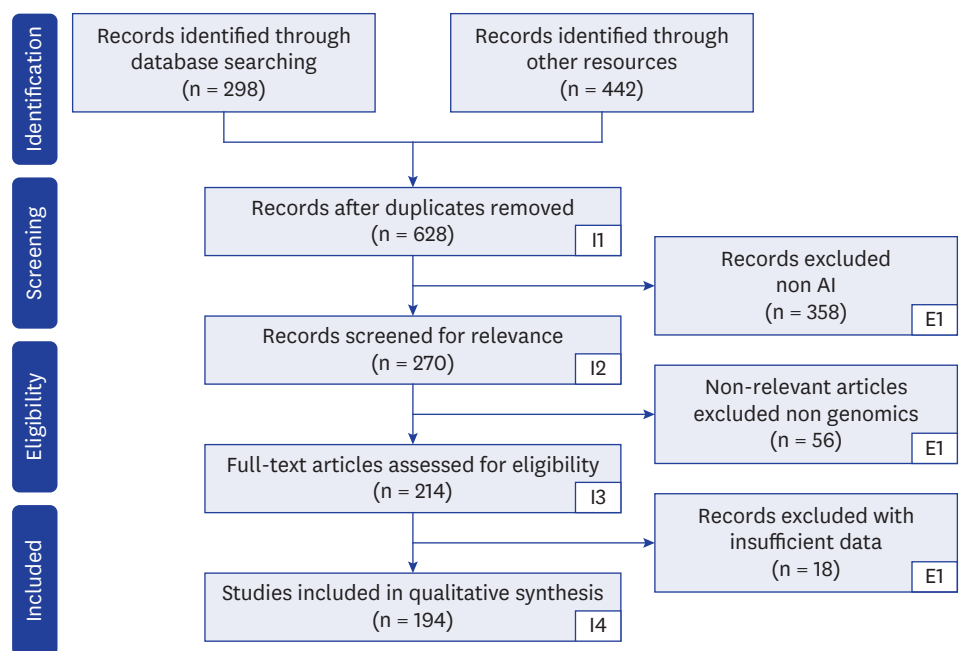
Fig. 2. PRISMA model for study selection.
AI = artificial intelligence.

screened pertinent papers." Among other electronic databases, PubMed, Google Scholar, and Science Direct were searched extensively. The following inclusion standards were chosen: papers written in English; (a) initial research publications or reviews that examined the application of AI in bioinformatics frameworks for forecasting CVD risk. Letters to the editor, conference proceedings, and publications irrelevant to the research subject were not considered. Out of the total of 740 studies identified, 112 were found to be duplicates. After removing the duplicates, the number of unique studies left was 628. Further screening was conducted, resulting in the exclusion of 358 studies that were unrelated to the research topic. Additionally, 56 studies were deemed irrelevant, and 18 studies lacked sufficient data. As a result, a total of 194 studies met the inclusion criteria and were selected for analysis. Following this search strategy, we aimed to discover the AI and genomics-based approaches to investigate the relationship between PRS and CVD.

Statistical distribution

Fig. 3 represents the statistical distribution of the various studies employed in the present article. **Fig. 3A** shows the comparison of PRS studies versus study type in the four different categories: 1) PRS with AI but without CVD: 25 studies involved using AI in conjunction with PRS but did not focus on CVD as the outcome. 2) PRS without AI but with CVD: 20 studies did not involve the use of AI but focused on CVD as the outcome using PRS. 3) PRS with both CVD and AI: 13 studies involved both PRS and CVD as the outcome and utilized AI techniques. 4) for PRS without CVD and without AI: 12 studies did not involve PRS or AI and did not focus on CVD as the outcome. **Fig. 3B** shows the number of studies that utilized PRS
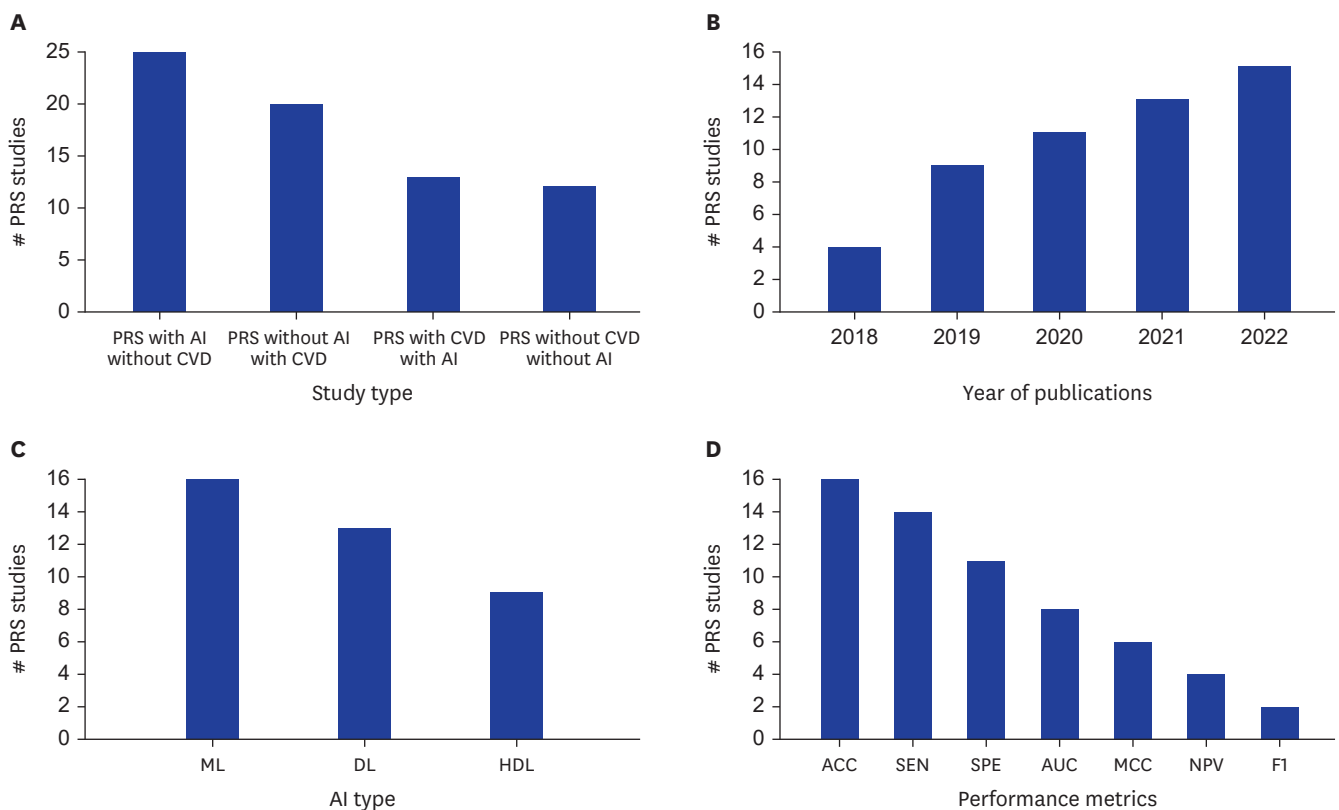
Fig. 3. Statistical distribution of various studies involved in PRS.
PRS = Polygenic Risk Score, AI = artificial intelligence, CVD = cardiovascular diseases, ML = machine learning, DL = deep learning, HDL = hybrid deep learning, ACC = accuracy, SEN = sensitivity, AUC = area under curve, MCC = Mathews orrelation coefficient, SPE = specificity, NPV = net present value.

as a variable across five years, from 2018 to 2022. The number of studies increased from 5 in 2018 to 16 in 2022, with the most significant increase occurring between 2018 and 2019. The trend suggests a growing interest in PRS research over time.

Fig. 3C represents the number of RA studies utilizing different AI techniques, including ML, DL, and hybrid deep learning (HDL). ML was used in 16 PRS studies, DL was used in 13 PRS studies, and HDL was used in 9 PRS studies. In our context, HDL refers to an approach that combines two different DL methods. Example of HDL are by combining two DL models.[55-63] This may be serial two DL connections or parallel connection of two DL models. In the literature of AI, HDL has also been called as the fusion of ML and DL models.[18] These results suggest that AI is increasingly utilized in PRS research, with ML being the most used technique. **Fig. 3D** shows the performance metrics used in studies, and the number of studies that utilized each metric. The performance metrics listed are Accuracy (ACC), Sensitivity (SEN), Specificity (SPE), Area-Under-the-Curve (AUC), Matthews Correlation Coefficient (MCC), Negative Predictive Value (NPV), and F1-Score (F1). The number of studies that reported these metrics are as follows: ACC (16 studies), SEN (14 studies), SPE (11 studies), AUC (8 studies), MCC (6 studies), NPV (4 studies), and F1 (2 studies). This figure provides an overview of the usage of different performance metrics in the studies, with ACC being the most reported metric and F1 being the least commonly reported metric.

CVD RISK PREDICTION AND MANAGEMENT USING GENOMIC APPROACHES

The study of genes and DNA in people and other living organisms, known as genomics, sheds light on their genetic make-up and the expression and regulation of features. As genetic changes can play a role in the development and risk of CVD, understanding gene activity correlations at the molecular level can assist in identifying risk factors and forecasting disease outcomes. Genomics in CVD involves identifying genetic variants, such as single nucleotide polymorphisms (SNPs), contributing to disease development.[64,65] By analyzing individual patient genomes, doctors can develop personalized treatment plans based on their genetic makeup, improving the chances of successful intervention and prevention.[66]

Genomic methods and technology have become effective tools in the research of CVD, providing fresh perspectives on disease causes and new treatment targets. These include next-generation sequencing, which has revolutionized the field of genomics by enabling the rapid and affordable sequencing of entire genomes or targeted regions of interest, thereby allowing for the identification of rare genetic variants associated with CVD. To find genetic variants linked to complicated diseases like CVD, GWAS examines the whole genomes of large populations.[67] The term "actual values in a large population" refers to the real-world outcomes related to the trait or disease being studied. This means gathering data on whether individuals in a large population develop the trait or disease over time. It is sometimes also called a gold standard or ground truth value.[68-71] In particular, GWAS is a tool that evaluates genetic information to uncover differences in genes related to traits or disorders in an individual.[9,10]

Additionally, the study of the whole transcriptome, proteome, and metabolome through the disciplines of transcriptomics, proteomics, and metabolomics has shed light on the patterns of gene and protein expression and metabolic pathways linked to CVD.[72-76] The function of epigenetic alterations in the onset and development of CVD has also been

clarified by epigenomics.[76,77] Finally, the function of genes linked to CVD has been studied and new treatments for the condition have been created using CRISPR/Cas9 gene editing technology.[78] Together, these genomic tools and methods have uncovered previously unattainable details about the molecular underpinnings of CVD, paving the way for discovering new therapeutic targets for the condition.

CONVENTIONAL PRS USING GWAS

PRS are statistical measures used to identify the likelihood of developing heritable diseases or traits, such as CVD. Conventional PRS techniques, like GWAS, assess DNA samples from enormous populations to locate specific SNPs linked to the target condition or trait.[79,80] In particular, GWAS examines the relationships between genetic variations, including SNPs, and the traits or diseases of interest.[81] **Fig. 4** shows elements for calculating PRS for CVD. A region in the genome known as an SNP indicates a distinction in an individual's genetic makeup by harboring one nucleotide in many forms.[82] In GWAS, DNA samples from a large population are usually gathered, the genetic variations present in each sample are identified, and the data are then analyzed to find SNPs associated with the characteristic or disease of interest. While GWAS has been used to examine a variety of traits and diseases, including cancer, type 2 diabetes (T2D), and CAD, it is crucial to remember that SNPs are only one of many potential risk factors for disease.[83-88] Finding a novel SNP does not necessarily mean that the disease will be caused by it. Additionally, each gene has two copies, or alleles, present in pairs at each chromosomal locus and inherited from each parent. The many gene combinations that can produce distinct traits dictate the expressiveness of each trait in the body. To identify individuals with SNPs of interest, researchers calculate the frequency of alleles by dividing the number of individuals with the disease or trait of interest by the total number of individuals in the genotyping experiment. Then weights are assigned to each SNPs based on their effect size, determining the strength of the association between the SNPs and the trait or disease. A positive effect size denotes that a specific allele is attributed to a higher risk of the trait or disease, while a negative effect size denotes that a specific allele is attributed to a lower risk.[89,90]

