Check for updates

Establishment and analysis of artificial neural network diagnosis model for coagulation-related molecular subgroups in coronary artery disease

Biwei Zheng[1]*[†], Yujing Li[2,3†] and Guoliang Xiong[2]

[1]Department of Cardiology, Dongguan Hospital of Integrated Chinese and Western Medicine Affiliated to Guangzhou University of Traditional Chinese Medicine, Dongguan, China, [2]Shenzhen Traditional Chinese Medicine Hospital, The Fourth Clinical Medical College of Guangzhou University of Chinese Medicine, Shenzhen, China, [3]Beijing University of Chinese Medicine Shenzhen Hospital (Longgang), Shenzhen, China

Background: Coronary artery disease (CAD) is the most common type of cardiovascular disease and cause significant morbidity and mortality. Abnormal coagulation cascade is one of the high-risk factors in CAD patients, but the molecular mechanism of coagulation in CAD is still limited.

Methods: We clustered and categorized 352 CAD paitents based on the expression patterns of coagulation-related genes (CRGs), and then we explored the molecular and immunological variations across the subgroups to reveal the underlying biological characteristics of CAD patients. The feature genes between CRG-subgroups were further identified using a random forest model (RF) and least absolute shrinkage and selection operator (LASSO) regression, and an artificial neural network prediction model was constructed.

Results: CAD patients could be divided into the C1 and C2 CRG-subgroups, with the C1 subgroup highly enriched in immune-related signaling pathways. The differential expressed genes between the two CRG-subgroups (DE-CRGs) were primarily enriched in signaling pathways connected to signal transduction and energy metabolism. Subsequently, 10 feature DE-CRGs were identified by RF and LASSO. We constructed a novel artificial neural network model using these 10 genes and evaluated and validated its diagnostic performance on a public dataset.

Conclusion: Diverse molecular subgroups of CAD patients may each have a unique gene expression pattern. We may identify subgroups using a few feature genes, providing a theoretical basis for the precise treatment of CAD patients with different molecular subgroups.

Abbreviations: CAD, Coronary artery disease; CRGs, Coagulation-related genes; DE-CRGs, Differential expressed genes between the two CRG-subgroups; GEO, Gene Expression Omnibus; GO, Gene Ontology; GSVA, Gene set variation analysis; KEGG, Kyoto Encyclopedia of Genes and Genomes; LASSO, Least absolute shrinkage and selection operator; PCA, Principal component analysis; RF, Random forest; ssGSEA, single-sample gene set enrichment analysis.

Background

Coronary artery disease (CAD) is a prevalent cardiac illness characterized by the narrowing or blockage of coronary arteries, which are the major vessels supplying blood to the heart. This restriction impedes the delivery of sufficient blood, oxygen, and nutrients to the heart muscle, leading to the accumulation of cholesterol deposits (plaque) and inflammation within the arterial walls (Libby et al., 2021). As a leading cause of death worldwide, CAD poses significant health risks and requires prompt intervention and management to mitigate its adverse effects. In 2019, CAD affected an estimated 197 million patients worldwide, resulting in 9.1 million deaths (16.1% of all deaths) (GBD, 2019 Demographics Collaborators, 2020; Roth et al., 2020). As with the majority of complicated disorders, A person's risk of suffering CAD is influenced by the interplay of inherited and lifestyle variables (Khera and Kathiresan, 2017). The latest epidemiological studies have shown that risk factors for the development of CAD include smoking, hypertension, dyslipidemia, and lack of physical activity, while the prevalence of CAD is increasing in elderly, diabetic, and obese populations (Duggan et al., 2022).

Recent research has shed light on the significant roles of coagulation Factors II (prothrombin), V, VII, and X in CAD. Dysregulation levels of these factors are associated with an increased risk of CAD and adverse cardiovascular events. High neutrophil and basophil blood cell counts, linked to enhanced factor II plasma coagulation activity, may predict mortality in clinically stable CAD patients, indicating underlying prothrombotic mechanisms (Pizzolo et al., 2021). Additionally, the Factor V Leiden mutation poses a risk for premature coronary artery disease, while elevated levels of the coagulation factor VIIa-antithrombin complex are associated with an increased risk of ischemic stroke/systemic thromboembolism (Paszek et al., 2022; Agosti et al., 2023; Valeriani et al., 2023). In related experiments, high-dose statin therapy has shown effectiveness in reducing levels of coagulation factors VII, VIII, and XI, all linked to thrombosis (Stępień et al., 2023). Notably, the reduction in factor XI levels corresponds to a less prothrombotic fibrin clot phenotype, suggesting additional antithrombotic effects in CAD patients (Stępień et al., 2023). Specifically, Factor II promotes thrombus formation, Factor V facilitates fibrin formation, Factor VII initiates the coagulation cascade, and Factor X promotes clot formation, offering potential therapeutic targets for CAD management (Redondo et al., 1999). Furthermore, targeting fibrinogen and factor XI has been demonstrated to decrease the risk of venous thromboembolism and ischemic stroke, supported by Mendelian randomization analysis (Yuan et al., 2021). Additionally, inhibiting factors V, VII, and X may reduce the risk of ischemic stroke (Yuan et al., 2021). These findings underscore promising therapeutic targets for mitigating cardiovascular disease risk associated with the inhibition of clotting factors.

The coagulation system significantly impacts the development of atherothrombotic diseases such as atherosclerosis. Coagulation factors also contribute to plaque instability, inflammatory responses, and thrombotic events within arterial walls, exacerbating atherosclerosis progression and elevating the risk of cardiovascular events like myocardial infarction and stroke (Ajjan and Grant, 2006; Keihanian et al., 2018). Therefore, targeting coagulation system regulation presents a promising strategy for preventing and treating atherothrombotic diseases.

Hence, our study employs artificial neural networks to analyze coagulation-related gene expression patterns in CAD. This innovative approach deepens our comprehension of CAD pathogenesis by revealing intricate molecular signatures and interactions within the coagulation pathway. Through the identification of potential biomarkers and therapeutic targets, our research endeavors to propel personalized treatment strategies for CAD forward.

Methods

Publicly available cohort datasets and preprocessing

The "GEOquery" R package (Davis and Meltzer, 2007) was used to download data, and obtain the expression profiles of chip datasets GSE20681 (Beineke et al., 2012), GSE20680 (Elashoff et al., 2011), and GSE12288 (Sinnaeve et al., 2009). The chip probes corresponding to the platform were taken from the Gene Expression Omnibus (GEO) database. The "org.Hs.eg.db" R package was used to conversion to gene symbols. We combined the GSE20681 and GSE20680 datasets as the training group, and the GSE12288 dataset as the validation group. The "sva" R package was used to remove the batch effect of the two datasets due to differences in time, personnel, and processing methods (Leek et al., 2012). Principal component analysis (PCA) was used to assess the distribution of the two expression matrices.

Coagulation pathways were gathered from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (https://www.genome.jp/kegg/), including hsa04611 (platelet activation) and hsa04610 (complement and coagulation cascades) (Kanehisa and Goto, 2000). There are 209 genes in all determined to be coagulation-related genes (CRGs) in the two pathways.

Consensus clustering analysis of CRG expression patterns

The k-means algorithm was used to cluster CAD samples with the same or similar expression levels of CRGs with 1000-times iteration for classification stability. We used the "ConsensusClusterPlus" R package to implement the algorithm for the optimal k-value (number of clusters) in the training cohort. PCA analysis was performed to reveal differences in the distribution of CRG-subgroups. We also used external datasets for validation.

Pathway characteristics and immune landscape of CRG-subgroups

The heatmap was plotted to display changes in biological functions between CAD subgroups, gene set variation analysis (GSVA) was carried out using the "GSVA" R package (Hänzelmann et al., 2013) to evaluate normalized enrichment

scores (NES) for pathway and functional annotations. The single-sample gene set enrichment analysis (ssGSEA) was used to quantify the degree of infiltration of 23 immune cell signatures in each coronary patient.

Comparison and enrichment analysis of CRG-subgroups

The "limma" R program was used to obtain differentially expressed genes between the C1 and C2 CRG-subgroups (DE-CRGs). The "org.hs.eg.db" R package was applied to annotate gene symbols as Entrez IDs, and the "cluster Profiler" R program was used to perform Gene Ontology (GO) and KEGG pathway analyses on DE-CRGs.

Identification and validation feature CRGs

First, the least absolute shrinkage and selection operator (LASSO) regression was used to filter feature DE-CRGs. The LASSO algorithm's variable selection and shrinkage were performed using the "glmnet" R package (Friedman et al., 2010). For the training cohort, CRG-subgroup (C1/C2) of CAD was the response variable in the regression, while the independent variable in the regression was the normalized expression matrix of potential feature genes (DE-CRGs). The penalty parameter (λ) of the model was determined by ten-fold cross-validation following the minimum criterion. Then, a random forest (RF) model of DE-CRGs was created using the "randomForest" R package (Guidi et al., 2013), and dimension important values were extracted from the RF model using the approach of decreasing accuracy (Gini coefficient method). For further analysis, disease-specific genes with an importance value ("MeanDecreaseGini" index) higher than 2.0 were chosen. Finally, the feature DE-CRGs were obtained by intersecting the particular genes provided by the two approaches.

