

## Research Article

# Differential Regulation of the Immune System in Peripheral Blood Following Ischemic Stroke

Wenhao Liu <sup>1</sup>, Xin-Zhuang Yang,<sup>2</sup> Dingding Zhang,<sup>2</sup> Xin He,<sup>1</sup> Qianlan Yu,<sup>1</sup> Xinquan Liu,<sup>1</sup> and Yi Dai <sup>1</sup>

<sup>1</sup>Department of Neurology, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

<sup>2</sup>Medical Research Center, State Key Laboratory of Complex Severe and Rare Diseases, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing, China

Correspondence should be addressed to Yi Dai; [pumchdy@pumch.cn](mailto:pumchdy@pumch.cn)

Received 22 December 2021; Accepted 9 May 2022; Published 8 June 2022

Academic Editor: Shtaywy S. Abdalla

Copyright © 2022 Wenhao Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

**AIM.** Previous studies have provided insights into complex immune system changes caused by ischemic stroke (IS), while detailed reports are lacking especially in peripheral blood. Here, we sought to identify genetic biomarkers in immune system which significantly associated with the occurrence of IS and explore candidate drugs that can regulate the process. We also investigated whether gene expression alternation of immune genes contributed to differential distribution of immune cells in peripheral blood following IS. **Method.** 108 IS samples and 47 matched controls were obtained from the GEO database. Immune-related genes (IRGs) and their associated drugs were collected from the ImmPort and PharmGBK databases, respectively. Random forest (RF) regression and least absolute shrinkage and selection operator (LASSO) logistic regression were applied to identify immune-related genetic biomarkers (IRGBs) of IS, and accuracy was verified using neural network models. Finally, proportion changes of various immune cells in peripheral blood of IS patients were evaluated using CIBERSORT and xCell and correlation analyses were performed between IRGBs and differentially distributed immune cells. **Results.** A total of 537 genes were differentially expressed between IS and control samples. Four immune-related differential expressed genes identified by regression analysis presented strong predictive power (AUC = 0.909) which we suggested them as immune-related genetic biomarkers (IRGBs). We also demonstrated six immune-related genes targeted by known drugs. In addition, post-IS immune system presented an increase in the proportion of innate immune cells and a decrease in adaptive immune cells in the peripheral circulation, and IRGBs showing significance were associated with this process. **Conclusion.** The study identified *CARD11*, *ICAM2*, *VIM*, and *CD19* as immune-related genetic biomarkers of IS. Six immune-related DEGs targeted by known drugs were found and provide new candidate drug targets for modulating the post-IS immune system. The innate immune cells and adaptive immune cells are diversified in the post-IS immune system, and IRGBs might play important role during this process.

## 1. Introduction

Stroke is the second leading cause of death and the main cause of disability among adults worldwide [1]. When an acute stroke (particularly ischemic stroke (IS)) is followed by a severe inflammatory attack, immune cells (such as microglia) inside the central nervous system (CNS) are activated and exert macrophage-like effects [2, 3]. More importantly, as the blood-brain barrier is disrupted, immune cells

in the peripheral blood circulation can invade the CNS [4]. Proinflammatory signals from immune-mediators not only activate immune cells in situ but also recruit inflammatory cells (neutrophils, monocytes, macrophages, various types of T cells, and other inflammatory cells) from the peripheral blood to the ischemic area and induce a systemic inflammatory response [5]. Conversely, this early activation of the immune system is quickly replaced by a state of systemic immunosuppression that predisposes individuals to

poststroke infections [4, 6]. In patients with stroke, the incidence of infectious complications, pneumonia, urinary tract infections, and other organ system infections is approximately 30%, and these symptoms are strongly associated with death and high recurrence rates in stroke survivors [7, 8].

Although changes in the immune system are relevant to the occurrence and development of IS, the pathogenesis underlying their variation is still unclear. Previous studies have reported that damaged brains can remodel peripheral immunity and inhibit the function of the peripheral immune system, leading to lymphopenia, reduced levels of inflammatory cytokines, and secondary lymphoid organ atrophy [3, 4, 9, 10]. The CNS can cause variations in immune function through complex humoral and neural pathways, including the sympathetic nervous system, vagus nerve, and hypothalamic-pituitary-adrenal (HPA) axis [9]. However, it remains unclear which signals and mechanisms trigger the sympathetic nervous system and the HPA axis to regulate the immune system following IS [6]. Comprehensive elucidation of the alteration in each component of the poststroke immune system is lacking, and studies on the molecular mechanisms underlying the changes are limited.