The statistical analysis calculates the *P* value, which represents the probability of the association between the SNP and the trait or disease occurring by chance. Researchers multiply the effect size by the genotype of individual *p* at SNP *i*, assigning weights to each
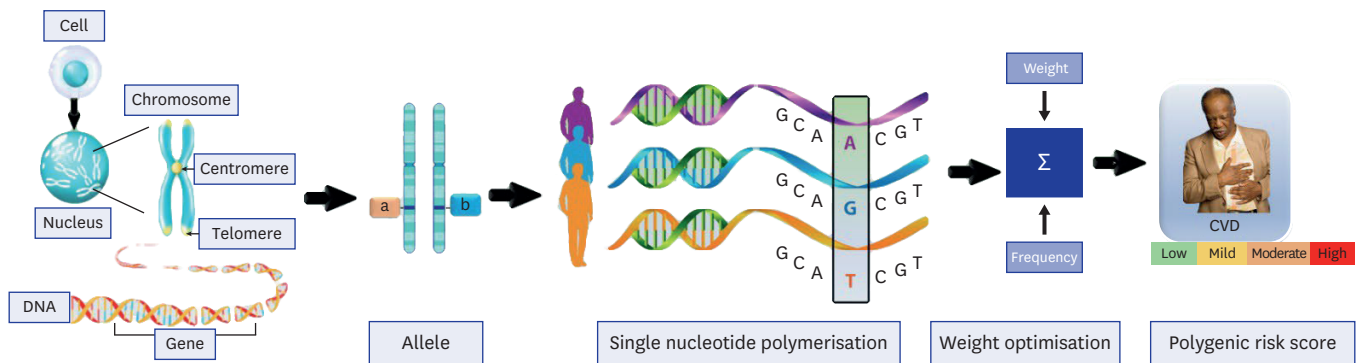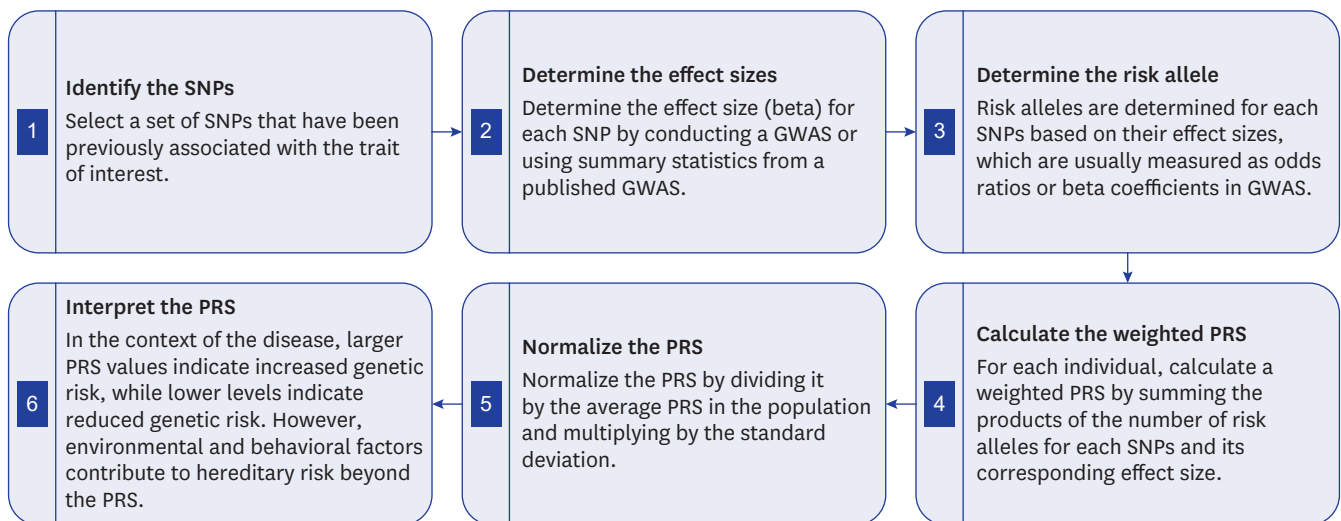

Fig. 4. Elements for calculating Polygenic Risk Score for CVD.
CVD = cardiovascular diseases.

SNP. These weights are then summed up to obtain the PRS for one participant. The accuracy of the PRS is assessed by comparing predicted risk scores with actual values in a large population, which can help to improve its predictive ability.

Where N is the total number of SNPs in the PRS,[INSERT FIGURE 002] signifies the effect size (or beta) of variant i, and [INSERT FIGURE 003]reflects the number of replicas of SNP i in the genotype of individual p.[53,91] After calculating the PRS we validate it by comparing the PRS with the actual value in many populations. It will assess the accuracy of PRS and can help to improve its predictive ability. PRS helps take clinical decisions regarding the concerned disease, however it should not be the only tool to identify the risk in patients because it does not consider the environmental factors, lifestyle, and family history, which could be different for every individual. If we combine the PRS with other risk factors, the accuracy and predictive capability of the PRS would increase manifolds.[92]

The flowchart in **Fig. 5** illustrates the many procedures needed to compute and decipher a weighted PRS using SNPs. In step 1, we select a collection of SNPs that have been previously associated with the trait or condition of interest. Consult published studies, databases, or consortia studies that have identified genetic variants linked to the specific trait to accomplish this. In step 2, one must determine the effect size (beta) of each SNP after identifying the pertinent SNPs. This can be done by conducting a GWAS specific to the trait or by leveraging the results of a previously published GWAS. The effect size (beta) reveals the strength of the association between the SNP and the desired attribute. The risk allele for each SNP can be chosen depending on the impact sizes you obtained in step 2. Risk alleles are frequently found by examining the direction of the effect magnitude. For instance, if the allele associated with that influence exhibited a higher risk for the trait, it would be referred to as the risk allele. For each person, the weighted PRS is calculated by multiplying the number of risk alleles for each SNP by the effect size (beta) that each of those risk alleles has. By summing the total of these products across all the chosen SNPs and dividing each

1 Identify the SNPs
Select a set of SNPs that have been previously associated with the trait of interest.

2 Determine the effect sizes
Determine the effect size (beta) for each SNP by conducting a GWAS or using summary statistics from a published GWAS.

3 Determine the risk allele
Risk alleles are determined for each SNPs based on their effect sizes, which are usually measured as odds ratios or beta coefficients in GWAS.

6 Interpret the PRS
In the context of the disease, larger PRS values indicate increased genetic risk, while lower levels indicate reduced genetic risk. However, environmental and behavioral factors contribute to hereditary risk beyond the PRS.

5 Normalize the PRS
Normalize the PRS by dividing it by the average PRS in the population and multiplying by the standard deviation.

4 Calculate the weighted PRS
For each individual, calculate a weighted PRS by summing the products of the number of risk alleles for each SNPs and its corresponding effect size.

Fig. 5. GWAS analysis for calculating and interpreting PRS. A simplified illustration of the related phases involved in GWAS analysis is provided by this flowchart. The first step in the procedure is to locate the SNPs associated with the desired phenotype (Box 1) and calculate their effect sizes (Box 2). The weighted PRS is then determined by identifying the risk allele (Box 3 and 4). To allow for inter-individual comparisons, the PRS is then normalized (Box 5). The trait under investigation is then taken into consideration when interpreting the PRS (Box 6). Please take note that although the steps are listed sequentially for clarity, they are actually interrelated and iterative in nature.
SNP = single nucleotide polymorphism, GWAS = Gene Wide Association Studies, PRS = Polygenic Risk Score.

SNP's impact size by the number of risk alleles a person carries for each SNP, this may be done. A genetic risk score for the desired characteristic or illness should be considered when interpreting the PRS. Higher PRS levels imply a greater genetic risk for the trait whereas lower readings indicate a reduced genetic risk. It is critical to keep in mind that the PRS is only one aspect of genetic risk and that environmental and behavioral factors that can increase the chance of getting a disease should be considered in addition to genetic risk.

ROLE OF AI IN PRS

In genomics, PRS is an effective instrument for predicting complex features like CVD.[93] The PRS is constructed by selecting markers from a preliminary training sample and then constructing a weighted sum of associated alleles within each participant. This approach has been implemented to develop risk prediction models for CVD and to determine a shared genetic foundation for associated diseases.[94] More complex models for predicting CVD based on PRS have been developed as a result of recent breakthroughs in AI.[95-97]

Determining which genetic variants are most useful for predicting disease risk is a significant use of AI in developing PRS.[95,98] The polymorphisms most strongly related to disease risks can be found using ML algorithms that analyze large-scale genomic data from varied populations.[99,100] By utilizing the expanding body of big data in medicine, AI may enable the best creation of patient-specific models for enhancing CVD diagnosis, intervention, and outcome. AI can also be utilized to generate advanced approaches, such as neural networks, and DL algorithms, for amalgamating genetic variations into a single PRS.[43,101-104]

Additionally, it can assist in improving the weighting of specific genetic variations in PRS by accounting for their effect magnitude and other variables including gene-gene interactions and environmental factors that could affect their influence on disease risk. This may result in more accurate and individualized risk evaluations.[105-107] AI may also be employed to create risk stratification tools that are more sophisticated and account for various risk factors as well as their interactions.[108] For instance, using AI, risk assessments and preventative plans can be tailored to subgroups of people with various risk profiles depending on their genetic makeup and environmental exposure.[109] AI can speed up the creation of new PRS models and increase their ability to generalize to various populations. AI can generate more precise and individualized risk models that can be used to guide therapeutic decision-making by analyzing massive and diverse datasets to find genetic and environmental factors that may be more significant in certain populations or subgroups. Based on a person's risk profile, AI-powered decision support tools can also assist in identifying effective screening and preventative initiatives.[110,111] For instance, those with a high risk of developing a particular type of cancer could be advised to undergo earlier and more frequent screening, whilst those with a low risk might be able to skip a few screening procedures. **Table 1**[112-120] lists details on various CVD risk assessment programs, including their risk ratings, citations, and distinguishing characteristics. Based on several variables like age, gender, blood pressure, cholesterol levels, family history, and ethnicity, these programs are intended to assess a person's risk of acquiring cardiovascular illnesses.

Additionally, AI-powered decision support solutions can offer patients educational materials and resources to aid in their understanding of their PRS and any associated health effects. Patients might be advised to make lifestyle changes, such as dietary and activity adjustments,

Table 1. Comparative analysis of patient characteristics and risk scores

No.	Citations	Year	Risk score	Disease	Age, yr	Gender	HDL-C	SBP	SS	DS	FH	Ethnicity	BMI, kg/m²
1	Kannel et al. [112]	1967	Framingham	CVD	35–64	Both	o[a]	o	o	o	o	Largely White	NR
2	Assmann and Schulte [113]	1988	PROCAM	CVD	35–65	Both	x[b]		o	NR	x	German	NR
3	Woodward et al. [114]	2007	ASSIGN	CVD	30–74	Both	o	o	NR	o	o	Scottish	NR
4	Lakoski et al. [115]	2007	MESA	CVD	30–85	Both	x	x	o	NR	x	African-American, Hispanic, Chinese-American, and Caucasian	18.5–40
5	Ridker et al. [116]	2007	Reynolds	CVD	45–80	Female	o	o	o	o	o	White	18.5–34.9
6	Hippisley-Cox et al. [117,118]	2008	QRISK2	CVD	25–84	Both	o	o	o	o	o	South Asian, Black African-Caribbean, Middle Eastern, or Eastern European	15–55
7	Goff et al. [119]	2013	ACC/AHA pooled cohort equations	CVD	20–59	Both	x	o	o	o	o	African-American	18.5–40
8	Conroy et al. [120]	2017	SCORE	CVD	45–64	Both	x	o	o	o	o	European	NR

HDL-C = high density, SBP = systolic blood pressure, SS = smoking status, DS = diabetes status, FH = family history, BMI = body mass index, CVD = cardiovascular disease, NR = not reported, PROCAM = Prospective Cardiovascular Münster, ASSIGN = ASsessing cardiovascular risk using SIGN guidelines, MESA = Multi-Ethnic Study of Atherosclerosis, QRISK = QRESEARCH cardiovascular risk algorithm(www.qresearch.org), SCORE = Systematic COronary Risk Evaluation, ACC = American College of Cardiology, AHA = American Heart Association.
[a]o = included, [b]x = not included.

that can lower their risk of contracting specific diseases. In summary, AI has the potential to revolutionize the use of PRS information to improve healthcare by delivering more individualized risk assessments, better-targeted screening and preventative efforts, and more successful patient engagement and education. **Fig. 6** shows the various ways AI is used in genomic medicine. It displays the various phases of the genomic medicine pipeline, such as the gathering and preprocessing genomic data, genomic analysis and interpretation, and clinical decision-making. AI approaches including forecasting, amalgamation, and genomic inference are applied to increase the precision and dependability of genetic analysis. Some of the clinical applications of AI in genomic medicine include routine mode practice, pharmacological trials, and system design. AI can also be used to analyze gene variations and PRS to forecast a person's chance of contracting a specific disease. The figure highlights ways AI is helping to advance genomic medicine and enhance patient outcomes.
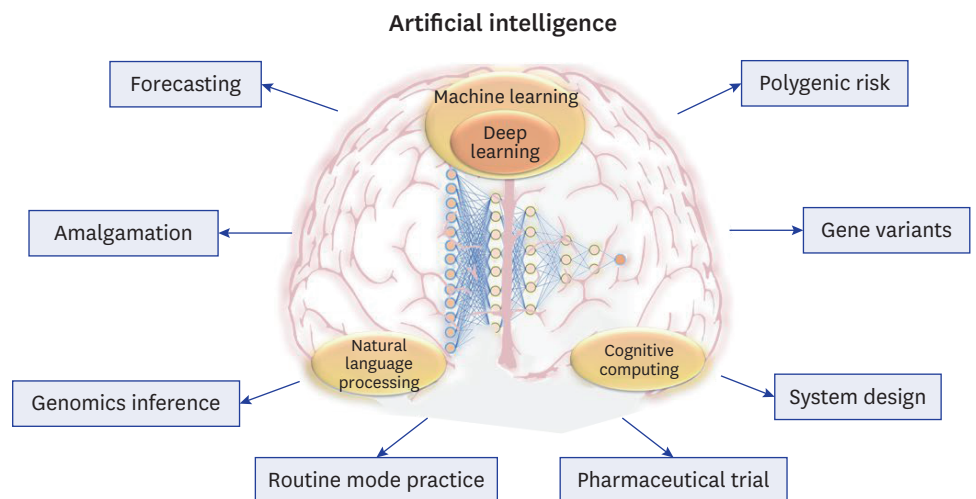
Fig. 6. AI-powered genomic risk assessment in personalized medicine.
AI = artificial intelligence.