Construction of CAD classification model by artificial neural network

Artificial neural network is a computing structure proposed based on the mechanism of biological neuron network, which is a kind of simulation, simplification and abstraction of biological neural network. Neurons (feature DE-CRGs) are the "nodes" of this network, the "processing units". We constructed a topological network with layered connections. The neural network with layered structure can be separated into input layer (reception of external input information), hidden layer (exchange and transmission of internal information) and output layer (output of information processing results). Each layer is connected in sequence, and the signal is transmitted in one direction.

We used the training set dataset to establish the neural network disease classification prediction model, and another external data set (validation group) is selected for neural network model validation. The model of feature DE-CRGs was constructed using the "neuralnet" R package (Beck, 2018). Prior to training the neural network, normalization and min-max processing were performed

on the two groups of data. In the neural network model, we set a hidden layer as a model parameter, and constructed a CAD classification model through the obtained gene weight information. In this model, the sum of the product of the weight score and the expression level of important genes is used as the disease classification score. The confusion matrix function was used to do the five-fold cross-validation and acquire the model accuracy results. The AUC classification performance verification results were calculated using the "pROC" software package. The accuracy, recall, precision, and F1 scores were evaluated to assess the validity and reliability of the model.

Statistical analysis

All data analysis in this study was based on R software (version 4.2.1). Pearson and Spearman correlation analysis was used to test the correlation between two variables. Bayesian testing with Benjamini–Hochberg procedure were used for differential analysis to screen the genes with significant differences between the two groups. All tests were two-sided, and $p < 0.05$ was considered statistically significant.

Results

Characteristics of CRG-subgroups with coronary artery disease

The workflow diagram of the study is displayed in Figure 1. We obtained 242 CAD patients (excluding controls) from GSE20680 and GSE20681 datasets as the training group, and 110 CAD patients from GSE12288 as the validation group. We combined the data from the two datasets in order to eliminate the batch effect and get a consistent classification for the training group. Before removing batch effects, samples were clustered across datasets according to the first two principal components (PCs) of unnormalized expression values (Figure 2A). In contrast, the scatterplots of PCA analysis based on normalized expressions showed that the batch effect produced by different platforms was significantly eliminated (Figure 2B). The outcomes demonstrated that batch effect removal via cross-platform normalization is successful.

Identification of CRG-subgroups in coronary artery disease

Two distinct expression patterns, comprising 117 instances in the coagulation-related cluster C1 and 125 cases in cluster C2, were found by employing an unsupervised clustering approach to analyze the expression levels of CRGs from CAD patients in the training group (Figures 3A, B). In accordance with the PCA analysis, all patients could be roughly divided into two parts, which further confirmed two distinct subgroups (Figure 3C). Furthermore, we performed subgroup identification in the validation dataset. Similarly, the validation dataset can also be divided into two different coagulation subgroups (Figures 3D–F).

FIGURE 1
Workflow diagram.
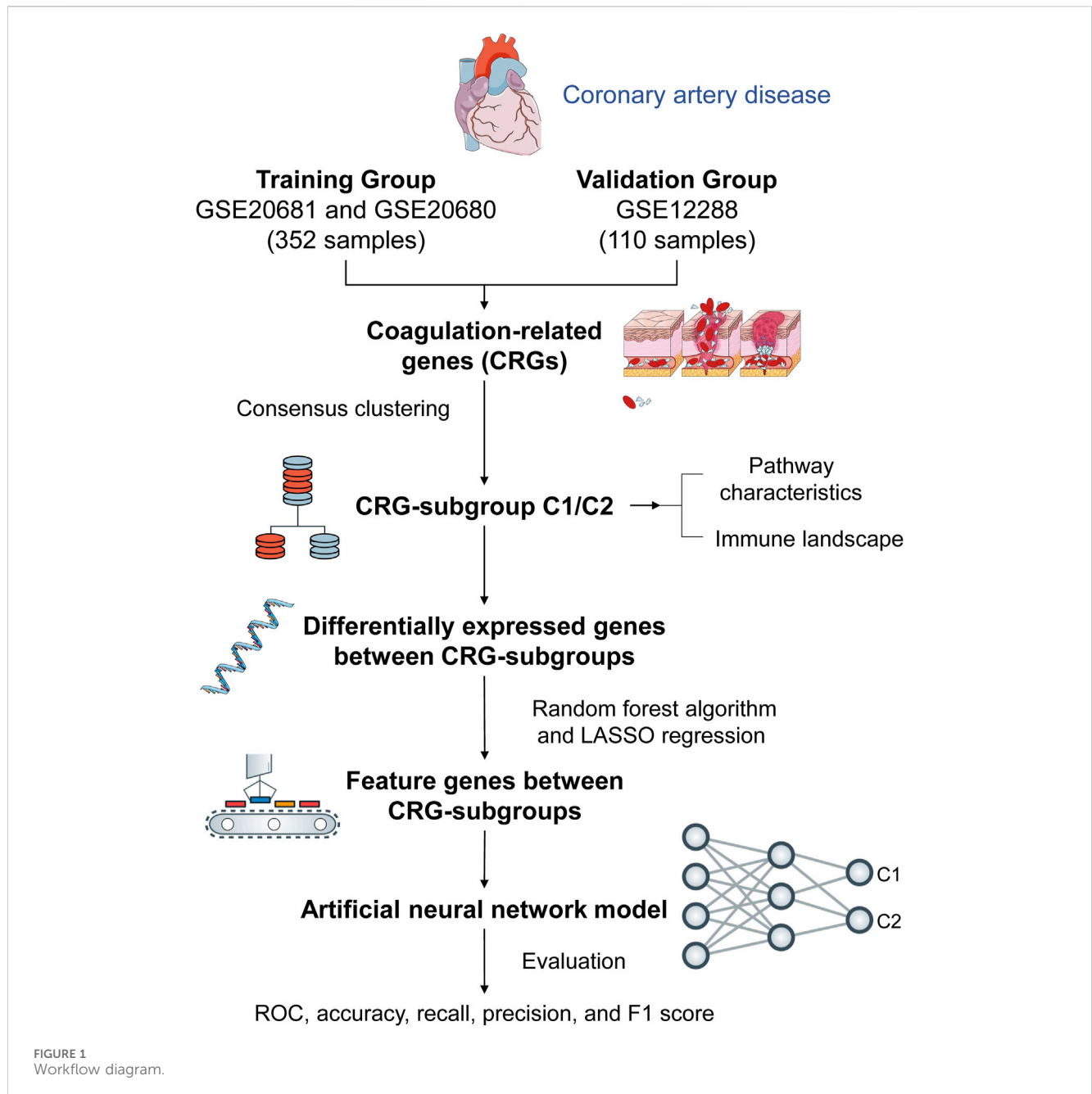
Immune landscape of CRG-subgroups

We used GSVA analysis to compare the regulatory pathway between the two coagulation subgroups in the training dataset, and we discovered that the two subgroups showed clear biological functional differences. The enrichment heatmap revealed that the signaling pathways involved in the metabolism of tyrosine, retinol, linoleic acid, and other biological compounds were significantly enriched in the C1 subgroup (Figure 4A). At the same time, we discovered that the C1 subgroup also had enriched calcium channels and ECM receptor interaction (Figure 4A). In 2022, it was proven that extracellular matrix proteins have a regulatory function on natural killer cells (Bunting et al., 2022). Calcium serves as both a signal and a nutrient in the regulation of numerous

immunological responses linked to B cells and plasma cells (Newman and Tolar, 2021). Taking into account the relationships between CRG-subgroups and the immune system in CAD, we used the ssGSEA method and Wilcoxon test to analyze the abundance of immune cell infiltration of two CRG-subgroups based on the CRGs expression of the training group. The C1 subgroup is characterized by a higher degree of infiltration of natural killer cells and type 17 T helper cells; while the C2 subgroup is characterized by a higher degree of infiltration of immune cells such as activated B cells, activated CD4 T cells, activated CD8 T cells, eosinophil, and immature B cells (Figure 4B). We also calculated immune-related indicators using ssGSEA and found that the expression levels of antigen-presenting cell (APC) co-stimulation, check-point-related

FIGURE 2
Principal component analysis (PCA) of the training group datasets. Visualization samples of the first two principal components before **(A)** and after **(B)** batch-effect removal.

FIGURE 3
Identification of coagulation-related subgroups. **(A, B)** Consensus clustering matrix for k = 2 (optimal cluster number) of the training group. **(C)** PCA analysis of the training group. Cluster analysis **(D, E)** and PCA analysis **(F)** of the validation group.

immune factors, and CCR gene family were higher in the C1 subgroup; the expression levels of HLA gene family, inflammation-promoting and parainflammation-related factors

were higher in C2 subgroup (Figure 4C). This suggests that different coagulation subgroups of CAD have different immune microenvironments.
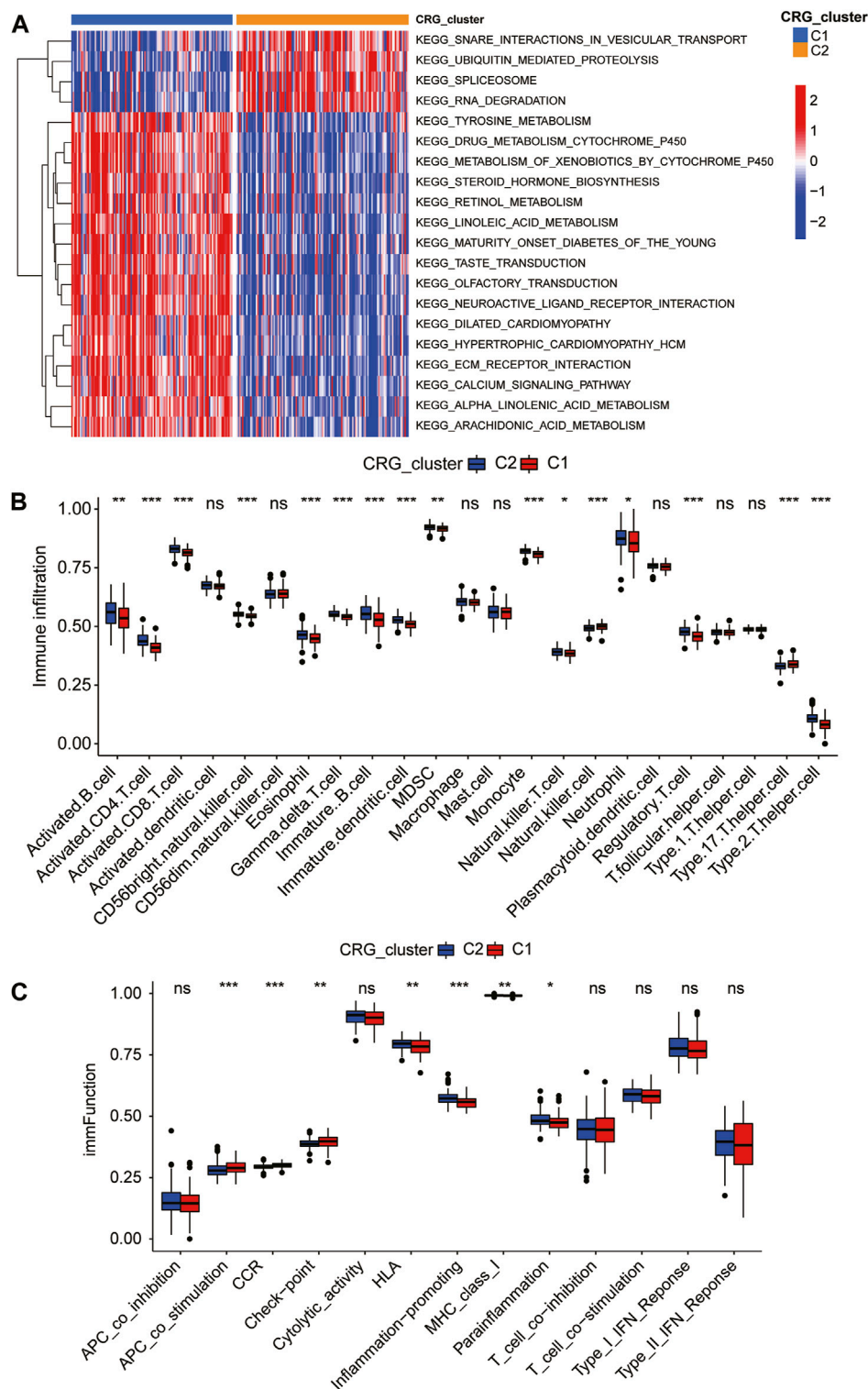
FIGURE 4
Immune landscape of coagulation subgroups. Gene set variation analysis (GSVA) **(A)** and differences in immune cell abundance **(B)** and immune indicators **(C)** between CRG-subgroups.

Functional enrichment analysis of DE-CRGs

DE-CRGs between the C1 and C2 subgroups of the training set dataset were identified by Bayesian testing using the "limma" R package. We screened with $|log2FC|>1$ and adj. $p < 0.05$ as the threshold to identify 95 DE-CRGs related to the coagulation function of CAD, and the DE-CRGs are shown in the heatmap (Figure 5A; Supplementary Table S1).

FIGURE 5
Functional analysis of DE-CRGs between two subgroups. Significance of difference analysis **(A)**, Gene Ontology (GO) enrichment analysis **(B)**, and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analysis **(C)** between C1 and C2 CRG-subgroups.
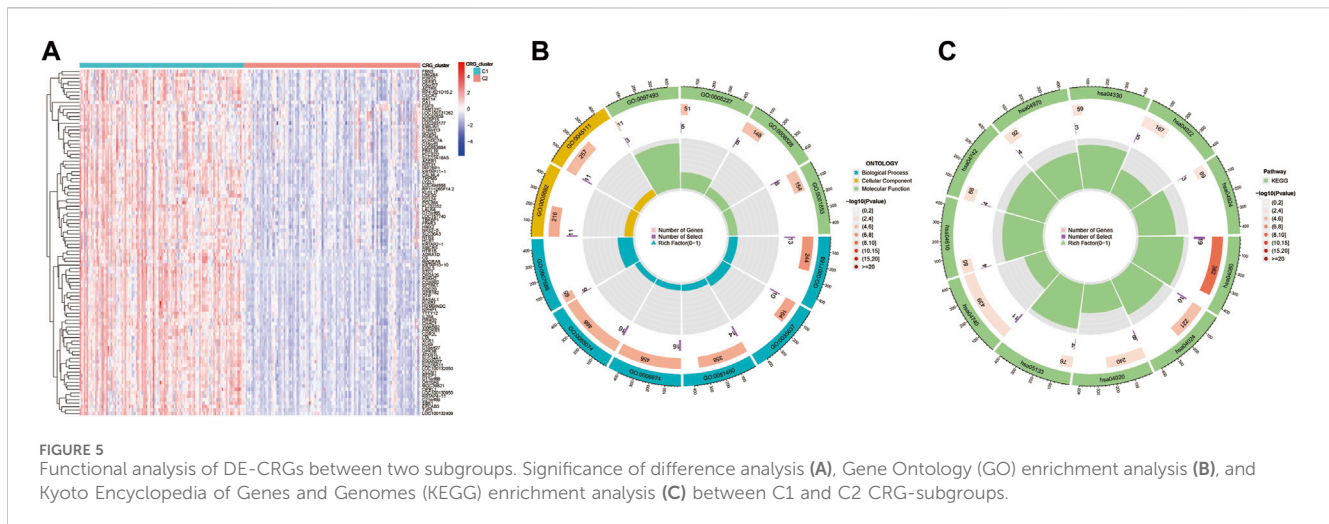
GO enrichment analysis of 95 significant DE-CRGs was implemented using the "clusterProfiler" R package with the thresholds at *p*-values <0.01 and FDR values < 0.01. The outcomes of the GO analysis revealed that these DE-CRGs were mostly enriched in signal transduction-related biological functions such as G protein–coupled peptide receptor activity, serine–type endopeptidase activity, intermediate filament cytoskeleton, calcium ion homeostasis, and transmission of nerve impulse; they were also enriched in regulation of blood pressure, blood vessel diameter maintenance, positive regulation of vasoconstriction related to coagulation and vascular blood pressure regulation (Figure 5B; Supplementary Table S2). According to KEGG analysis, it was found that 95 DE-CRGs were enriched in neuroactive ligand-receptor interaction, cAMP signaling pathway, calcium signaling pathway, notch signaling pathway, complement and coagulation cascades and other signaling pathways (Figure 5C; Supplementary Table S3).