Deep learning model dimensionality reduction for gene analysis

In this section, we discuss the diverse methodologies and strategies researchers have employed to mitigate the challenge posed by high-dimensional gene data, which comprises many variables or features.[121] The intricacies of analyzing and interpreting such data require dimensionality reduction, whereby the number of variables is reduced to a more manageable level, thereby enhancing the performance of DL models.[122]

Dimensionality reduction has been well established in the field of medical imaging before.[123,124] They offer several advantages in our current context. First, by lowering the number of features, these strategies boost computational performance, enabling a quicker and more effective analysis of high-dimensional gene data. Given the prevalence of large-scale datasets in gene analysis, such a paradigm is very helpful. Second, dimensionality reduction improves DL model performance by reducing the dimensionality curse. These methods increase the generalization and prediction accuracy of the models by concentrating on the most instructive aspects and removing noise and extraneous data. Additionally, dimensionality reduction aids in tackling multicollinearity, which can develop when strongly associated predictor variables. Dimensionality reduction assures that the DL models are built on independent and uncorrelated variables by finding and removing correlated characteristics, producing more consistent and understandable results. Moreover, these techniques facilitate visualization and interpretation of high-dimensional gene data. By reducing the data to a lower-dimensional space, patterns and relationships within the data become easier to comprehend. This helps in gaining insights into the underlying structure and mechanisms of gene-related phenomena. In conclusion, dimensionality reduction methods in gene analysis include benefits like increased computing effectiveness, improved model performance, addressing multicollinearity, facilitating visualization and interpretation, and extracting significant features. In our research we explore the details on the various dimensionality reduction techniques implemented in DL models for gene analysis, like GWAS-based dimensionality reduction, Binary Particle Swarm Optimization (BPSO)-based dimensionality reduction, and Random Walk Restart-CNN-based dimensionality reduction.[125,126] Each of these approaches involves distinctive strategies for selecting the most pertinent features, transforming the data into a lower-dimensional format, and utilizing DL models to explore the data. Several methods have been applied for dimensionality reduction for ML/DL models. Previously principal component analysis (PCA)[127,128] and PCA-polling[128,129] methods have been applied. In genomics-based research, statistical test has been for feature dimensionality reduction.[130] BPSO and Random Walk Research-CNN based approaches are evolutionary methods mainly for compression (or dimensionality reduction) objective and iterative in nature,[131] unlike PCA and PCA-polling methods are less iterative in nature and straight forward to implement. BPSO have shown to have different applications in pruning AI systems and yields higher dimensionality reduction ratio. Peng et al.[132] proposed the deep PRS model which encodes the genotype information into feature vectors, which are then sent into a deep neural network with a Bi-LSTM layer to capture long-distance interactions between genes. This method reduced the dimensionality of genetic data.

According to the deep PRS risk analysis, people with high deep PRS are more likely to develop the disease than people with low deep PRSs. Deep PRS can be employed as a disease early warning indication for screening and disease prevention in high-risk populations. For Alzheimer's disease (AD), inflammatory bowel disease (IBD), T2D, and breast cancer (BC), deep PRS performed better than two other state-of-the-art approaches using the UK Biobank

dataset. Also, it assessed deep PRS ability to produce PRSs for these diseases. The findings demonstrated that deep PRS performed well in predicting disease risk. The AUC values for AD, IBD, T2D, and BC were 0.7245, 0.6517, 0.6508, and 0.6227, respectively. Furthermore, when Deep PRS was combined with clinical features, the AUC values improved to 0.8624, 0.6585, 0.7316, and 0.6660 for AD, IBD, T2D, and BC, respectively. These results demonstrate the strong performance of Deep PRS when evaluated in conjunction with clinical characteristics. Deep PRS also outperforms methods that use genotype weight estimates from GWAS and requires less prior knowledge than conventional methods.

Khalifa et al.[125] employed the BPSO algorithm as a feature selection technique to isolate the most pertinent genes from high-dimensional gene expression data. These genes were then transformed into a low-dimensional image format and fed into a CNN for tumor classification. Preprocessing is the initial step, which entails utilizing BPSO-DT to choose the best features from the high-dimensional RNA-seq data and transform the selected features into 2D pictures. The second stage is augmentation, which multiplies the initial dataset of 2,086 samples by five while having the least impact on the image's features. This aids in overcoming the issue of overfitting and educates the model to be more accurate. The third stage is the deep CNN phase, which has two primary convolutional layers for feature extraction and two fully connected layers for classification. The proposed method was tested on five different cancer types: uterine corpus endometrial carcinoma, lung squamous cell carcinoma, lung adenocarcinoma, and kidney renal clear cell carcinoma. The findings demonstrate that the suggested method exceeded comparable works in testing accuracy for the five kinds of cancer, achieving an overall testing accuracy of 96.90%. Additionally, the suggested method requires less memory and is less complicated.

Peng et al.[126] devised a model for predicting gene function that used semi AE, a unique version of AE which stands for autoencoders, a class of neural networks frequently employed for unsupervised learning. They proposed an innovative approach for predicting gene function based on several heterogeneous networks, known as Deep MNE-CNN. The Deep MNE-CNN design was primarily composed of two components. The semi AE was employed in the first section to combine various networks and produce low-dimensional feature representations of the genes. The semi AE was trained in a semi-supervised manner using the labelled and unlabeled data. The association between genes was captured using the pairwise correlation coefficients, which further increased the precision of the feature learning procedure. On yeast and human datasets, the Deep MNE-CNN method's performance was assessed and contrasted with that of four cutting-edge techniques. The outcomes demonstrated that the Deep MNE-CNN method performed better in prediction accuracy than the other methods.

Xu et al.[133] used the autoencoder dimensionality is reduced by feeding the autoencoder a low-dimensional representation of the input gene expression data, which is then used for matrix factorization. This method emphasizes the potential of dimensionality reduction methods based on autoencoders for handling high-dimensional gene expression data. The autoencoder can identify significant patterns and correlations within the data while minimizing noise and redundancy by compressing the data into a lower-dimensional representation.

Zeng et al.[134] proposed a DL framework, the deep matrix factorization model DMFLDA, to forecast lncRNA-disease relationships. Deep autoencoders, feature learning, and disease semantic similarity are all used in the framework to effectively learn the low-dimensional

representations of lncRNA and disease features. The learned representations predict the associations between lncRNAs and disease and then fed into a classification model. The research illustrates the promise of DL-based techniques for analyzing high-dimensional genomic data and demonstrates the efficacy of DMFLDA in identifying lncRNA-disease connections. The authors used AUC to assess the model's performance. The DMFLDA model had strong predictive performance, evidenced by its AUC score of 0.8393. In addition, the DMFLDA model beat numerous cutting-edge approaches, including RWRMDA, SimNMF, and LRLSLDA, to predict lncRNA-disease relationships. According to the comparative analysis, DMFLDA has much higher prediction accuracy, sensitivity, specificity, and precision.

Zhao et al.[135] proposed an ML method utilizing k-means dimensional reduction for predicting survival outcomes in BC patients. The recommended method reduces the high-dimensional gene expression data to a lower-dimensional framework using k-means clustering. The support vector machine (SVM) classifier is trained on the reduced data to forecast patient survival outcomes.

Furey et al.[136] used PCA and linear discriminant analysis for dimensionality reduction. The authors train an SVM classifier for cancer classification using the reduced-dimensional data as input. The suggested strategy is assessed and contrasted with other cutting-edge approaches on several benchmark datasets. The outcomes demonstrate that the suggested strategy performs well in accuracy and, in some circumstances, outperforms alternative approaches.

Gu et al.[137] used heterogeneous graph neural networks (HGNN), a technique for integrating various forms of genomic data, such as protein-protein interaction networks and gene expression networks, using an HGNN. By mixing information from several modalities and using the connections between various forms of data, the HGNN develops a low-dimensional representation of the genomic data.

CRITICAL DISCUSSION

First, by combining various genetic variations and other risk factors, using aiPRS for CVD risk prediction can improve the accuracy of predicting a person's likelihood of developing CVD. Second, aiPRS models can get beyond the limits of conventional PRS calculators by using a more extensive range of genetic and non-genetic factors in the risk assessment process, leading to more accurate and unique risk estimates. Third, by employing AI approaches to reduce the dimensionality of massive genomic datasets, disease risk prediction models can be improved and made more effective.

Benchmarking analysis

Table 2 provides an overview of studies conducted on different statistical/AI/ML algorithms used for the construction of PRS for different diseases.[84,132,138-143]

Vilhjálmsson et al.[138] proposed a novel statistical method called LD pred-funct. The purpose of their study was to improve the accuracy of PRS by incorporating linkage disequilibrium (LD) information. They used LDpred to analyse several sizable datasets and evaluated their effectiveness in comparison to other approaches. They concentrated on phenotypic prediction for conditions such as type 1 diabetes, type 2 diabetes, rheumatoid arthritis, CAD, Crohn's disease, hypertension, and bipolar disorder. When predicting these traits, LDpred

Table 2. Benchmarking analysis of studies

No.	Citations	Year	PRS	Dataset	Disease	Statistical/AI/ML	Primary features	Secondary features	Challenges	TR
1	Vilhjálmsson et al. [138]	2015	LDpred	WTCCC	BD, CAD, CD, HT, RA, T1D, T2D	LD	Use of a novel statistical method called LD pred-funct, testing the method on several large-scale datasets, and comparison of LDpred-funct with other methods.	Modelling LD can improve the accuracy of polygenic risk scores; the study focused on predicting height, body mass index, and breast cancer risk, using both European and African populations.	LDpred may overfit on the training data; it requires accurate estimates of LD, it may not perform well for small sample sizes, it may not perform well for rare variants.	65
2	Privé et al. [139]	2019	PRS	UKB	BRCA	PLR	Efficient implementation of penalized regression, and fast computation, can handle a large number of genetic variants.	Easy-to-use, can incorporate external information as priors, and improved performance compared to other methods.	Limited flexibility in modelling complex genetic effects, and difficulty in selecting optimal tuning parameters.	39
3	Leonenko et al. [140]	2019	PRS/PHS	GERAD, IGAP	AD	CR	PRS and PHS models for AD prediction, comparison of PRS and PHS with APOE status, validation in independent cohorts.	Use summary statistics from large GWAS studies, construction of different PHS models, and analysis of gene-environment interactions.	Limited prediction accuracy, lack of understanding of underlying biology, and ethical concerns regarding use of genetic information.	22
4	Mavaddat et al. [84]	2019	PRS-LPR	UKB	BRCA	LPR	Development of PRS models for breast cancer and subtypes, use of large-scale GWAS data, and evaluation of PRS performance in different populations.	Investigating the contribution of individual SNPs to PRS and comparing PRS with clinical risk factors.	Limited representation of non-European populations, and limited information on environmental factors.	34
5	Choi and O'Reilly [141]	2019	PRSice-2	Various large-scale datasets	Various	C+T	PRSice-2 software for polygenic risk scoring, support for biobank-scale data, incorporation of LD information, and multiple testing correction.	Support for different reference panels, plotting functions and customizable thresholds, and parallelization for faster computation.	Limited support for non-European populations, dependence on LD reference panels, and interpretation of PRS results can be compelling.	31
6	Ge et al. [142]	2019	PRS-CS	PHCB, UKB	BRCA, CAD, DEP, IBD, RA, T2D	BR	Continuous shrinkage prior, multivariate block update of effect sizes, improved LD adjustment.	Automatic learning of shrinkage, genome partitioning, tuning parameter \xcf\x95.	Optimizing global shrinkage, block updates, and capturing LD.	66
7	Huang et al. [143]	2021	DL-PRS	UKB	COPD	FCDNN	A novel deep learning approach for PRS, utilizes dense neural networks.	Achieves higher accuracy than traditional PRS methods and can handle large-scale datasets.	Requiring a large amount of training data may not be feasible for small-scale studies.	26
8	Peng et al. [132]	2021	DeepPRS	UKB	IBD, T2D, AD, BRCA	Bi-LSTM	Deep learning model, GWAS data, identification of individuals at risk for common diseases.	Data pre-processing, model training and testing, and comparison with traditional PRS methods.	Limited availability of large-scale, diverse datasets, interpretability of the model, and ethical considerations regarding genetic risk prediction.	45

PRS = Polygenic Risk Score, AI = artificial intelligence, ML = machine learning, TR = total references, LD = linkage disequilibrium, WTCCC = Wellcome Trust Case Control Consortium, BD = bipolar disorder, CAD = coronary artery disease, CD = Crohn's disease, HT = hypertension, RA= rheumatoid arthritis, T1D = type 1 diabetes mellitus, T2D = type 2 diabetes mellitus, UKB = United Kingdom Biobank, BRCA = breast cancer, LPR = Lasso penalized regression, PLR = penalized logistic regression, GERAD = Genetic and Environmental Risk in Alzheimer's Disease Consortium, IGAP = International Genomics of Alzheimer's Project, AD = Alzheimer's disease, CR = cox regression, PHS = Polygenic Hazard Score, APOE = apolipoprotein E, GWAS = Gene Wide Association Studies, SNP = single nucleotide polymorphism, C+T = clumping+thresholding, PRS-CS = Polygenic Risk Score–Continuous Shrinkage, PHCB = Partners HealthCare Biobank, DEP = depression, IBD = inflammatory bowel disease, BR = Bayesian regression, FCDNN = fully connected deep neural network, Bi-LSTM = bidirectional long short term memory.

outperformed other approaches due to the inclusion of LD information. They used both European and African populations to predict height, body mass index, and BC risk. Modelling LD patterns—the non-random connection of alleles at various loci—allows LDpred to act on the genome. The goal of LDpred is to enhance the predictive accuracy of PRS by capturing the associated effects of genetic variations. However, the study highlighted some LDpred flaws. First, LDpred might overfit the training data, which would mean that it might perform incredibly well there but struggle to generalise to new data. Correct LD calculations are necessary for the strategy to function. Additionally, LDpred could not perform well with small sample sizes since the little data might make it impossible to faithfully replicate LD patterns. Finally, LDpred might not be the ideal technique for predicting uncommon variations because they might have different LD patterns than common ones.

Privé et al.[139] calculated PRS on UK Biobank dataset using the Penalized Logistic Regression. The goal of the study was to manage a high number of genetic variants while increasing the computing efficiency of PRS computations. They were able to lessen the processing strain and speed up the computation process by adopting penalized regression. Their strategy's user-friendliness was a major plus. The technique was created to be simple to apply, allowing researchers to calculate PRS rather quickly. Additionally, their method permitted the insertion of outside data as priors, which might improve the precision of PRS predictions. Compared to other approaches, their method performed better, which suggested that it accurately identified the genetic contributions to BC risk. However, the study highlighted some flaws. Their strategy was not very adaptable for simulating intricate genetic processes. Although it handled a lot of genetic variants well, it might not have captured more complex linkages or interactions between genetic factors. Additionally, it is challenging to choose the best tuning settings for the penalised regression model. The level of penalization and regularisation used throughout the model fitting phase is controlled by tuning parameters. To get the optimum performance, choosing the proper parameters can be difficult and requires significant thought.

Leonenko et al.[140] evaluated the age-specific genetic risk for AD using the PRS and compare the findings with the Polygenic Hazard Score (PHS). They used the data from IGAP (IGAP: International Genomics of Alzheimer's Project) and GERAD (Genetic and Environmental Risk in Alzheimer's Disease Consortium). Their goal was to measure individual variations in age-specific genetic risk for AD and to assess how well PRS and PHS can predict the development of the disease. Based on SNPs found in GWAS that were related with AD, PRS and PHS were determined for each participant. The age-specific genetic risk for AD with PRS was quantified using Cox Regression analysis, and PHS scores were calculated for the same individuals. The usefulness of PRS and PHS in determining the genetic risk for AD that is age-specific was established in this study. Even though the results showed that PRS based on genome-wide significant SNPs showed the strongest association, more research is required to examine the specific benefits and restrictions of PHS compared to PRS in the prediction of AD risk, especially when considering various SNP selection criteria and the Cox Proportional Hazard Regression model.

Mavaddat et al.[84] used the Lasso Penalized Regression (LPR) technique to choose pertinent factors and account for overfitting when constructing the PRS. LPR finds the most useful SNPs related to BC. While minimising the impact of noise or irrelevant genetic variants, the PRS-LPR approach's usage of LPR enabled for the identification of a subset of SNPs that contribute the most to the prediction of BC risk. They built PRS models using extensive

GWAS data and assessed how well they performed at predicting the risk of BC. The study showed PRS increased risk prediction beyond only clinical risk variables. However, there were drawbacks, such as the need for more thorough environmental data and the inadequate representation of non-European groups.

Choi and O'Reilly[141] made a software PRSice-2 similar to PRSice to calculate the PRS by using the "C+T" approach, which clumps single-nucleotide polymorphisms (SNPs) based on LD and *P* value thresholding. The key characteristics of PRSice-2 are its capacity to conduct large-scale PRS analyses on genotyped and imputed data, compute empirical association *P* values to address overfitting, analyse numerous target phenotypes concurrently, and provide options for imputing missing genotypes. It supports various inheritance models (additive, dominant, recessive, and heterozygous) and automatically creates dummy variables for categorical covariates. The complexity of PRSice-2, which offers a variety of features and options for PRS analysis, is one possible drawback. Because of this intricacy, users may need to have a certain level of experience in genomics and statistical analysis in order to completely understand and use the application. Additionally, because PRSice-2 uses several different parameters and calculations, including clumping and *P* value thresholding, comprehending the findings it produces can be difficult. To achieve proper interpretation and valuable insights from the analysis, adequate knowledge of these factors is required.

Ge et al.[142] used the Bayesian Regression framework and GWAS summary statistics to develop PRS-CS. The inclusion of a continuous shrinkage (CS) prior on SNP effect sizes is a crucial component of PRS-CS. The heterogeneity in genomic layouts across many traits and disorders is addressed by this prior. The PRS-CS method improves resilience in its predictions and offers computational advantages over other approaches by utilising the CS prior. The Partners HealthCare Biobank was used to use PRS-CS to estimate the probability of six prevalent complicated diseases and six quantitative features. In comparison to other approaches, the results demonstrated PRS-CS's superiority in terms of prediction accuracy. By adding a CS prior and a Bayesian regression framework, PRS-CS represents a development in the field of polygenic prediction. The accuracy of risk prediction is improved by its capacity to capture local patterns of LD and adapt to various genomic architectures.

Huang et al.[143] designed a DL neural network strategy for GWAS and PRS. The condition known as chronic obstructive pulmonary disease (COPD) is complicated and varied, impacted by both genetic and environmental factors. Traditional techniques, such as GWAS and PRS, have proved effective in locating risk variations, but they operate under the supposition that each allele's effects are independent and unaffected by other variables. To get over these constraints DL-PRS was made. It makes less assumptions regarding the genetic effects that are being modelled since DL models can capture gene-gene interactions and non-additive effects. Several populations based GWAS studies of COPD were used by the researchers to apply the DL technique to genetic association data. Comparing the DL-PRS method to other PRS methods, it showed superior predictive ability, expanding the ranges of risk prediction and establishing a stronger connection with lung function measurements. However, DL-PRS implementation requires a significant investment in processing power and DL technique knowledge. This study emphasizes how ML methods could improve risk prediction models for difficult-to-diagnose conditions like COPD.

Peng et al.[132] developed a DL model called DeepPRS using the GWAS data obtained from UK Biobank to identify people at risk of four different diseases such as IBD, BRCA, T2D, AD

using the Bi-LSTM. It was trained to discover the intricate patterns and relationships between genetic variations and disease risk using the GWAS data as its input. The process included several crucial components. The GWAS data first underwent pre-processing, which included quality control steps to guarantee data dependability and integrity. To model the links between genetic variations and disease outcomes, the Bi-LSTM architecture was then used. A vast amount of labelled data was used to train the model where everyone's genetic makeup and disease state were known. They evaluated DeepPRS effectiveness by contrasting its prognostic capabilities with those of conventional PRS techniques. They assessed DeepPRS's precision, sensitivity, and specificity in predicting the risk of developing IBD, T2D, AD, and BRCA diseases. To learn more about the genetic factors impacting disease susceptibility, they also looked at the model's interpretability. However, there are some challenges associated with it. The lack of large-scale, diverse datasets was one of the primary drawbacks, which could have affected the model's resilience and generalizability. The DL model's interpretability presented another difficulty because of its complicated design, which may make it challenging to comprehend the precise genetic elements influencing illness risk estimates.

A short note on AI for PRS

The PRS is a technique for estimating a person's genetic risk for contracting a complex disease. PRS creates a composite score that can be used to calculate an individual's total genetic susceptibility to the disease by combining data from numerous genetic variants, each of which has a negligible impact on disease risk. Due to improvements in genotyping technology and the accessibility of extensive genetic data from GWAS, PRS has grown in popularity in recent years.[9] However, determining PRS precisely can be difficult, especially when working with big datasets including millions of genetic variants. AI is useful in this situation. The use of AI techniques has improved the precision and effectiveness of PRS estimates, especially ML models.[144] The creation of more precise PRS models is made possible by the ability of these models to analyze massive amounts of genetic data and pinpoint the key genetic variants for determining disease risk.

The gradient boosting machine (GBM) is one ML method applied to PRS computation.[145] Even in noisy or correlated data, the potent ML algorithm GBM can precisely identify the most pertinent genetic variations for predicting disease risk. For a variety of complicated diseases, including T2D,[146,147] BC,[84] and AD,[148] GBM has been utilized to create PRS models. The neural network is another ML method that has demonstrated potential in PRS computation.[149] A DL algorithm known as a neural network can recognize intricate patterns in genetic data and produce precise forecasts of disease risk. A PRS model for colorectal cancer was created in one study using a neural network, and it performed more accurately than conventional PRS models.[150] AI has been used to find new genetic variations linked to illness risk and increase PRS accuracy. For instance, a recent study that examined GWAS data for schizophrenia using a deep neural network discovered numerous unique genetic variations linked to the condition.[151]

The role of bias and variance in AI

The assessment of bias and variance in AI models has gained significant importance in recent years.[152,153] Previous computer-aided diagnosis techniques have revealed shortcomings not only in evaluating bias but also in managing variance effectively.[154] To address these challenges comprehensively, a multifaceted approach is essential. To mitigate both bias and variance, a range of strategies can be employed.[70,155] Utilizing large sample sizes can help in reducing bias by ensuring a more representative dataset, while also aiding in mitigating

variance by providing a broader data spectrum. Conducting appropriate clinical testing is crucial for evaluating model performance under different conditions, thereby tackling both bias and variance issues.

Additionally, utilizing big data configurations can assist in minimizing bias by capturing a more comprehensive view of the data distribution, while it also has implications for managing variance by introducing higher dimensionality and variability. Analyzing unseen data is vital to uncover both bias and variance, as it helps identify how well the model generalizes to new, unobserved cases. Finally, scientifically validating the training model design plays a crucial role in reducing both bias and variance, as a well-designed model is less prone to systematic errors and overfitting. Key steps in patient risk stratification encompass assessing not only the AI risk of bias[155-157] but also considering the AI risk of variance (RoV). It is essential to appropriately modify diagnostics and treatment plans to account for both bias and variance in AI models. This holistic approach ensures that AI-based medical decisions are not only fair but also consistently reliable across different patient populations and scenarios. In summary, addressing bias and variance in AI models is essential for achieving both fairness and reliability in healthcare applications. By implementing a combination of strategies and considering both aspects, we can enhance the quality of AI-driven diagnostics and patient risk stratification.

The role of explainability in AI

Understanding how AI's "black box" functions are critical for effective AI design. The role of AI Explainability practitioners is more likely to comprehend this "black box" if the results it produces can be interpreted and questioned.[158] By employing tools such as Local Interpretable Model-Agnostic Explanations and Shapley Additive Explanations, AI models can provide insights into complex disorders, which has garnered trust among medical professionals.[159,160] Additionally, techniques like GradCAM, GradCAM+, or GradCAM++ can be utilized to visualize carotid lesions and facilitate wider acceptance of AI models in the medical domain.[161] This emphasis on interpretability enables the improvement and cost-effectiveness of AI devices.[162]

The role of cloud-based paradigms

Cloud-based XAI, can be used for calculating the PRS.[163] The PRS is a score calculated based on an individual's genetic information to estimate their risk of developing certain diseases or conditions.[164] Cloud-based XAI refers to a system where ML algorithms are hosted on cloud servers and can be accessed remotely through the internet.[165,166] Consequently, processing massive amounts of data—like genetic data from a vast population—can be done more effectively and flexibly. In order to calculate PRS using cloud-based XAI, the algorithm would examine a sizable dataset of genetic data to pinpoint the precise genetic markers linked to a higher risk for a specific disease or condition.[167] Based on each person's unique genetic profile, the algorithm would utilise this data to determine a PRS for each person in the dataset. Using cloud based XAI for PRS computation is advantageous since it makes the findings more transparent and understandable. By using explainable AI, the algorithm can provide a detailed explanation of how it arrived at a particular PRS for an individual, making it easier for healthcare providers and patients to understand and make informed decisions about their health. Overall, cloud based XAI for PRS calculation can help to improve personalized medicine by providing more accurate and transparent risk assessments based on an individual's genetic information.

The role of pruning in AI systems

With the development of the internet and cloud-based systems, edge devices are becoming increasingly important.[168] In mobile frameworks, these devices are crucial for utilising trained AI models for future predictions or illness risk stratifications.[169] Compressed models must be used because it may not be possible to install huge data models on edge devices.[131] Image-based DL models such as Fully Convolutional Networks or Segmentation Networks can be pruned using evolutionary algorithms like particle swarm optimization, genetic algorithms, wolf optimization, and differential evolution.[170] The future of genetically-based paradigms and radiomics-based CVD risk stratification can be compacted and implemented on edge devices to serve rural areas, especially in developing countries.[171]

The role of big data

The focus of study has shifted from radiography and pathology to genetics as a result of the development of radiogenomics.[172] Through the fusion of radiomics, genetic information, and clinical records, this change has led to the expansion of the radiogenomics field over the past ten years.[173] By developing novel algorithms, procedures, and techniques, DL and big data programming have considerably advanced radiogenomics research.[174] The creation of a completely automated system that interfaces with a radiological process in the big data realm is a significant advancement in the area.[175] By decreasing the amount of time spent on tedious and repetitive tasks, this approach has increased productivity.[176,177] The ability to compare numerous images from the database concurrently allows for real-time therapy monitoring.[175,177]

Special note on generalization of aiPRS

The use of AI in medical diagnosis, such as the detection and prediction of CVD, is one example of how generalisation in AI models can go beyond specialized domains and include a wide range of topics. The objective of generalization in this context is to create models capable of efficiently learning from a wide range of patient data, such as symptoms, medical history, and diagnostic tests, and producing precise predictions about unrecognized CVD cases. These AI models could help medical professionals by assisting in the early diagnosis, risk assessment, and personalized treatment recommendations for individuals with CVD. Generalization is vital in other AI disciplines in addition to medical applications. For instance, in natural language processing, even if a generalized language model has never encountered a particular sentence before, it may comprehend and produce coherent statements across a variety of themes.[178] The same is true in computer vision, where a well-generalized model can identify items or patterns in photos, such as skin tones, objects in various surroundings, and more.[124,179-181] Effective generalization enables AI models to behave consistently in real-world circumstances outside the bounds of their training data. It is important to note that finding the ideal balance between memorization and generalization is essential in the development of AI.[182] A model's ability to perform well on fresh, untested data might be hampered by an excessive reliance on memorization without generalization. While some memorizing may be useful in capturing details or personalized preferences. AI systems can adapt and make precise predictions in a variety of changing situations by aiming for generalization.[183,184] To promote generalization and reduce overfitting, which is a type of memorizing, researchers and practitioners use strategies including regularization, cross-validation, and early stopping. AI solutions have a stronger and better ability to generalize compared to conventional-based methods, making them suitable for composite data analysis in CVD risk assessment. aiPRS leverage the power of AI on composite data for CVD risk assessment.[37,40,185] This composite risk combines the PRS values derived from gene data and a combination of other biomarkers such as office-based, laboratory-based,

radiomics-based, and genomics-based.[186] Office-based biomarkers are the conventional risk biomarkers that constitute basic biomarkers namely height, weight, body mass index, hypertension values (systolic and diastolic blood pressure), smoking conditions, and family conditions. Laboratory-based biomarkers are blood-based conventional biomarkers namely cholesterol levels such as LDL, HDL, Total Cholesterol, triglycerides, fasting glucose, renal biomarkers like eGFR, arthritis biomarkers namely ESR, and homocysteine. Radiomics-based biomarkers, it includes the actual image-based biomarkers namely direct or surrogate biomarkers namely carotid plaque burden, carotid-intima thickness (maximum, minimum, and average), maximum carotid plaque height, and intima-media thickness variability. Genomics is the other risky genetic atherosclerotic biomarker.