Screening and validation of feature DE-CRGs

First, LASSO regression analysis was performed on 95 DE-CRGs, and the cross-validation method was used for iterative analysis. The results showed that the model's root means square error was lowest when there were 19 variables (Figure 6A). Then, we performed recurrent random forest classification on all possible numbers in 95 variables and calculated the average error rate of the model. Referring to the model error graph, it was found that when the number of classification trees is around 100, the error in the model tends to remain stable (Figure 6B). As the random forest model is being created, the Gini coefficient method was used to reduce the precision and mean square error, and the top 30 characteristic genes of variable importance were output, and the 22 characteristic genes whose "MeanDecreaseGini" index was greater than 2.0 were analyzed (Figure 6C). Finally, we intersected the genes acquired by the two methods to obtain 10 feature DE-CRGs (Figure 6D).
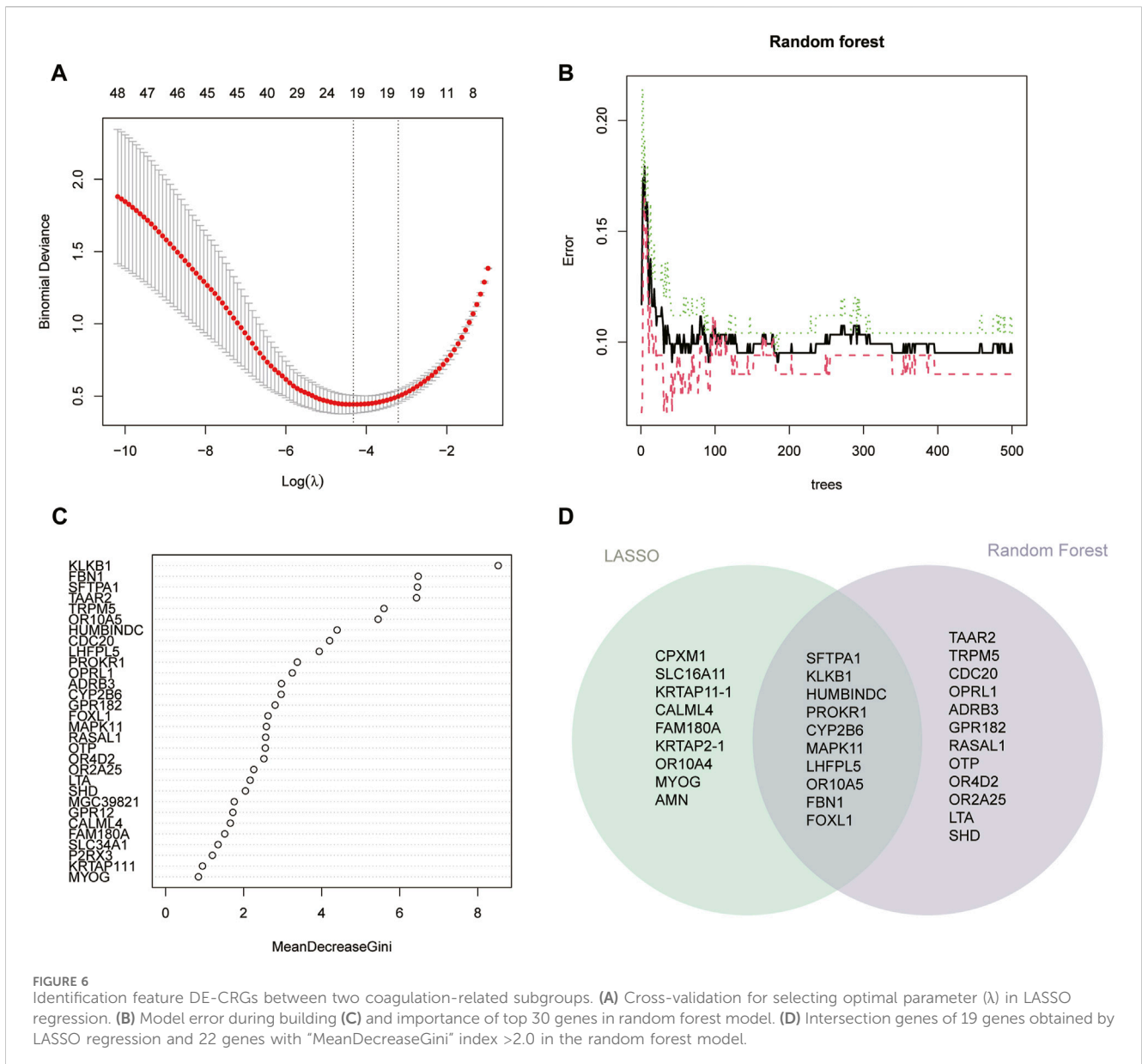
We used unsupervised clustering to compare the expression levels of 10 feature DE-CRGs in CAD patients, and the findings revealed that C1 subgroup patients had high gene expression, whereas C2 subgroup patients had low gene expression. (Figure 7A). Subsequently, we constructed ROC curves for 10 feature DE-CRGs one by one to predict CAD coagulation-related subgroups, and found that the AUC values of 10 genes were all greater than 0.9 (Figures 7B–K). These results show that our feature DE-CRGs have an excellent ability to diagnose and predict molecular subgroups.

Construction of artificial neural network model

We extracted a matrix of 10 feature DE-CRGs expression levels and CAD outcome variables (C1/C2) of 243 samples in the training group to establish a neural network prediction model (Figure 8A). 10 input layers, 5 hidden layers, and 2 output layers are set up for the artificial neural network. The area under the ROC curve (AUC) of the five-fold cross-validation results was 0.999 (Figure 8B). The accuracy, recall, precision, and F1 score of the training group were 0.979, 0.984, 0.976, and 0.980 (Figure 8C). Similarly, the classification effectiveness of the model scoring model created using gene expression and gene weights was confirmed using the validation group, and the AUC value of the ROC curve of the validation group also reached 0.999 (Figure 8D). The accuracy, recall, precision, and F1 score of the validation group were 0.927, 0.951, 0.921, and 0.936 (Figure 8E), which is confirmed that the artificial neural network model we established has excellent predictive robustness for the classification of C1/C2 CRG-subgroups in CAD patients.

We performed Spearman correlation analysis on the gene expression levels of 10 feature DE-CRGs for constructing artificial neural networks and the relative abundance and immune function of immune cells. Among them, 10 genes were positively correlated with the contents of natural killer cells and type-17 T helper cells, and positively regulated the expression level of CCR family, the functions of APC co-stimulation and check-point; while 10 genes are negatively correlated with γδ T cells and type-2 T helper cells, and antagonize the immune function associated with inflammation-promoting and so on. All 10 genes

FIGURE 6
Identification feature DE-CRGs between two coagulation-related subgroups. **(A)** Cross-validation for selecting optimal parameter (λ) in LASSO regression. **(B)** Model error during building **(C)** and importance of top 30 genes in random forest model. **(D)** Intersection genes of 19 genes obtained by LASSO regression and 22 genes with "MeanDecreaseGini" index >2.0 in the random forest model.
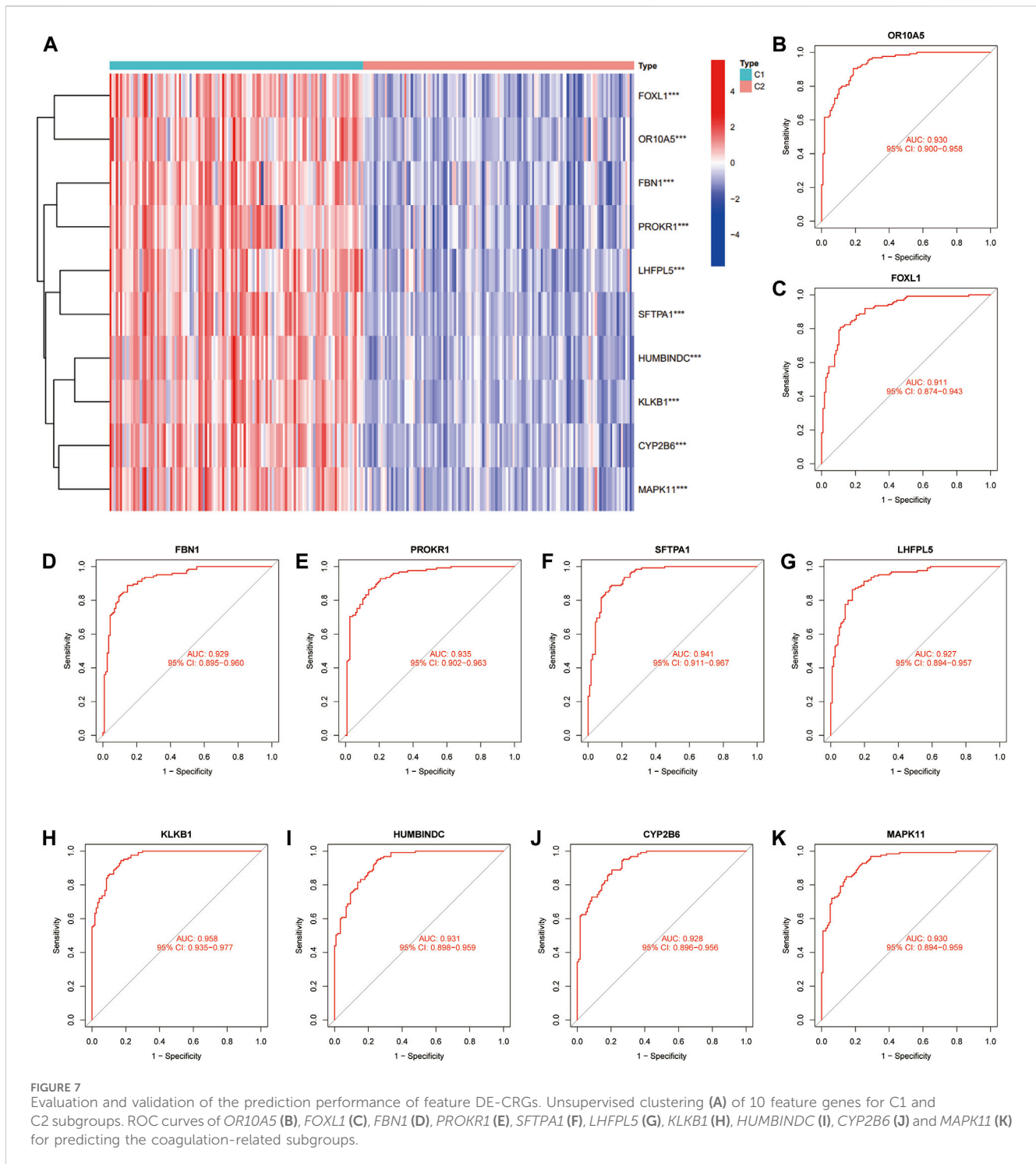
were weakly correlated with immune-related indicators (|cor|<0.5, Figures 9A, B). Moreover, the correlation coefficients between the expression levels of different feature DE-CRGs and the content of a certain immune marker follow the nearly same pattern (a row of similar colors). We speculate that these 10 genes may collaboratively contribute to the immunological program of CAD patients.