When all such data are composited, the system becomes nonlinear and that is when the supervised DL systems are well fit for CVD risk prediction. Thus, it does not matter if the AI system is fed by the UK bio-bank data sets or ERIC database, or NIH-based data sets. It is worth noting that the DL layers allow to extract the powerful features from the composite data which can then be used for precise CVD risk prediction. It is the forward and backward propagation of the neural networks which allows to reduce the error between the weight propagation values and the gold standard. The number of iterations during the propagation (so-called epochs), besides the batch size in the DL system are supportive ingredients for a preventive, precise, and personalized (aiP^3) approach for CVD risk estimation.[41,45,187,188]

Strengths, weaknesses, and extensions

Conventional PRS is a straightforward and economical method that uses genetic data already available, but it is limited in its capacity to predict complex features since it cannot take complicated data interactions into account. Contrarily, aiPRS can effectively analyze multidimensional data, consider complicated data interactions, and include genetic, clinical, and imaging data. But it necessitates specialized knowledge and depends on reliable and readily available data. The usage of aiPRS may also raise privacy and ethical concerns. The use of AI in PRS computing and analysis offers enormous potential to improve our understanding of complex diseases and identify novel genetic targets for disease prevention and treatment. Despite the benefits of AI-based risk stratification, it is necessary to note its significant flaws and problems. The potential lack of generalizability of AI models is a serious flaw. AI models may not perform as well when applied to diverse populations or various circumstances if they were designed using certain datasets or populations.[55,178]

To evaluate the generalizability and dependability of AI models in practical settings, comprehensive validation experiments involving a variety of populations must be conducted. The scant scientific evidence that supports the efficacy of AI-based risk prediction models is another area of weakness. While AI has potential, a thorough scientific examination is required to determine its therapeutic utility and dependability. It is crucial to conduct prospective validation studies comparing AI models to current risk prediction techniques and evaluating how well they function in actual clinical situations. These research projects will offer the required scientific proof to back up the adoption and application of AI-based risk prediction models. To enable the ethical incorporation of AI in risk classification, these flaws must be addressed. The limits of AI-based risk prediction models can be better recognized and reduced by undertaking validation studies across varied populations and producing empirical evidence of their efficacy. As a result, AI models will become more applicable and trustworthy, which will increase their ability to help with illness management and preventative plans.

More research is necessary, however, to accurately evaluate the dependability and accuracy of aiPRS models and ensure that these models are developed and applied ethically and openly. Lastly, we do need to understand the changes in aiPRS for CVD risk when comorbidities are involved such as viruses.[189] Even though aiPRS technology is point data, but as things evolve, we will see gene data to get converted to images and screening tools can be applied,[190,191] such as usage of the classification tools.[192] Further, correlations need to be developed between the aiPRS and cardiovascular outcomes.[124,193]

CONCLUSIONS

aiPRS modeling approaches offer enormous potential to enhance CVD personalized treatment approaches. Integrating multi-omics data, including genomes, transcriptomics, proteomics, and metabolomics, can offer a more thorough understanding of the molecular pathways driving CVD and increase the precision of PRS models.[194] These AI-based techniques are especially beneficial for drawing out intricate patterns and connections from highly dimensional data genetic data. This makes it easier and more precise for scientists and doctors to find real connections between genetic variations and the likelihood of developing diseases. Additionally, AI algorithms may continuously learn from fresh data and enhance their performance, producing more accurate and unique risk evaluations over time.

The need for more representative and diverse data, new data analysis techniques, and ethical considerations remain major obstacles. More funding must be put into CVD genomics research and personalized medicine techniques to overcome these obstacles and improve patient outcomes. We can enhance our understanding of the intricate disease mechanisms underlying CVD and create more efficient methods for prevention, diagnosis, and therapy by fostering large-scale genomic data resources that are openly accessible and available.

ACKNOWLEDGMENTS

REFERENCES

1. Virani SS, Alonso A, Benjamin EJ, Bittencourt MS, Callaway CW, Carson AP, et al. Heart disease and stroke statistics—2020 update: a report from the American Heart Association. *Circulation* 2020;141(9):e139-596.
   PUBMED | CROSSREF

2. Roth GA, Johnson CO, Abate KH, Abd-Allah F, Ahmed M, Alam K, et al. The burden of cardiovascular diseases among US states, 1990–2016. *JAMA Cardiol* 2018;3(5):375-89.
   PUBMED | CROSSREF

3. O'Donnell CJ, Nabel EG. Genomics of cardiovascular disease. *N Engl J Med* 2011;365(22):2098-109.
   PUBMED | CROSSREF

4. Gluba A, Banach M, Mikhailidis DP, Rysz J. Genetic determinants of cardiovascular disease: the renin-angiotensin-aldosterone system, paraoxonases, endothelin-1, nitric oxide synthase and adrenergic receptors. *In Vivo* 2009;23(5):797-812.
   PUBMED

5. Barua JD, Omit SB, Rana HK, Podder NK, Chowdhury UN, Rahman MH. Bioinformatics and system biological approaches for the identification of genetic risk factors in the progression of cardiovascular disease. *Cardiovasc Ther* 2022;2022:9034996.
PUBMED | **CROSSREF**

6. Phan JH, Quo CF, Wang MD. Cardiovascular genomics: a biomarker identification pipeline. *IEEE Trans Inf Technol Biomed* 2012;16(5):809-22.
PUBMED | **CROSSREF**

7. Maniruzzaman M, Jahanur Rahman M, Ahammed B, Abedin MM, Suri HS, Biswas M, et al. Statistical characterization and classification of colon microarray gene expression data using multiple machine learning paradigms. *Comput Methods Programs Biomed* 2019;176:173-93.
PUBMED | **CROSSREF**

8. Shah S, Henry A, Roselli C, Lin H, Sveinbjörnsson G, Fatemifar G, et al. Genome-wide association and Mendelian randomisation analysis provide insights into the pathogenesis of heart failure. *Nat Commun* 2020;11(1):163.
PUBMED | **CROSSREF**

9. Visscher PM, Brown MA, McCarthy MI, Yang J. Five years of GWAS discovery. *Am J Hum Genet* 2012;90(1):7-24.
PUBMED | **CROSSREF**

10. Gallagher MD, Chen-Plotkin AS. The post-GWAS era: from association to function. *Am J Hum Genet* 2018;102(5):717-30.
PUBMED | **CROSSREF**

11. Nolte IM, Munoz ML, Tragante V, Amare AT, Jansen R, Vaez A, et al. Genetic loci associated with heart rate variability and their effects on cardiac disease risk. *Nat Commun* 2017;8(1):15805.
PUBMED | **CROSSREF**

12. Wu JH, Lemaitre RN, Manichaikul A, Guan W, Tanaka T, Foy M, et al. Genome-wide association study identifies novel loci associated with concentrations of four plasma phospholipid fatty acids in the de novo lipogenesis pathway: results from the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium. *Circ Cardiovasc Genet* 2013;6(2):171-83.
PUBMED | **CROSSREF**

13. Ripatti S, Tikkanen E, Orho-Melander M, Havulinna AS, Silander K, Sharma A, et al. A multilocus genetic risk score for coronary heart disease: case-control and prospective cohort analyses. *Lancet* 2010;376(9750):1393-400.
PUBMED | **CROSSREF**

14. El-Baz AS, Suri JS. *Cardiovascular and Coronary Artery Imaging: Volume 1*. Cambridge, MA, USA: Academic Press; 2021.

15. Khanna NN, Maindarkar M, Puvvula A, Paul S, Bhagawati M, Ahluwalia P, et al. Vascular implications of COVID-19: role of radiological imaging, artificial intelligence, and tissue characterization: a special report. *J Cardiovasc Dev Dis* 2022;9(8):268.
PUBMED | **CROSSREF**

16. Elliott P, Chambers JC, Zhang W, Clarke R, Hopewell JC, Peden JF, et al. Genetic Loci associated with C-reactive protein levels and risk of coronary heart disease. *JAMA* 2009;302(1):37-48.
PUBMED | **CROSSREF**

17. Khanna NN, Maindarkar M, Saxena A, Ahluwalia P, Paul S, Srivastava SK, et al. Cardiovascular/stroke risk assessment in patients with erectile dysfunction-a role of carotid wall arterial imaging and plaque tissue characterization using artificial intelligence paradigm: a narrative review. *Diagnostics (Basel)* 2022;12(5):1249.
PUBMED | **CROSSREF**

18. Suri JS, Paul S, Maindarkar MA, Puvvula A, Saxena S, Saba L, et al. Cardiovascular/stroke risk stratification in Parkinson's disease patients using atherosclerosis pathway and artificial intelligence paradigm: a systematic review. *Metabolites* 2022;12(4):312.
PUBMED | **CROSSREF**

19. Cahill TJ, Ashrafian H, Watkins H. Genetic cardiomyopathies causing heart failure. *Circ Res* 2013;113(6):660-75.
PUBMED | **CROSSREF**

20. Hucker WJ, Saini H, Lubitz SA, Ellinor PT. Atrial fibrillation genetics: is there a practical clinical value now or in the future? *Can J Cardiol* 2016;32(11):1300-5.
PUBMED | **CROSSREF**

21. Weiss JC, Natarajan S, Peissig PL, McCarty CA, Page D. Machine learning for personalized medicine: predicting primary myocardial infarction from electronic health records. *AI Mag* 2012;33(4):33-33.
CROSSREF

22. Shameer K, Johnson KW, Glicksberg BS, Dudley JT, Sengupta PP. Machine learning in cardiovascular medicine: are we there yet? *Heart* 2018;104(14):1156-64.
    PUBMED | CROSSREF

23. Kagiyama N, Shrestha S, Farjo PD, Sengupta PP. Artificial intelligence: practical primer for clinical research in cardiovascular disease. *J Am Heart Assoc* 2019;8(17):e012788.
    PUBMED | CROSSREF

24. Alimadadi A, Manandhar I, Aryal S, Munroe PB, Joe B, Cheng X. Machine learning-based classification and diagnosis of clinical cardiomyopathies. *Physiol Genomics* 2020;52(9):391-400.
    PUBMED | CROSSREF

25. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015, 18th International Conference; October 5-9, 2015; Munich, Germany. Berlin, Germany: Springer; 2015, 234-41.

26. Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; June 7-12, 2015; Boston, MA, USA. Washington, D.C., USA: IEEE Computer Society; 2015, 3431-40.

27. Badrinarayanan V, Kendall A, Cipolla R. Segnet: a deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans Pattern Anal Mach Intell* 2017;39(12):2481-95.
    PUBMED | CROSSREF

28. Noh H, Hong S, Han B. Learning deconvolution network for semantic segmentation. Proceedings of the IEEE International Conference on Computer Vision; December 7-13, 2015; Santiago, Chile. Piscataway, NJ, USA: Institute of Electrical and Electronics Engineers; 2015, 1520-8.

29. Karim F, Majumdar S, Darabi H, Harford S. Multivariate LSTM-FCNs for time series classification. *Neural Netw* 2019;116:237-45.
    PUBMED | CROSSREF

30. Xia M, Yan W, Huang Y, Guo Y, Zhou G, Wang Y. Extracting membrane borders in IVUS images using a multi-scale feature aggregated u-net. 2020. 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC); July 20-24, 2020; Montreal, Canada. Piscataway, NJ, USA: Institute of Electrical and Electronics Engineers; 2020, 1650-3.

31. Azad R, Asadi-Aghbolaghi M, Fathy M, Escalera S. Bi-directional ConvLSTM U-Net with densley connected convolutions. Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops; October 27-28, 2019; Seoul, Korea. Piscataway, NJ, USA: Institute of Electrical and Electronics Engineers; 2019, 406-15.

32. Wollmann T, Gunkel M, Chung I, Erfle H, Rippe K, Rohr K. GRUU-Net: Integrated convolutional and gated recurrent neural network for cell segmentation. *Med Image Anal* 2019;56:68-79.
    PUBMED | CROSSREF

33. Adak A, Pradhan B, Shukla N, Alamri A. Unboxing deep learning model of food delivery service reviews using explainable artificial intelligence (XAI) technique. *Foods* 2022;11(14):2019.
    PUBMED | CROSSREF

34. Deif MA, Solyman AA, Kamarposhti MA, Band SS, Hammam RE. A deep bidirectional recurrent neural network for identification of SARS-CoV-2 from viral genome sequences. *Math Biosci Eng* 2021;18(6):8933-50.
    PUBMED | CROSSREF

35. Suri JS, Bhagawati M, Agarwal S, Paul S, Pandey A, Gupta SK, et al. UNet deep learning architecture for segmentation of vascular and non-vascular images: a microscopic look at UNet components buffered with pruning, explainable artificial intelligence, and bias. *IEEE Access* 2022;11:595-645.
    CROSSREF

36. Libiseller-Egger J, Phelan JE, Attia ZI, Benavente ED, Campino S, Friedman PA, et al. Deep learning-derived cardiovascular age shares a genetic basis with other cardiac phenotypes. *Sci Rep* 2022;12(1):22625.
    PUBMED | CROSSREF

37. Johri AM, Mantella LE, Jamthikar AD, Saba L, Laird JR, Suri JS. Role of artificial intelligence in cardiovascular risk prediction and outcomes: comparison of machine-learning and conventional statistical approaches for the analysis of carotid ultrasound features and intra-plaque neovascularization. *Int J Cardiovasc Imaging* 2021;37(11):3145-56.
    PUBMED | CROSSREF

38. Krittanawong C, Johnson KW, Choi E, Kaplin S, Venner E, Murugan M, et al. Artificial intelligence and cardiovascular genetics. *Life (Basel)* 2022;12(2):279.
    PUBMED | CROSSREF

39. El-Baz A, Suri JS. *Big Data in Multimodal Medical Imaging*. Boca Raton, FL, USA: CRC Press; 2019.

40. Jamthikar AD, Gupta D, Mantella LE, Saba L, Laird JR, Johri AM, et al. Multiclass machine learning vs. conventional calculators for stroke/CVD risk assessment using carotid plaque predictors with coronary angiography scores as gold standard: a 500 participants study. *Int J Cardiovasc Imaging* 2021;37(4):1171-87.
    PUBMED | CROSSREF

41. Jamthikar A, Gupta D, Khanna NN, Saba L, Laird JR, Suri JS. Cardiovascular/stroke risk prevention: a new machine learning framework integrating carotid ultrasound image-based phenotypes and its harmonics with conventional risk factors. *Indian Heart J* 2020;72(4):258-64.
PUBMED | CROSSREF

42. Steinfeldt J, Buergel T, Loock L, Kittner P, Ruyoga G, Zu Belzen JU, et al. Neural network-based integration of polygenic and clinical information: development and validation of a prediction model for 10-year risk of major adverse cardiac events in the UK Biobank cohort. *Lancet Digit Health* 2022;4(2):e84-94.
PUBMED | CROSSREF

43. Johri AM, Singh KV, Mantella LE, Saba L, Sharma A, Laird JR, et al. Deep learning artificial intelligence framework for multiclass coronary artery disease prediction using combination of conventional risk factors, carotid ultrasound, and intraplaque neovascularization. *Comput Biol Med* 2022;150:106018.
PUBMED | CROSSREF

44. Konstantonis G, Singh KV, Sfikakis PP, Jamthikar AD, Kitas GD, Gupta SK, et al. Cardiovascular disease detection using machine learning and carotid/femoral arterial imaging frameworks in rheumatoid arthritis patients. *Rheumatol Int* 2022;42(2):215-39.
PUBMED | CROSSREF

45. Jamthikar A, Gupta D, Johri AM, Mantella LE, Saba L, Suri JS. A machine learning framework for risk prediction of multi-label cardiovascular events based on focused carotid plaque B-Mode ultrasound: a Canadian study. *Comput Biol Med* 2022;140:105102.
PUBMED | CROSSREF

46. Ho DS, Schierding W, Wake M, Saffery R, O'Sullivan J. Machine learning SNP based prediction for precision medicine. *Front Genet* 2019;10:267.
PUBMED | CROSSREF

47. O'Sullivan JW, Raghavan S, Marquez-Luna C, Luzum JA, Damrauer SM, Ashley EA, et al. Polygenic risk scores for cardiovascular disease: a scientific statement from the American Heart Association. *Circulation* 2022;146(8):e93-118.
PUBMED | CROSSREF

48. Kuanr M, Mohapatra P, Mittal S, Maindarkar M, Fauda MM, Saba L, et al. Recommender system for the efficient treatment of COVID-19 using a convolutional neural network model and image similarity. *Diagnostics (Basel)* 2022;12(11):2700.
PUBMED | CROSSREF

49. Natarajan P, Young R, Stitziel NO, Padmanabhan S, Baber U, Mehran R, et al. Polygenic risk score identifies subgroup with higher burden of atherosclerosis and greater relative benefit from statin therapy in the primary prevention setting. *Circulation* 2017;135(22):2091-101.
PUBMED | CROSSREF

50. Quazi S. Artificial intelligence and machine learning in precision and genomic medicine. *Med Oncol* 2022;39(8):120.
PUBMED | CROSSREF

51. Inouye M, Abraham G, Nelson CP, Wood AM, Sweeting MJ, Dudbridge F, et al. Genomic risk prediction of coronary artery disease in 480,000 adults: implications for primary prevention. *J Am Coll Cardiol* 2018;72(16):1883-93.
PUBMED | CROSSREF

52. Fritzsche MC, Akyüz K, Cano Abadía M, McLennan S, Marttinen P, Mayrhofer MT, et al. Ethical layering in AI-driven polygenic risk scores: new complexities, new challenges. *Front Genet* 2023;14:1098439.
PUBMED | CROSSREF

53. Aragam KG, Natarajan P. Polygenic scores to assess atherosclerotic cardiovascular disease risk: clinical perspectives and basic implications. *Circ Res* 2020;126(9):1159-77.
PUBMED | CROSSREF

54. Dudbridge F. Power and predictive accuracy of polygenic risk scores. *PLoS Genet* 2013;9(3):e1003348.
PUBMED | CROSSREF

55. Dubey AK, Chabert GL, Carriero A, Pasche A, Danna PS, Agarwal S, et al. Ensemble deep learning derived from transfer learning for classification of COVID-19 patients on hybrid deep-learning-based lung segmentation: a data augmentation and balancing framework. *Diagnostics (Basel)* 2023;13(11):1954.
PUBMED | CROSSREF

56. Suri JS, Agarwal S, Saba L, Chabert GL, Carriero A, Paschè A, et al. Multicenter study on COVID-19 lung computed tomography segmentation with varying glass ground opacities using unseen deep learning artificial intelligence paradigms: COVLIAS 1.0 validation. *J Med Syst* 2022;46(10):62.
PUBMED | CROSSREF

57. Suri JS, Agarwal S, Chabert GL, Carriero A, Paschè A, Danna PS, et al. COVLIAS 2.0-cXAI: cloud-based explainable deep learning system for COVID-19 lesion localization in computed tomography scans. *Diagnostics (Basel)* 2022;12(6):1482.
    PUBMED | **CROSSREF**

58. Das S, Nayak GK, Saba L, Kalra M, Suri JS, Saxena S. An artificial intelligence framework and its bias for brain tumor segmentation: a narrative review. *Comput Biol Med* 2022;143:105273.
    PUBMED | **CROSSREF**

59. Suri JS, Agarwal S, Carriero A, Paschè A, Danna PS, Columbu M, et al. COVLIAS 1.0 vs. MedSeg: artificial intelligence-based comparative study for automated COVID-19 computed tomography lung segmentation in Italian and Croatian cohorts. *Diagnostics (Basel)* 2021;11(12):2367.
    PUBMED | **CROSSREF**

60. Suri JS, Agarwal S, Elavarthi P, Pathak R, Ketireddy V, Columbu M, et al. Inter-variability study of COVLIAS 1.0: hybrid deep learning models for COVID-19 lung segmentation in computed tomography. *Diagnostics (Basel)* 2021;11(11):2025.
    PUBMED | **CROSSREF**

61. Jena B, Saxena S, Nayak GK, Saba L, Sharma N, Suri JS. Artificial intelligence-based hybrid deep learning models for image classification: the first narrative review. *Comput Biol Med* 2021;137:104803.
    PUBMED | **CROSSREF**

62. Jain PK, Sharma N, Giannopoulos AA, Saba L, Nicolaides A, Suri JS. Hybrid deep learning segmentation models for atherosclerotic plaque in internal carotid artery B-mode ultrasound. *Comput Biol Med* 2021;136:104721.
    PUBMED | **CROSSREF**

63. Skandha SS, Nicolaides A, Gupta SK, et al. A hybrid deep learning paradigm for carotid plaque tissue characterization and its validation in multicenter cohorts using a supercomputer framework. *Comput Biol Med* 2022;141:105131.
    PUBMED | **CROSSREF**

64. Zheng H, Wang H, Azuaje F. Incorporation of ontology-driven biological knowledge into cardiovascular genomics. 2011 Computing in Cardiology; September 18-21, 2011; Hangzhou, China. Piscataway, NJ, USA: Institute of Electrical and Electronics Engineers; 2011, 565-8.

65. Wung SF, Hickey KT, Taylor JY, Gallek MJ. Cardiovascular genomics. *J Nurs Scholarsh* 2013;45(1):60-8.
    PUBMED | **CROSSREF**

66. Young WJ, Ramírez J, van Duijvenboden S, et al. Will genetic data significantly change cardiovascular risk prediction in daily practice? 2020 Computing in Cardiology; September 13-16, 2020; Rimini, Italy. Piscataway, NJ, USA: Institute of Electrical and Electronics Engineers; 2020, 1-4.

67. Ganesh SK, Arnett DK, Assimes TL, Basson CT, Chakravarti A, Ellinor PT, et al. Genetics and genomics for the prevention and treatment of cardiovascular disease: update: a scientific statement from the American Heart Association. *Circulation* 2013;128(25):2813-51.
    PUBMED | **CROSSREF**

68. Tandel GS, Tiwari A, Kakde OG, Gupta N, Saba L, Suri JS. Role of ensemble deep learning for brain tumor classification in multiple magnetic resonance imaging sequence Data. *Diagnostics (Basel)* 2023;13(3):481.
    PUBMED | **CROSSREF**

69. Terrada O, Cherradi B, Raihani A, Bouattane O. Classification and prediction of atherosclerosis diseases using machine learning algorithms. In: *2019 5th International Conference on Optimization and Applications (ICOA)*. Piscataway, NJ, USA: Institute of Electrical and Electronics Engineers; 2019, 10056.

70. Suri JS, Bhagawati M, Paul S, Protogeron A, Sfikakis PP, Kitas GD, et al. Understanding the bias in machine learning systems for cardiovascular disease risk assessment: The first of its kind review. *Comput Biol Med* 2022;142:105204.
    PUBMED | **CROSSREF**

71. Jamthikar AD, Gupta D, Mantella LE, et al. Multiclass machine learning vs. conventional calculators for stroke/CVD risk assessment using carotid plaque predictors with coronary angiography scores as gold standard: a 500 participants study. *Int J Cardiovasc Imaging* 2021;37(4):1171-87.
    PUBMED | **CROSSREF**

72. Reel PS, Reel S, Pearson E, Trucco E, Jefferson E. Using machine learning approaches for multi-omics data analysis: a review. *Biotechnol Adv* 2021;49:107739.
    PUBMED | **CROSSREF**

73. Vakili D, Radenkovic D, Chawla S, Bhatt DL. Panomics: new databases for advancing cardiology. *Front Cardiovasc Med* 2021;8:587768.
    PUBMED | **CROSSREF**

74. Picard M, Scott-Boyer MP, Bodein A, Périn O, Droit A. Integration strategies of multi-omics data for machine learning analysis. *Comput Struct Biotechnol J* 2021;19:3735-46.
   PUBMED | **CROSSREF**

75. Pan Y, Lei X, Zhang Y. Association predictions of genomics, proteinomics, transcriptomics, microbiome, metabolomics, pathomics, radiomics, drug, symptoms, environment factor, and disease networks: a comprehensive approach. *Med Res Rev* 2022;42(1):441-61.
   PUBMED | **CROSSREF**

76. Hamamoto R, Komatsu M, Takasawa K, Asada K, Kaneko S. Epigenetics analysis and integrated analysis of multiomics data, including epigenetic data, using artificial intelligence in the era of precision medicine. *Biomolecules* 2019;10(1):62.
   PUBMED | **CROSSREF**

77. Schnabel RB, Baccarelli A, Lin H, Ellinor PT, Benjamin EJ. Next steps in cardiovascular disease genomic research--sequencing, epigenetics, and transcriptomics. *Clin Chem* 2012;58(1):113-26.
   PUBMED | **CROSSREF**

78. Jacinto FV, Link W, Ferreira BI. CRISPR/Cas9-mediated genome editing: From basic research to translational medicine. *J Cell Mol Med* 2020;24(7):3766-78.
   PUBMED | **CROSSREF**

79. Wang Z, Emmerich A, Pillon NJ, Moore T, Hemerich D, Cornelis MC, et al. Genome-wide association analyses of physical activity and sedentary behavior provide insights into underlying mechanisms and roles in disease prevention. *Nat Genet* 2022;54(9):1332-44.
   PUBMED | **CROSSREF**

80. Tahir UA, Katz DH, Avila-Pachecho J, Bick AG, Pampana A, Robbins JM, et al. Whole genome association study of the plasma metabolome identifies metabolites linked to cardiometabolic disease in black individuals. *Nat Commun* 2022;13(1):4923.
   PUBMED | **CROSSREF**

81. Marees AT, de Kluiver H, Stringer S, Vorspan F, Curis E, Marie-Claire C, et al. A tutorial on conducting genome-wide association studies: quality control and statistical analysis. *Int J Methods Psychiatr Res* 2018;27(2):e1608.
   PUBMED | **CROSSREF**

82. Katz DH, Tahir UA, Bick AG, Pampana A, Ngo D, Benson MD, et al. Whole genome sequence analysis of the plasma proteome in black adults provides novel insights into cardiovascular disease. *Circulation* 2022;145(5):357-70.
   PUBMED | **CROSSREF**

83. Rahman MH, Peng S, Hu X, Chen C, Rahman MR, Uddin S, et al. A network-based bioinformatics approach to identify molecular biomarkers for type 2 diabetes that are linked to the progression of neurological diseases. *Int J Environ Res Public Health* 2020;17(3):1035.
   PUBMED | **CROSSREF**

84. Mavaddat N, Michailidou K, Dennis J, Lush M, Fachal L, Lee A, et al. Polygenic risk scores for prediction of breast cancer and breast cancer subtypes. *Am J Hum Genet* 2019;104(1):21-34.
   PUBMED | **CROSSREF**

85. Pjanic M, Miller CL, Wirka R, Kim JB, DiRenzo DM, Quertermous T. Genetics and genomics of coronary artery disease. *Curr Cardiol Rep* 2016;18(10):102.
   PUBMED | **CROSSREF**

86. Musunuru K, Hershberger RE, Day SM, Klinedinst NJ, Landstrom AP, Parikh VN, et al. Genetic testing for inherited cardiovascular diseases: a scientific statement from the American Heart Association. *Circ Genom Precis Med* 2020;13(4):e000067.
   PUBMED | **CROSSREF**

87. Miyazawa K, Ito K. Genetic analysis for coronary artery disease toward diverse populations. *Front Genet* 2021;12:766485.
   PUBMED | **CROSSREF**

88. Kwon OS, Hong M, Kim TH, Hwang I, Shim J, Choi EK, et al. Genome-wide association study-based prediction of atrial fibrillation using artificial intelligence. *Open Heart* 2022;9(1):e001898.
   PUBMED | **CROSSREF**

89. Choi SW, Mak TS, O'Reilly PF. Tutorial: a guide to performing polygenic risk score analyses. *Nat Protoc* 2020;15(9):2759-72.
   PUBMED | **CROSSREF**

90. Chatterjee N, Shi J, García-Closas M. Developing and evaluating polygenic risk prediction models for stratified disease prevention. *Nat Rev Genet* 2016;17(7):392-406.
   PUBMED | **CROSSREF**

91. Collister JA, Liu X, Clifton L. Calculating polygenic risk scores (PRS) in UK Biobank: a practical guide for epidemiologists. *Front Genet* 2022;13:818574.
PUBMED | **CROSSREF**

92. Wand H, Lambert SA, Tamburro C, Iacocca MA, O'Sullivan JW, Sillari C, et al. Improving reporting standards for polygenic scores in risk prediction studies. *Nature* 2021;591(7849):211-9.
PUBMED | **CROSSREF**

93. Martin AR, Kanai M, Kamatani Y, Okada Y, Neale BM, Daly MJ. Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat Genet* 2019;51(4):584-91.
PUBMED | **CROSSREF**

94. Hindy G, Aragam KG, Ng K, Chaffin M, Lotta LA, Baras A, et al. Genome-wide polygenic score, clinical risk factors, and long-term trajectories of coronary artery disease. *Arterioscler Thromb Vasc Biol* 2020;40(11):2738-46.
PUBMED | **CROSSREF**

95. de Marvao A, Dawes TJ, O'Regan DP. Artificial intelligence for cardiac imaging-genetics research. *Front Cardiovasc Med* 2020;6:195.
PUBMED | **CROSSREF**

96. Öztornaci RO, Coşgun E, Çolak C, Taşdelen B. Prediction of Polygenic Risk Score by machine learning and deep learning methods in genome-wide association studies. *bioRxiv*. January 3, 2023. https://doi.org/10.1101/2022.12.30.522280.
CROSSREF

97. Li L, Huang Y, Han Y, Jiang J. Use of deep learning genomics to discriminate Alzheimer's disease and healthy controls. 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC); November 1-5, 2021. Piscataway, NJ, USA: Institute of Electrical and Electronics Engineers; 2021, 5788-91.