Discussion

In recent years, research on molecular subgroups and particular illness biomarkers has been intensively conducted due to the advancement of whole-genome sequencing technology and the demand for individualized treatment. For example, molecular characterization and typing of triple-negative breast cancer (Bareche et al., 2018; Zhao et al., 2020), colorectal cancer (Menter et al., 2019; Hu et al., 2021), lung adenocarcinoma (Wang et al., 2020), pancreatic ductal carcinoma (Topham et al., 2021), and other solid tumors based on tumor multi-omics datasets from the Cancer Genome Atlas (TCGA) database (Tomczak et al., 2015); molecular typing of non-neoplastic diseases are commonly used in immune-related diseases such as HIV (Patil et al., 2020; Amer et al., 2021). However, other diseases are limited by the lack of large-scale sequencing data, and molecular diagnosis and typing are still in the preliminary development. Therefore, we focused on the molecular characteristics of a special physiological activity (coagulation) in CAD patients using gene expression data from individuals in the GEO database. It is hoped that from the molecular level, as a supplement to traditional clinical diagnosis methods, it can improve the prediction accuracy of patient prognosis and diagnosis, and provide clinicians with improved decision-making tools.

In our study, an artificial neural network is introduced, and the model can accurately predict the molecular subgroups of CAD

FIGURE 7
Evaluation and validation of the prediction performance of feature DE-CRGs. Unsupervised clustering **(A)** of 10 feature genes for C1 and C2 subgroups. ROC curves of *OR10A5* **(B)**, *FOXL1* **(C)**, *FBN1* **(D)**, *PROKR1* **(E)**, *SFTPA1* **(F)**, *LHFPL5* **(G)**, *KLKB1* **(H)**, *HUMBINDC* **(I)**, *CYP2B6* **(J)** and *MAPK11* **(K)** for predicting the coagulation-related subgroups.

patients. An artificial neural network is an algorithm based on artificial intelligence and machine learning that consists of a densely linked network of computer processors that were inspired by biological nerve systems (Steimann, 2001). Backpropagation and Bayesian inference techniques are used in data mining and machine learning for artificial intelligence to handle gathered medical data (Salari et al., 2014). Artificial intelligence facilitates the clinical diagnostic and prognosis prediction processes by classifying and organizing medical knowledge and clinical data

(Wong and Monaco, 1995; Jiang et al., 2017). The combined model of machine learning and artificial neural networks utilizing genetic polymorphisms in this study outperforms previous ANN models (Cheng et al., 2022) and other machine learning approaches (Peng et al., 2022) based solely on clinical features when diagnosing CAD patient subgroups. This enhanced diagnostic performance underscores the importance of integrating genetic information into predictive models for CAD. By leveraging genetic polymorphisms alongside ANN technology, this joint model
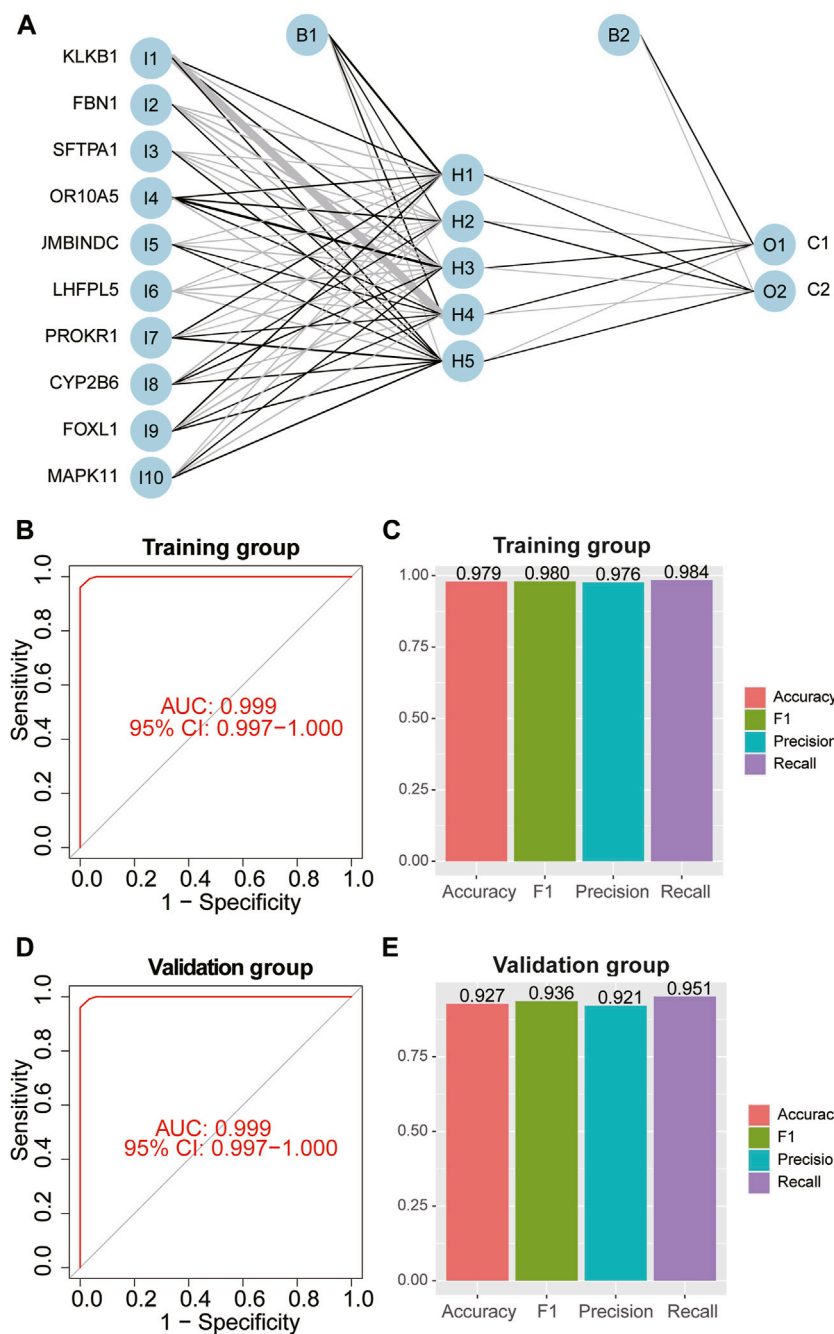
FIGURE 8
Construction and validation of the artificial neural network. Artificial neural network pattern plot **(A)** for predicting coagulation-related C1/C2 subgroups. ROC curves of training **(B)** and validation groups **(D)** for the model. The accuracy, F1-score, precision and recall of training **(C)** and validation groups **(E)** for the model.

offers improved accuracy in identifying CAD patients, thus potentially advancing personalized diagnostic approaches in clinical settings.

However, the limitations of the model are mainly reflected in the fact that the model may have overfitting problems when the training sample size is relatively small; at the same time, the initialization parameters of the neural network have a certain impact on the performance of the model, and the way to set the parameters is a non-deterministic polynomial problem. Therefore, one of the main

directions that need to be explored in the next step is how to set the optimal initialization parameters of the model. This may involve combining regularization techniques with swarm intelligence optimization algorithms and ensemble methods to develop more reliable and generalizable models for genomic analysis. For example, the initial parameters of the model can be further optimized through the swarm intelligence optimization algorithm, such as the whale optimization algorithm (Brodzicki et al., 2021), Harris Hawks optimization algorithm (Qu et al., 2021), and wolf pack
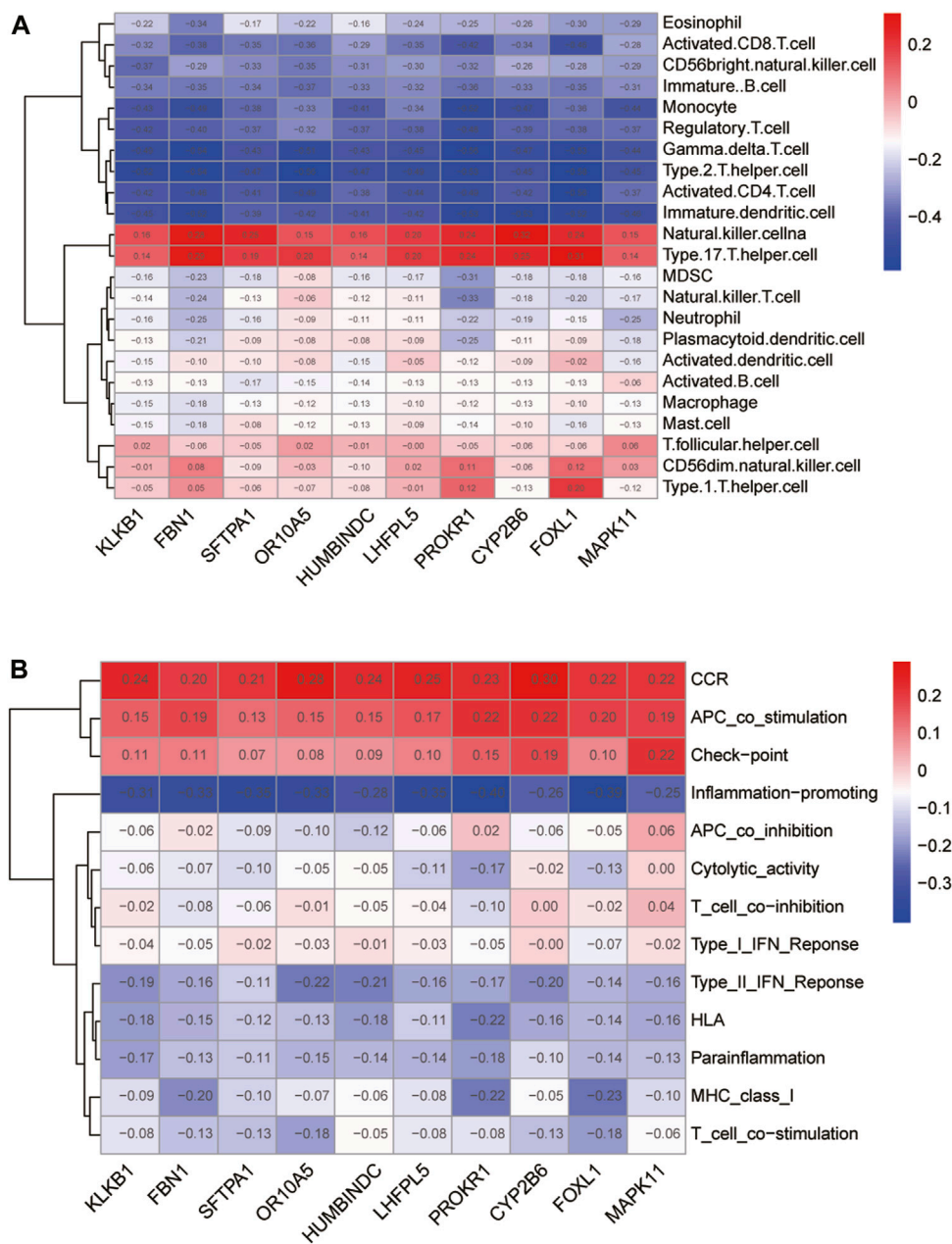
FIGURE 9
Immunological characterization of 10 feature DE-CRGs in artificial neural networks. Correlation of the 10 genes with immune cell abundance **(A)** and immune function **(B)**.

optimization algorithm (Chen et al., 2021), to obtain more accurate results. Additionally, efforts to improve data quality and increase sample sizes can help reduce the risk of overfitting and enhance the robustness of artificial neural network-based predictive models in clinical applications.

Among the feature genes for constructing artificial neural networks, *KLKB1* and *PROKR1* have been confirmed to be related to coagulation function and the angiogenesis process. *KLKB1* is usually synthesized in hepatocytes and secreted into the blood and is involved in the surface-dependent activation of blood coagulation, fibrinolysis, kinin production, and biological processes of inflammation, which can reflect the severity of liver

injury (Che et al., 2021). The TBX20-PROK2-PROKR1 pathway may also be a target for the treatment of diseases associated with dysregulation of angiogenesis, benefit on patients with ischemic heart failure (Lichtenauer and Jung, 2018). Comparisons with similar studies might involve investigations into other receptor genes, such as *PROK2* (Lichtenauer and Jung, 2018), *EDN1* (Liang et al., 2018), and *NOS3* (Teralı and Ergören, 2019), which play roles in vascular function and inflammation regulation and may have similar implications in CAD.

SFTPA1, *FOXL1,* and *MAPK11* are tumor-characteristic molecular markers. *SFTPA1* variant carriers are at increased risk of inherited lung disease (Benusiglio et al., 2021), and this

gene may be a viable prognostic biomarker since it is connected to immune cell infiltration and the effectiveness of immunotherapy in lung cancer (Yuan et al., 2022). It has been established that FOXL1 is intricately linked to the onset and progression of glioma (Chen et al., 2019), renal cancer (Yang et al., 2014), and pancreatic cancer (Zhang et al., 2013). *MAPK11* plays a role in a variety of female tumors (breast cancer (He et al., 2014), uterine endometrial cancer (Li et al., 2019), cervical cancer, ovarian cancer, and uterine carcinosarcoma), and its expression levels are significantly reduced (Katopodis et al., 2021). In addition, *CYP2B6* is the only gene encoding a functional enzyme in the human CYP2B subfamily (Desta et al., 2021), genetic variation in this gene locus affects the metabolism or bioactivation of clinically important drugs bupropion (Kirchheiner et al., 2003) and efavirenz (Haas et al., 2004; Desta et al., 2007). Pathogenic mutations in FBN1 are the cause of Marfan syndrome, a life-threatening autosomal dominant disorder of connective tissue (Wang et al., 2022). Lipoma HMGIC fusion partner-like 5 (*LHFPL5*) is an important molecule in the normal auditory system involved in mechanotransduction pathways in sensory hair cells of the ear (Yu et al., 2020).

It is noteworthy that these feature genes have been implicated in various biological processes relevant to immune function and CAD pathogenesis. For instance, *KLKB1* encodes for plasma kallikrein, which plays a role in the kinin-kallikrein system and has been associated with inflammation and thrombosis (Hayama et al., 2016). *PROKR1* has been linked to angiogenesis and vascular development, both of which are closely intertwined with immune response modulation (Goryszewska et al., 2020). Furthermore, *MAPK11* is involved in the MAPK signaling pathway, which regulates immune cell activation and cytokine production (Roche et al., 2020). By elucidating the interplay between these genes and immune cells, we gain insights into the complex immunological mechanisms underlying CAD development and progression. This understanding may inform the development of novel immunomodulatory therapies and precision medicine approaches targeting immune-inflammatory pathways in CAD.

The evolution of precision therapeutics in the context of disease genomics offers promising avenues for enhancing patient care in various medical conditions, including coronary artery disease (CAD). By leveraging extensive data analysis and molecular classification, precision medicine approaches aim to tailor treatments to individual patients based on their specific genetic makeup and disease characteristics.

In the study mentioned, the use of artificial neural networks represents a novel approach to identifying characteristic genes associated with CAD from large-scale genomic data. These genes can serve as diagnostic biomarkers, allowing for more accurate diagnosis and even risk prediction of CAD. The ability of artificial neural networks to analyze complex gene interactions enhances our understanding of the genetic mechanisms underlying CAD, thereby improving diagnostic accuracy, particularly in patients with diverse genetic backgrounds. However, despite the potential benefits, challenges remain. Large-scale, high-quality genomic data are essential for training and optimizing artificial neural network models, highlighting the need for continued investment in data collection and curation efforts. Additionally, further validation of the effectiveness and reliability of the identified feature genes in real-world clinical settings is necessary to ensure their utility in improving patient outcomes.

Future research directions could explore integrating multi-omics data, such as proteomics, to enhance artificial neural network model's recognition capabilities further. Additionally, combining artificial neural networks with other advanced technologies like single-cell sequencing and gene editing may offer synergistic advantages in achieving more accurate and personalized diagnosis and treatment of CAD. In summary, precision therapeutics driven by advancements in disease genomics, coupled with innovative approaches like artificial neural networks, hold great promise for revolutionizing the diagnosis and treatment of CAD. Continued research efforts and technological advancements are crucial for overcoming existing challenges and realizing the full potential of precision medicine in cardiovascular healthcare.

Conclusion

In summary, our study identified two distinct molecular subgroups in coronary artery disease (CAD) related to coagulation function through gene expression profiling of CAD patients. We further investigated the biological function and immunological characteristics of these subgroups, revealing differing immunological roles between them. Utilizing LASSO and RF, we screened feature genes associated with coagulation function and developed an artificial neural network model for subgroup classification. The model exhibited excellent prediction accuracy, providing a theoretical framework for precision medicine in CAD by identifying patients with different molecular subgroups and suggesting novel medication therapy targets. This research significantly advances precision medicine in CAD by aligning with personalized treatment strategies and offering new avenues for improving patient outcomes. Future directions involve validating these molecular subgroups in larger patient cohorts and exploring their implementation in clinical settings to realize the full potential of precision medicine in CAD management.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

Author contributions

BZ: Data curation, Software, Visualization, Writing–original draft, Writing–review and editing. YL: Data curation, Formal Analysis, Methodology, Writing–original draft. GX: Conceptualization, Data curation, Project administration, Writing–review and editing.