98. Bhadri K, Karnik N, Dhatrak P. Current advancements in cardiovascular disease management using artificial intelligence and machine learning models: current scenario and challenges. 2022 10th International Conference on Emerging Trends in Engineering and Technology-Signal and Information Processing (ICETET-SIP-22); April 29-30, 2022; Nagpur, India. Piscataway, NJ, USA: Institute of Electrical and Electronics Engineers; 2022, 1-6.

99. Dai H, Younis A, Kong JD, Puce L, Jabbour G, Yuan H, et al. Big data in cardiology: State-of-art and future prospects. *Front Cardiovasc Med* 2022;9:844296.
PUBMED | **CROSSREF**

100. Dai J, Lv J, Zhu M, Wang Y, Qin N, Ma H, et al. Identification of risk loci and a polygenic risk score for lung cancer: a large-scale prospective cohort study in Chinese populations. *Lancet Respir Med* 2019;7(10):881-91.
PUBMED | **CROSSREF**

101. Zekavat SM, Raghu VK, Trinder M, Ye Y, Koyama S, Honigberg MC, et al. Deep learning of the retina enables phenome-and genome-wide analyses of the microvasculature. *Circulation* 2022;145(2):134-50.
PUBMED | **CROSSREF**

102. Westerlund AM, Hawe JS, Heinig M, Schunkert H. Risk prediction of cardiovascular events by exploration of molecular data with explainable artificial intelligence. *Int J Mol Sci* 2021;22(19):10291.
PUBMED | **CROSSREF**

103. Rukhsar L, Bangyal WH, Ali Khan MS, Ag Ibrahim AA, Nisar K, Rawat DB. Analyzing RNA-seq gene expression data using deep learning approaches for cancer classification. *Applied Sciences* 2022;12(4):1850.
CROSSREF

104. Mathur P, Srivastava S, Xu X, Mehta JL. Artificial intelligence, machine learning, and cardiovascular disease. *Clin Med Insights Cardiol* 2020;14:1179546820927404.
PUBMED | **CROSSREF**

105. Suri JS, Maindarkar MA, Paul S, Ahluwalia P, Bhagawati M, Saba L, et al. Deep learning paradigm for cardiovascular disease/stroke risk stratification in Parkinson's disease affected by COVID-19: a narrative review. *Diagnostics (Basel)* 2022;12(7):1543.
PUBMED | **CROSSREF**

106. Suri JS, Bhagawati M, Paul S, Protogerou AD, Sfikakis PP, Kitas GD, et al. A powerful paradigm for cardiovascular risk stratification using multiclass, multi-label, and ensemble-based machine learning paradigms: a narrative review. *Diagnostics (Basel)* 2022;12(3):722.
PUBMED | **CROSSREF**

107. Weng SF, Reps J, Kai J, Garibaldi JM, Qureshi N. Can machine-learning improve cardiovascular risk prediction using routine clinical data? *PLoS One* 2017;12(4):e0174944.
PUBMED | **CROSSREF**

108. Schiano C, Franzese M, Geraci F, Zanfardino M, Maiello C, Palmieri V, et al. Machine learning and bioinformatics framework integration to potential familial DCM-related markers discovery. *Genes (Basel)* 2021;12(12):1946.
    PUBMED | CROSSREF

109. Saba L, Tiwari A, Biswas M, Gupta SK, Godia-Cuadrado E, Chaturvedi A, et al. Wilson's disease: a new perspective review on its genetics, diagnosis and treatment. *Front Biosci (Elite Ed)* 2019;11(1):166-85.
    PUBMED | CROSSREF

110. Liu B, Fang L, Xiong Y, Du Q, Xiang Y, Chen X, et al. A machine learning model based on genetic and traditional cardiovascular risk factors to predict premature coronary artery disease. *Front Biosci (Landmark Ed)* 2022;27(7):211.
    PUBMED | CROSSREF

111. Ordikhani M, Saniee Abadeh M, Prugger C, Hassannejad R, Mohammadifard N, Sarrafzadegan N. An evolutionary machine learning algorithm for cardiovascular disease risk prediction. *PLoS One* 2022;17(7):e0271723.
    PUBMED | CROSSREF

112. Kannel WB, McGee D, Gordon T. A general cardiovascular risk profile: the Framingham Study. *Am J Cardiol* 1976;38(1):46-51.
    PUBMED | CROSSREF

113. Assmann G, Schulte H. The Prospective Cardiovascular Münster (PROCAM) study: prevalence of hyperlipidemia in persons with hypertension and/or diabetes mellitus and the relationship to coronary heart disease. *Am Heart J* 1988;116(6 Pt 2):1713-24.
    PUBMED | CROSSREF

114. 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(2):172-6.
    PUBMED | CROSSREF

115. Lakoski SG, Greenland P, Wong ND, Schreiner PJ, Herrington DM, Kronmal RA, et al. Coronary artery calcium scores and risk for cardiovascular events in women classified as "low risk" based on Framingham risk score: the multi-ethnic study of atherosclerosis (MESA). *Arch Intern Med* 2007;167(22):2437-42.
    PUBMED | CROSSREF

116. Ridker PM, Buring JE, Rifai N, Cook NR. Development and validation of improved algorithms for the assessment of global cardiovascular risk in women: the Reynolds Risk Score. *JAMA* 2007;297(6):611-9.
    PUBMED | CROSSREF

117. Hippisley-Cox J, Coupland C, Vinogradova Y, Robson J, Minhas R, Sheikh A, et al. Predicting cardiovascular risk in England and Wales: prospective derivation and validation of QRISK2. *BMJ* 2008;336(7659):1475-82.
    PUBMED | CROSSREF

118. 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(7611):136.
    PUBMED | CROSSREF

119. Goff DC Jr, Lloyd-Jones DM, Bennett G, Coady S, D'Agostino RB, Gibbons R, et al. 2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *Circulation* 2014;129(25 Suppl 2):S49-73.
    PUBMED | CROSSREF

120. Conroy RM, Pyörälä K, Fitzgerald AP, Sans S, Menotti A, De Backer G, et al. Estimation of ten-year risk of fatal cardiovascular disease in Europe: the SCORE project. *Eur Heart J* 2003;24(11):987-1003.
    PUBMED | CROSSREF

121. Khandelwal M, Kumar Rout R, Umer S, Mallik S, Li A. Multifactorial feature extraction and site prognosis model for protein methylation data. *Brief Funct Genomics* 2023;22(1):20-30.
    PUBMED | CROSSREF

122. Pasha SN, Ramesh D, Mohmmad S, Harshavardhan A. Cardiovascular disease prediction using deep learning techniques. *IOP Conf Ser Mater Sci Eng* 2020;981(2):022006.
    CROSSREF

123. Banchhor SK, Londhe ND, Araki T, Saba L, Radeva P, Laird JR, et al. Wall-based measurement features provides an improved IVUS coronary artery risk assessment when fused with plaque texture-based features during machine learning paradigm. *Comput Biol Med* 2017;91:198-212.
    PUBMED | CROSSREF

124. Araki T, Ikeda N, Shukla D, Jain PK, Londhe ND, Shrivastava VK, et al. PCA-based polling strategy in machine learning framework for coronary artery disease risk assessment in intravascular ultrasound: a link between carotid and coronary grayscale plaque morphology. *Comput Methods Programs Biomed* 2016;128:137-58.
    PUBMED | CROSSREF

125. Khalifa NE, Taha MH, Ali DE, Slowik A, Hassanien AE. Artificial intelligence technique for gene expression by tumor RNA-Seq data: a novel optimized deep learning approach. *IEEE Access* 2020;8:22874-83.
CROSSREF

126. Peng J, Xue H, Wei Z, Tuncali I, Hao J, Shang X. Integrating multi-network topology for gene function prediction using deep neural networks. *Brief Bioinform* 2021;22(2):2096-105.
PUBMED | CROSSREF

127. Jamthikar A, Gupta D, Khanna NN, Saba L, Araki T, Viskovic K, et al. A low-cost machine learning-based cardiovascular/stroke risk assessment system: integration of conventional factors with image phenotypes. *Cardiovasc Diagn Ther* 2019;9(5):420-30.
PUBMED | CROSSREF

128. Shrivastava VK, Londhe ND, Sonawane RS, Suri JS. A novel and robust Bayesian approach for segmentation of psoriasis lesions and its risk stratification. *Comput Methods Programs Biomed* 2017;150:9-22.
PUBMED | CROSSREF

129. Araki T, Ikeda N, Dey N, Chakraborty S, Saba L, Kumar D, et al. A comparative approach of four different image registration techniques for quantitative assessment of coronary artery calcium lesions using intravascular ultrasound. *Comput Methods Programs Biomed* 2015;118(2):158-72.
PUBMED | CROSSREF

130. Shrivastava VK, Londhe ND, Sonawane RS, Suri JS. Reliable and accurate psoriasis disease classification in dermatology images using comprehensive feature space in machine learning paradigm. *Expert Syst Appl* 2015;42(15-16):6184-95.
CROSSREF

131. Agarwal M, Agarwal S, Saba L, Chabert GL, Gupta S, Carriero A, et al. Eight pruning deep learning models for low storage and high-speed COVID-19 computed tomography lung segmentation and heatmap-based lesion localization: a multicenter study using COVLIAS 2.0. *Comput Biol Med* 2022;146:105571.
PUBMED | CROSSREF

132. Song S, Jiang W, Hou L, Zhao H. Leveraging effect size distributions to improve polygenic risk scores derived from summary statistics of genome-wide association studies. *PLOS Comput Biol* 2020;16(2):e1007565.
PUBMED | CROSSREF

133. Xu Y, Wang Y, Xie X, Wang F, Chen Q, Sun H. An autoencoder-based matrix factorization approach to estimating cell proportion from bulk tumor RNA-seq data. 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM); December 9-12, 2021. Piscataway, NJ, USA: Institute of Electrical and Electronics Engineers; 2021, 562-7.