Funding

Acknowledgments

We are deeply appreciative of the GEO database for providing publicly available clinical datasets that served as the data basis for our research.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2024.1351774/full#supplementary-material

References

Agosti, P., Mancini, I., Sadeghian, S., Pagliari, M. T., Abbasi, S. H., Pourhosseini, H., et al. (2023). Factor V Leiden but not the factor II 20210G>A mutation is a risk factor for premature coronary artery disease: a case-control study in Iran. *Res. Pract. thrombosis haemostasis* 7, 100048. doi:10.1016/j.rpth.2023.100048

Ajjan, R., and Grant, P. J. (2006). Coagulation and atherothrombotic disease. *Atheroscler. 186* 186, 240–259. doi:10.1016/j.atherosclerosis.2005.10.042

Amer, A. N., Gaballah, A., Emad, R., Ghazal, A., and Attia, N. (2021). Molecular epidemiology of HIV-1 virus in Egypt: a major change in the circulating subtypes. *Curr. HIV Res.* 19, 448–456. doi:10.2174/1570162x19666210805091742

Bareche, Y., Venet, D., Ignatiadis, M., Aftimos, P., Piccart, M., Rothe, F., et al. (2018). Unravelling triple-negative breast cancer molecular heterogeneity using an integrative multiomic analysis. *Ann. Oncol. official J. Eur. Soc. Med. Oncol.* 29, 895–902. doi:10.1093/annonc/mdy024

Beck, M. W. (2018). NeuralNetTools: visualization and analysis tools for neural networks. *J. Stat. Softw.* 85, 1–20. doi:10.18637/jss.v085.i11

Beineke, P., Fitch, K., Tao, H., Elashoff, M. R., Rosenberg, S., Kraus, W. E., et al. (2012). A whole blood gene expression-based signature for smoking status. *BMC Med. genomics* 5, 58. doi:10.1186/1755-8794-5-58

Benusiglio, P. R., Fallet, V., Sanchis-Borja, M., Coulet, F., and Cadranel, J. (2021). Lung cancer is also a hereditary disease. *Eur. Respir. Rev. official J. Eur. Respir. Soc.* 30, 210045. doi:10.1183/16000617.0045-2021

Brodzicki, A., Piekarski, M., and Jaworek-Korjakowska, J. (2021). The whale optimization algorithm approach for deep neural networks. *Sensors Basel, Switz.* 21, 8003. doi:10.3390/s21238003

Bunting, M. D., Vyas, M., Requesens, M., Langenbucher, A., Schiferle, E. B., Manguso, R. T., et al. (2022). Extracellular matrix proteins regulate NK cell function in peripheral tissues. *Sci. Adv.* 8, eabk3327. doi:10.1126/sciadv.abk3327

Che, Y. Q., Zhang, Y., Li, H. B., Shen, D., and Cui, W. (2021). Serum KLKB1 as a potential prognostic biomarker for hepatocellular carcinoma based on data-independent acquisition and parallel reaction monitoring. *J. Hepatocell. carcinoma* 8, 1241–1252. doi:10.2147/jhc.S325629

Chen, A., Zhong, L., and Lv, J. (2019). FOXL1 overexpression is associated with poor outcome in patients with glioma. *Oncol. Lett.* 18, 751–757. doi:10.3892/ol.2019.10351

Chen, X., Cheng, F., Liu, C., Cheng, L., and Mao, Y. (2021). An improved Wolf pack algorithm for optimization problems: design and evaluation. *PloS one* 16, e0254239. doi:10.1371/journal.pone.0254239

Cheng, X., Han, W., Liang, Y., Lin, X., Luo, J., Zhong, W., et al. (2022). Risk prediction of coronary artery stenosis in patients with coronary heart disease based on logistic regression and artificial neural network. *Comput. Math. methods Med.* 2022, 3684700. doi:10.1155/2022/3684700

Davis, S., and Meltzer, P. S. (2007). GEOquery: a bridge between the gene expression Omnibus (GEO) and BioConductor. *Bioinforma. Oxf. Engl.* 23, 1846–1847. doi:10.1093/bioinformatics/btm254

Desta, Z., El-Boraie, A., Gong, L., Somogyi, A. A., Lauschke, V. M., Dandara, C., et al. (2021). PharmVar GeneFocus: CYP2B6. *Clin. Pharmacol. Ther.* 110, 82–97. doi:10.1002/cpt.2166

Desta, Z., Saussele, T., Ward, B., Blievernicht, J., Li, L., Klein, K., et al. (2007). Impact of CYP2B6 polymorphism on hepatic efavirenz metabolism *in vitro*. *Pharmacogenomics* 8 (6), 547–558. doi:10.2217/14622416.8.6.547

Duggan, J. P., Peters, A. S., Trachiotis, G. D., and Antevil, J. L. (2022). Epidemiology of coronary artery disease. *Surg. Clin. N. Am.* 102, 499–516. doi:10.1016/j.suc.2022.01.007

Elashoff, M. R., Wingrove, J. A., Beineke, P., Daniels, S. E., Tingley, W. G., Rosenberg, S., et al. (2011). Development of a blood-based gene expression algorithm for assessment of obstructive coronary artery disease in non-diabetic patients. *BMC Med. genomics* 4, 26. doi:10.1186/1755-8794-4-26

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* 33, 1–22. doi:10.18637/jss.v033.i01

GBD 2019 Demographics Collaborators (2020). Global age-sex-specific fertility, mortality, healthy life expectancy (HALE), and population estimates in 204 countries and territories, 1950-2019: a comprehensive demographic analysis for the Global Burden of Disease Study 2019. *Lancet London, Engl.* 396, 1160–1203. doi:10.1016/s0140-6736(20)30977-6

Goryszewska, E., Kaczynski, P., Balboni, G., and Waclawik, A. (2020). Prokineticin 1-prokineticin receptor 1 signaling promotes angiogenesis in the porcine endometrium during pregnancy†. *Biol. reproduction* 103, 654–668. doi:10.1093/biolre/ioaa066

Guidi, G., Pettenati, M. C., Miniati, R., and Iadanza, E. (2013). Random Forest for automatic assessment of heart failure severity in a telemonitoring scenario. *Eng. Med. Biol. Soc. Annu. Int. Conf.* 2013, 3230–3233. doi:10.1109/embc.2013.6610229

Haas, D. W., Ribaudo, H. J., Kim, R. B., Tierney, C., Wilkinson, G. R., Gulick, R. M., et al. (2004). Pharmacogenetics of efavirenz and central nervous system side effects: an Adult AIDS Clinical Trials Group study. *Aids* 18, 2391–2400.

Hänzelmann, S., Castelo, R., and Guinney, J. (2013). GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinforma.* 14, 7. doi:10.1186/1471-2105-14-7

Hayama, T., Kamio, N., Okabe, T., Muromachi, K., and Matsushima, K. (2016). Kallikrein promotes inflammation in human dental pulp cells via protease-activated receptor-1. *J. Cell. Biochem.* 117, 1522–1528. doi:10.1002/jcb.25437

He, Z., He, J., Liu, Z., Xu, J., Yi, S. F., Liu, H., et al. (2014). MAPK11 in breast cancer cells enhances osteoclastogenesis and bone resorption. *Biochimie* 106, 24–32. doi:10.1016/j.biochi.2014.07.017

Hu, F., Wang, J., Zhang, M., Wang, S., Zhao, L., Yang, H., et al. (2021). Comprehensive analysis of subtype-specific molecular characteristics of colon cancer: specific genes, driver genes, signaling pathways, and immunotherapy responses. *Front. Cell Dev. Biol.* 9, 758776. doi:10.3389/fcell.2021.758776

Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., et al. (2017). Artificial intelligence in healthcare: past, present and future. *Stroke Vasc. neurology* 2, 230–243. doi:10.1136/svn-2017-000101

Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic acids Res.* 28, 27–30. doi:10.1093/nar/28.1.27

Katopodis, P., Kerslake, R., Zikopoulos, A., Beri, N., and Anikin, V. (2021). p38β-MAPK11 and its role in female cancers. *J. ovarian Res.* 14, 84. doi:10.1186/s13048-021-00834-9

Keihanian, F., Saeidinia, A., Bagheri, R. K., Johnston, T. P., and Sahebkar, A. (2018). Curcumin, hemostasis, thrombosis, and coagulation. *J. Cell. physiology* 233, 4497–4511. doi:10.1002/jcp.26249

Khera, A. V., and Kathiresan, S. (2017). Genetics of coronary artery disease: discovery, biology and clinical translation. *Nat. Rev. Genet.* 18, 331–344. doi:10.1038/nrg.2016.160

Kirchheiner, J., Klein, C., Meineke, I., Sasse, J., Zanger, U. M., Mürdter, T. E., et al. (2003). Bupropion and 4-OH-bupropion pharmacokinetics in relation to genetic polymorphisms in CYP2B6. *Pharmacogenetics Genomics* 13, 619–626. doi:10.1097/00008571-200310000-00005

Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E., and Storey, J. D. (2012). The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinforma. Oxf. Engl.* 28, 882–883. doi:10.1093/bioinformatics/bts034

Li, Y., Kong, C., Wu, C., Wang, Y., Xu, B., Liang, S., et al. (2019). Knocking down of LINC01220 inhibits proliferation and induces apoptosis of endometrial carcinoma through silencing MAPK11. *Biosci. Rep.* 39. doi:10.1042/bsr20181794

Liang, L. L., Chen, L., Zhou, M. Y., Cai, M. Y., Cheng, J., Chen, Y., et al. (2018). Genetic susceptibility of five tagSNPs in the endothelin-1 (EDN1) gene to coronary artery disease in a Chinese Han population. *Biosci. Rep.* 38. doi:10.1042/bsr20171320

Libby, P., Bonow, R. O., Tomaselli, G. F., Bhatt, D., Solomon, S. D., et al. (2021). Braunwald's heart disease-E-book: a textbook of cardiovascular medicine. *Elsevier Health Sci.*

Lichtenauer, M., and Jung, C. (2018). TBX20 and the PROK2-PROKR1 pathway-new kid on the block in angiogenesis research. *Ann. Transl. Med.* 6, S8. doi:10.21037/atm.2018.08.41

Menter, D. G., Davis, J. S., Broom, B. M., Overman, M. J., Morris, J., and Kopetz, S. (2019). Back to the colorectal cancer Consensus molecular subtype future. *Curr. Gastroenterol. Rep.* 21, 5. doi:10.1007/s11894-019-0674-9

Newman, R., and Tolar, P. (2021). Chronic calcium signaling in IgE(+) B cells limits plasma cell differentiation and survival. *Immunity* 54, 2756–2771.e10. doi:10.1016/j.immuni.2021.11.006

Paszek, E., Pociask, E., Ząbczyk, M., Butenas, S., and Undas, A. (2022). Activated factor XI is associated with increased factor VIIa - antithrombin complexes in stable coronary artery disease: impact on cardiovascular outcomes. *Eur. J. Clin. investigation* 52, e13857. doi:10.1111/eci.13857

Patil, A., Elwitigala, J. P., Rajapaksa, L., Gangakhedkar, R., Chaturbhuj, D., Pendse, R., et al. (2020). HIV-1 pol gene diversity and molecular dating of subtype C from Sri Lanka. *PloS one 15* 15, e0234133. doi:10.1371/journal.pone.0234133

Peng, W., Sun, Y., and Zhang, L. (2022). Construction of genetic classification model for coronary atherosclerosis heart disease using three machine learning methods. *BMC Cardiovasc. Disord.* 22, 42. doi:10.1186/s12872-022-02481-4

Pizzolo, F., Castagna, A., Olivieri, O., Girelli, D., Friso, S., Stefanoni, F., et al. (2021). Basophil blood cell count is associated with enhanced factor II plasma coagulant activity and increased risk of mortality in patients with stable coronary artery disease: not only neutrophils as prognostic marker in ischemic heart disease. *J. Am. Heart Assoc.* 10, e018243. doi:10.1161/jaha.120.018243

Qu, C., Zhang, L., Li, J., Deng, F., Tang, Y., Zeng, X., et al. (2021). Improving feature selection performance for classification of gene expression data using Harris Hawks optimizer with variable neighborhood learning. *Briefings Bioinforma.* 22, bbab097. doi:10.1093/bib/bbab097

Redondo, M., Watzke, H. H., Stucki, B., Sulzer, I., Biasiutti, F. D., Binder, B. R., et al. (1999). Coagulation factors II, V, VII, and X, prothrombin gene 20210G-->A transition, and factor V Leiden in coronary artery disease: high factor V clotting activity is an independent risk factor for myocardial infarction. *Arteriosclerosis, thrombosis, Vasc. Biol.* 19, 1020–1025. doi:10.1161/01.atv.19.4.1020

Roche, O., Fernández-Aroca, D. M., Arconada-Luque, E., García-Flores, N., Mellor, L. F., Ruiz-Hidalgo, M. J., et al. (2020). p38β and cancer: the beginning of the road. *Int. J. Mol. Sci.* 21, 7524. doi:10.3390/ijms21207524

Roth, G. A., Mensah, G. A., Johnson, C. O., Addolorato, G., Ammirati, E., Baddour, L. M., et al. (2020). Global burden of cardiovascular diseases and risk factors, 1990-2019: update from the GBD 2019 study. *J. Am. Coll. Cardiol.* 76, 2982–3021. doi:10.1016/j.jacc.2020.11.010

Salari, N., Shohaimi, S., Najafi, F., Nallappan, M., and Karishnarajah, I. (2014). A novel hybrid classification model of genetic algorithms, modified k-Nearest Neighbor and developed backpropagation neural network. *PloS one* 9, e112987. doi:10.1371/journal.pone.0112987

Sinnaeve, P. R., Donahue, M. P., Grass, P., Seo, D., Vonderscher, J., Chibout, S. D., et al. (2009). Gene expression patterns in peripheral blood correlate with the extent of coronary artery disease. *PloS one* 4, e7037. doi:10.1371/journal.pone.0007037

Steimann, F. (2001). On the use and usefulness of fuzzy sets in medical AI. *Artif. Intell. Med.* 21, 131–137. doi:10.1016/s0933-3657(00)00077-4

Stępień, K., Siudut, J., Konieczyńska, M., Nowak, K., Zalewski, J., and Undas, A. (2023). Effect of high-dose statin therapy on coagulation factors: lowering of factor XI as a modifier of fibrin clot properties in coronary artery disease. *Vasc. Pharmacol.* 149, 107153. doi:10.1016/j.vph.2023.107153

Teralı, K., and Ergören, M. (2019). The contribution of NOS3 variants to coronary artery disease: a combined genetic epidemiology and computational biochemistry perspective. *Int. J. Biol. Macromol.* 123, 494–499. doi:10.1016/j.ijbiomac.2018.11.128

Tomczak, K., Czerwińska, P., and Wiznerowicz, M. (2015). The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemp. Oncol. Poznan, Pol.* 19, A68–A77. doi:10.5114/wo.2014.47136

Topham, J. T., Karasinska, J. M., Lee, M. K. C., Csizmok, V., Williamson, L. M., Jang, G. H., et al. (2021). Subtype-discordant pancreatic ductal adenocarcinoma tumors show intermediate clinical and molecular characteristics. *Clin. cancer Res. official J. Am. Assoc. Cancer Res.* 27, 150–157. doi:10.1158/1078-0432.Ccr-20-2831

Valeriani, E., Pastori, D., Astorri, G., Porfidia, A., Menichelli, D., and Pignatelli, P. (2023). Factor V Leiden, prothrombin, MTHFR, and PAI-1 gene polymorphisms in patients with arterial disease: a comprehensive systematic-review and meta-analysis. *Thrombosis Res.* 230, 74–83. doi:10.1016/j.thromres.2023.08.006

Wang, J. J., Yu, B., Sun, Y., Song, X., Wang, D. W., and Li, Z. (2022). FBN1 splice-altering mutations in marfan syndrome: a case report and literature review. *Genes* 13, 1842. doi:10.3390/genes13101842

Wang, Q., Li, M., Yang, M., Yang, Y., Song, F., Zhang, W., et al. (2020). Analysis of immune-related signatures of lung adenocarcinoma identified two distinct subtypes: implications for immune checkpoint blockade therapy. *Aging* 12, 3312–3339. doi:10.18632/aging.102814

Wong, B. K., and Monaco, J. A. (1995). Expert system applications in business: a review and analysis of the literature (1977–1993). *Inf. Manag.* 29, 141–152. doi:10.1016/0378-7206(95)00023-p

Yang, F. Q., Yang, F. P., Li, W., Liu, M., Wang, G. C., Che, J. P., et al. (2014). Foxl1 inhibits tumor invasion and predicts outcome in human renal cancer. *Int. J. Clin. Exp. pathology* 7, 110–122.

Yu, X., Zhao, Q., Li, X., Chen, Y., Tian, Y., Liu, S., et al. (2020). Deafness mutation D572N of TMC1 destabilizes TMC1 expression by disrupting LHFPL5 binding. *Proc. Natl. Acad. Sci. U. S. A.* 117, 29894–29903. doi:10.1073/pnas.2011147117

Yuan, L., Wu, X., Zhang, L., Yang, M., Wang, X., Huang, W., et al. (2022). SFTPA1 is a potential prognostic biomarker correlated with immune cell infiltration and response to immunotherapy in lung adenocarcinoma. *Cancer Immunol. Immunother. CII* 71, 399–415. doi:10.1007/s00262-021-02995-4

Yuan, S., Burgess, S., Laffan, M., Mason, A. M., Dichgans, M., Gill, D., et al. (2021). Genetically proxied inhibition of coagulation factors and risk of cardiovascular disease: a mendelian randomization study. *J. Am. Heart Assoc.* 10, e019644. doi:10.1161/jaha.120.019644

Zhang, G., He, P., Gaedcke, J., Ghadimi, B. M., Ried, T., Yfantis, H. G., et al. (2013). FOXL1, a novel candidate tumor suppressor, inhibits tumor aggressiveness and predicts outcome in human pancreatic cancer. *Cancer Res.* 73, 5416–5425. doi:10.1158/0008-5472.Can-13-0362

Zhao, S., Ma, D., Xiao, Y., Li, X. M., Ma, J. L., Zhang, H., et al. (2020). Molecular subtyping of triple-negative breast cancers by immunohistochemistry: molecular basis and clinical relevance. *Oncol.* 25, e1481–e1491. doi:10.1634/theoncologist.2019-0982