134. Zeng M, Lu C, Fei Z, Wu FX, Li Y, Wang J, et al. DMFLDA: a deep learning framework for predicting lncRNA–disease associations. *IEEE/ACM Trans Comput Biol Bioinformatics* 2021;18(6):2353-63.
PUBMED | CROSSREF

135. Zhao M, Tang Y, Kim H, Hasegawa K. Machine learning with k-means dimensional reduction for predicting survival outcomes in patients with breast cancer. *Cancer Inform* 2018;17:1176935118810215.
PUBMED | CROSSREF

136. Furey TS, Cristianini N, Duffy N, Bednarski DW, Schummer M, Haussler D. Support vector machine classification and validation of cancer tissue samples using microarray expression data. *Bioinformatics* 2000;16(10):906-14.
PUBMED | CROSSREF

137. Gu Y, Zheng S, Yin Q, Jiang R, Li J. REDDA: Integrating multiple biological relations to heterogeneous graph neural network for drug-disease association prediction. *Comput Biol Med* 2022;150:106127.
PUBMED | CROSSREF

138. Vilhjálmsson BJ, Yang J, Finucane HK, Gusev A, Lindström S, Ripke S, et al. Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *Am J Hum Genet* 2015;97(4):576-92.
PUBMED | CROSSREF

139. Privé F, Aschard H, Blum MG. Efficient implementation of penalized regression for genetic risk prediction. *Genetics* 2019;212(1):65-74.
PUBMED | CROSSREF

140. Leonenko G, Sims R, Shoai M, Frizzati A, Bossù P, Spalletta G, et al. Polygenic risk and hazard scores for Alzheimer's disease prediction. *Ann Clin Transl Neurol* 2019;6(3):456-65.
PUBMED | CROSSREF

141. Choi SW, O'Reilly PF. PRSice-2: Polygenic Risk Score software for biobank-scale data. *Gigascience* 2019;8(7):giz082.
PUBMED | CROSSREF

142. Ge T, Chen CY, Ni Y, Feng YA, Smoller JW. Polygenic prediction via Bayesian regression and continuous shrinkage priors. *Nat Commun* 2019;10(1):1776.
PUBMED | CROSSREF

143. Song L, Horvath S. Predicting COPD status with a random generalized linear model. *Syst Biomed (Austin)* 2013;1(4):261-7.
CROSSREF

144. Khera AV, Chaffin M, Zekavat SM, Collins RL, Roselli C, Natarajan P, et al. Whole-genome sequencing to characterize monogenic and polygenic contributions in patients hospitalized with early-onset myocardial infarction. *Circulation* 2019;139(13):1593-602.
PUBMED | CROSSREF

145. Lloyd-Jones LR, Zeng J, Sidorenko J, Yengo L, Moser G, Kemper KE, et al. Improved polygenic prediction by Bayesian multiple regression on summary statistics. *Nat Commun* 2019;10(1):5086.
PUBMED | CROSSREF

146. Mahajan A, Taliun D, Thurner M, Robertson NR, Torres JM, Rayner NW, et al. Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat Genet* 2018;50(11):1505-13.
PUBMED | CROSSREF

147. Munjral S, Maindarkar M, Ahluwalia P, Puvvula A, Jamthikar A, Jujaray T, et al. Cardiovascular risk stratification in diabetic retinopathy via atherosclerotic pathway in COVID-19/non-COVID-19 frameworks using artificial intelligence paradigm: a narrative review. *Diagnostics (Basel)* 2022;12(5):1234.
PUBMED | CROSSREF

148. Tan CH, Bonham LW, Fan CC, Mormino EC, Sugrue LP, Broce IJ, et al. Polygenic hazard score, amyloid deposition and Alzheimer's neurodegeneration. *Brain* 2019;142(2):460-70.
PUBMED | CROSSREF

149. Mamoshina P, Vieira A, Putin E, Zhavoronkov A. Applications of deep learning in biomedicine. *Mol Pharm* 2016;13(5):1445-54.
PUBMED | CROSSREF

150. Kavitha MS, Gangadaran P, Jackson A, Venmathi Maran BA, Kurita T, Ahn BC. Deep neural network models for colon cancer screening. *Cancers (Basel)* 2022;14(15):3707.
PUBMED | CROSSREF

151. Wang D, Liu S, Warrell J, Won H, Shi X, Navarro FC, et al. Comprehensive functional genomic resource and integrative model for the human brain. *Science* 2018;362(6420):eaat8464.
PUBMED | CROSSREF

152. Vlachopoulos C, Aznaouridis K, Ioakeimidis N, Rokkas K, Vasiliadou C, Alexopoulos N, et al. Unfavourable endothelial and inflammatory state in erectile dysfunction patients with or without coronary artery disease. *Eur Heart J* 2006;27(22):2640-8.
PUBMED | CROSSREF

153. Gandaglia G, Briganti A, Jackson G, Kloner RA, Montorsi F, Montorsi P, et al. A systematic review of the association between erectile dysfunction and cardiovascular disease. *Eur Urol* 2014;65(5):968-78.
PUBMED | CROSSREF

154. Suri JS, Agarwal S, Gupta S, Puvvula A, Viskovic K, Suri N, et al. Systematic review of artificial intelligence in acute respiratory distress syndrome for COVID-19 lung patients: a biomedical imaging perspective. *IEEE J Biomed Health Inform* 2021;25(11):4128-39.
PUBMED | CROSSREF

155. Paul S, Maindarkar M, Saxena S, Saba L, Turk M, Kalra M, et al. Bias investigation in artificial intelligence systems for early detection of Parkinson's disease: a narrative review. *Diagnostics (Basel)* 2022;12(1):166.
PUBMED | CROSSREF

156. Suri JS, Agarwal S, Jena B, Saxena S, El-Baz A, Agarwal V, et al. Five strategies for bias estimation in artificial intelligence-based hybrid deep learning for acute respiratory distress syndrome COVID-19 lung infected patients using AP(ai)Bias 2.0: a systematic review. *IEEE Trans Instrum Meas.* Forthcoming 2022.

157. Kariuki JK, Stuart-Shor EM, Leveille SG, Hayman LL. Evaluation of the performance of existing non-laboratory based cardiovascular risk assessment algorithms. *BMC Cardiovasc Disord* 2013;13:123.
PUBMED | CROSSREF

158. Slack D, Hilgard S, Jia E, Singh S, Lakkaraju H. Fooling LIME and SHAP: adversarial attacks on post hoc explanation methods. In: *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*. New York, NY, USA: Association for Computing Machinery; 2020, 180-6.

159. Biswas M, Kuppili V, Saba L, Edla DR, Suri HS, Cuadrado-Godia E, et al. State-of-the-art review on deep learning in medical imaging. *Front Biosci (Landmark Ed)* 2019;24(3):392-426.
PUBMED | CROSSREF

160. Jena B, Saxena S, Nayak GK, Balestrieri A, Gupta N, Khanna NN, et al. Brain tumor characterization using radiogenomics in artificial intelligence framework. *Cancers (Basel)* 2022;14(16):4052.
PUBMED | CROSSREF

161. Sanagala SS, Nicolaides A, Gupta SK, Koppula VK, Saba L, Agarwal S, et al. Ten fast transfer learning models for carotid ultrasound plaque tissue characterization in augmentation framework embedded with heatmaps for stroke risk stratification. *Diagnostics (Basel)* 2021;11(11):2109.
PUBMED | CROSSREF

162. Khanna NN, Maindarkar MA, Viswanathan V, Fernandes JF, Paul S, Bhagawati M, et al. Economics of artificial intelligence in healthcare: diagnosis vs. treatment. *Healthcare (Basel)* 2022;10(12):2493.
PUBMED | CROSSREF

163. Pennisi M, Kavasidis I, Spampinato C, Schinina V, Palazzo S, Salanitri FP, et al. An explainable AI system for automated COVID-19 assessment and lesion categorization from CT-scans. *Artif Intell Med* 2021;118:102114.
PUBMED | CROSSREF

164. Langlotz CP, Allen B, Erickson BJ, Kalpathy-Cramer J, Bigelow K, Cook TS, et al. A roadmap for foundational research on artificial intelligence in medical imaging: from the 2018 NIH/RSNA/ACR/The Academy Workshop. *Radiology* 2019;291(3):781-91.
PUBMED | CROSSREF

165. Collin CB, Gebhardt T, Golebiewski M, Karaderi T, Hillemanns M, Khan FM, et al. Computational models for clinical applications in personalized medicine-guidelines and recommendations for data integration and model validation. *J Pers Med* 2022;12(2):166.
PUBMED | CROSSREF

166. Khanna NN, Maindarkar MA, Viswanathan V, Puvvula A, Paul S, Bhagawati M, et al. Cardiovascular/stroke risk stratification in diabetic foot infection patients using deep learning-based artificial intelligence: an investigative study. *J Clin Med* 2022;11(22):6844.
PUBMED | CROSSREF

167. Haque AK, Arifuzzaman BM, Siddik SA, Kalam A, Shahjahan TS, Saleena TS, et al. Semantic web in healthcare: a systematic literature review of application, research gap, and future research avenues. *Int J Clin Pract* 2022;2022:6807484.
PUBMED | CROSSREF

168. Panwar A, Semwal G, Goel S, Gupta S. Stratification of the lesions in color fundus images of diabetic retinopathy patients using deep learning models and machine learning classifiers. In: Patgiri R, Bandyopadhyay S, Borah MD, Emilia Balas V, editors. *Edge Analytics. Lecture Notes in Electrical Engineering*. Singapore: Springer; 2022, 653-66.

169. Garg I, Panda P, Roy K. A low effort approach to structured CNN design using PCA. *IEEE Access* 2019;8:1347-60.
CROSSREF

170. Acharya UR, Mookiah MR, Vinitha Sree S, Yanti R, Martis RJ, Saba L, et al. Evolutionary algorithm-based classifier parameter tuning for automatic ovarian cancer tissue characterization and classification. *Ultraschall Med* 2014;35(3):237-45.
PUBMED

171. Xuan J, Jiang H, Hu Y, Ren Z, Zou W, Luo Z, et al. Towards effective bug triage with software data reduction techniques. *IEEE Trans Knowl Data Eng* 2014;27(1):264-80.
CROSSREF

172. Shui L, Ren H, Yang X, Li J, Chen Z, Yi C, et al. The era of radiogenomics in precision medicine: an emerging approach to support diagnosis, treatment decisions, and prognostication in oncology. *Front Oncol* 2021;10:570465.
PUBMED | CROSSREF

173. Panayides AS, Pattichis MS, Leandrou S, Pitris C, Constantinidou A, Pattichis CS. Radiogenomics for precision medicine with a big data analytics perspective. *IEEE J Biomed Health Inform* 2019;23(5):2063-79.
PUBMED | CROSSREF

174. Liu Z, Keller PJ. Emerging imaging and genomic tools for developmental systems biology. *Dev Cell* 2016;36(6):597-610.
PUBMED | CROSSREF

175. Abdel Razek AA, Alksas A, Shehata M, AbdelKhalek A, Abdel Baky K, El-Baz A, et al. Clinical applications of artificial intelligence and radiomics in neuro-oncology imaging. *Insights Imaging* 2021;12(1):152.
PUBMED | CROSSREF

176. Rudie JD, Rauschecker AM, Bryan RN, Davatzikos C, Mohan S. Emerging applications of artificial intelligence in neuro-oncology. *Radiology* 2019;290(3):607-18.
PUBMED | CROSSREF

177. Gu X, Yu X, Shi G, Li Y, Yang L. Can PD-L1 expression be predicted by contrast-enhanced CT in patients with gastric adenocarcinoma? A preliminary retrospective study. *Abdom Radiol (NY)* 2023;48(1):220-8.
PUBMED | CROSSREF

178. Srivastava SK, Singh SK, Suri JS. Effect of incremental feature enrichment on healthcare text classification system: a machine learning paradigm. *Comput Methods Programs Biomed* 2019;172:35-51.
PUBMED | CROSSREF

179. Banchhor SK, Londhe ND, Araki T, Saba L, Radeva P, Laird JR, et al. Well-balanced system for coronary calcium detection and volume measurement in a low resolution intravascular ultrasound videos. *Comput Biol Med* 2017;84:168-81.
PUBMED | CROSSREF

180. Khalil RA, Saeed N, Masood M, Fard YM, Alouini MS, Al-Naffouri TY. 2 Deep learning in the industrial internet of things: Potentials, challenges, and emerging applications. *IEEE Internet Things J* 2021;8(14):11016-40.
CROSSREF

181. Shrivastava VK, Londhe ND, Sonawane RS, Suri JS. Computer-aided diagnosis of psoriasis skin images with HOS, texture and color features: a first comparative study of its kind. *Comput Methods Programs Biomed* 2016;126:98-109.
PUBMED | CROSSREF

182. Karniadakis GE, Kevrekidis IG, Lu L, Perdikaris P, Wang S, Yang L. Physics-informed machine learning. *Nat Rev Phys* 2021;3(6):422-40.
CROSSREF

183. Biswas M, Kuppili V, Edla DR, Suri HS, Saba L, Marinhoe RT, et al. Symtosis: a liver ultrasound tissue characterization and risk stratification in optimized deep learning paradigm. *Comput Methods Programs Biomed* 2018;155:165-77.
PUBMED | CROSSREF

184. Roslan RB, Razly IN, Sabri N, Ibrahim Z. Evaluation of psoriasis skin disease classification using convolutional neural network. *IAES Int J Artif Intell* 2020;9(2):349.

185. Jamthikar AD, Gupta D, Saba L, Khanna NN, Viskovic K, Mavrogeni S, et al. Artificial intelligence framework for predictive cardiovascular and stroke risk assessment models: a narrative review of integrated approaches using carotid ultrasound. *Comput Biol Med* 2020;126:104043.
PUBMED | CROSSREF

186. Jamthikar AD, Gupta D, Johri AM, Mantella LE, Saba L, Kolluri R, et al. Low-cost office-based cardiovascular risk stratification using machine learning and focused carotid ultrasound in an Asian-Indian cohort. *J Med Syst* 2020;44(12):208.
PUBMED | CROSSREF

187. Jamthikar A, Gupta D, Saba L, Khanna NN, Araki T, Viskovic K, et al. Cardiovascular/stroke risk predictive calculators: a comparison between statistical and machine learning models. *Cardiovasc Diagn Ther* 2020;10(4):919-38.
PUBMED | CROSSREF

188. Bartels S, Franco AR, Rundek T. Carotid intima-media thickness (cIMT) and plaque from risk assessment and clinical use to genetic discoveries. *Perspect Med* 2012;1(1-12):139-45.
CROSSREF

189. Suri JS, Puvvula A, Biswas M, Majhail M, Saba L, Faa G, et al. COVID-19 pathways for brain and heart injury in comorbidity patients: a role of medical imaging and artificial intelligence-based COVID severity classification: a review. *Comput Biol Med* 2020;124:103960.
PUBMED | CROSSREF

190. Liu K, Suri JS. *Automatic Vessel Indentification for Angiographic Screening. Patent No.: US6845260B2*. Alexandria, VA, USA: U.S. Patent and Trademark Office; 2005.

191. El-Baz A, Gimel'farb G, Suri JS. *Stochastic Modeling for Medical Image Analysis*. Boca Raton, FL, USA: CRC Press; 2015.

192. Gupta N, Gupta SK, Pathak RK, Jain V, Rashidi P, Suri JS. Human activity recognition in artificial intelligence framework: a narrative review. *Artif Intell Rev* 2022;55(6):4755-808.
PUBMED | CROSSREF

193. Upton R, Mumith A, Beqiri A, Parker A, Hawkes W, Gao S, et al. Automated echocardiographic detection of severe coronary artery disease using artificial intelligence. *JACC Cardiovasc Imaging* 2022;15(5):715-27.
PUBMED | CROSSREF

194. Fu Y, Xu J, Tang Z, Wang L, Yin D, Fan Y, et al. A gene prioritization method based on a swine multi-omics knowledgebase and a deep learning model. *Commun Biol* 2020;3(1):502.
PUBMED | CROSSREF