In this study, we identified immune-related genes whose expression highly associated with ischemic stroke and analyzed the correlation between these genes and different components of immune cells, in an attempt to understand the molecular mechanisms of immune system changes following IS and provide more evidence for the development of drugs for immune response in the future.

## 2. Materials and Methods

**2.1. Microarray Data and Clinical Information.** The microarray datasets used in this study were obtained from the GEO database (<http://www.ncbi.nlm.nih.gov/geo/>). The criteria for retrieving the datasets were (A) human peripheral whole blood samples, (B) all case samples were collected within acute phase of ischemic stroke, (C) gene expression profiling, (D) datasets containing both patients and healthy people without a history of stroke, where all patients were clinically diagnosed with ischemic stroke using medical imaging techniques (magnetic resonance imaging or X-ray computed tomography), and (E) datasets of patients with hemorrhagic stroke were excluded.

To ensure consistency and completeness of the datasets, we manually identified relevant literature using keyword filters. Finally, two datasets (GSE16561 [11–13] and GSE58294 [14]) were included and treated as “training set.” These datasets were merged, and batch-effects between different datasets were corrected using the “combat” function in the SVA package (version: 3.38.0) [11]. Next, we normalized the merged dataset and completed covariate adjustment, using the “NormalizeBetweenArrays” and “removeBatchEffect” function in the limma package (version: 3.46.0) [15, 16]. For validation of biomarker genes, we downloaded GSE22255 [17], GSE37587 [18], and GSE46480 [19] datasets

which satisfied most criteria while only containing IS or normal cohorts (Table 1).

**2.2. Identification of Differential Expressed Genes and Functional Annotation.** To identify differentially expressed genes (DEGs) in peripheral blood samples from ischemic stroke patients and controls, we performed differential expression analysis using the limma package (version: 3.46.0), controlling for age. The threshold for screening DEGs was  $|\log_2 \text{FC (fold change)}| > 0.5$  and  $\text{FDR} < 0.001$ . Enrichment analysis of Gene Ontology (GO) and Disease Ontology (DO) was performed on DEGs using the clusterProfiler package (version: 3.18.1) [20]. Reference to the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (<http://www.genome.jp/kegg/>) and gene set enrichment analysis (GSEA) was carried out on the gene expression matrix [21, 22]. The significance of KEGG signaling pathways was set at  $\text{FDR} < 0.05$ .

**2.3. Immune-Related Genes and Drug Targets.** Immune-related genes (IRGs) and “variant, gene, and drug relationship” datasets were downloaded from the ImmPort database (<https://www.immport.org/shared/genelists>) and PharmGBK database (<https://www.pharmgkb.org/downloads>), respectively [23, 24]. The intersection of DEGs, IRGs, and drug target genes (DTGs) was then used to generate immune-related DEGs targeted by drugs and potential drugs that may contribute to the changes in the post-IS immune system.

**2.4. Selection and Validation of Immune-Related Genetic Biomarkers.** Four IRGBs in IS, namely, *CARD11*, *ICAM2*, *VIM*, and *CD19*, were identified from the immune-related DEGs using LASSO logistic regression and random forest (RF) regression algorithms with training datasets. The LASSO algorithm was derived from the glmnet package (version: 4.1-1) [25]. The neural network was built using tensorflow2 framework (version: 2.3.0) using Python language and based on the merged dataset [26, 27]. GSE37587 and GSE46480 were used as test sets to verify the sensitivity and efficiency of IRGBs in IS diagnosis. Although the data sample is derived from peripheral blood mononuclear cells (PBMCs), GSE22255 was also used as an independent test set for additional validation.

**2.5. Immune Cell Infiltration Evaluation.** CIBERSORT immune system analysis tool and the deconvolute\_Xcell function in the immunedeconv package (version: 2.0.4) were used to generate immune cell composition profiles for all samples [28, 29]. The CIBERSORT immune system analysis tool results in an expression matrix of 22 immune cells in all samples of the merged dataset, while the deconvolute\_Xcell function results in an expression matrix of 67 immune-related variables versus all samples. We then used *t*-test to analyze the differences in immune cell components between IS patients and healthy people. Finally, Spearman’s correlation analysis was performed between IRGBs and significantly differentiated immune cells using the ggstatsplot package (version: 0.7.2). The ggplot2 package (version: 3.3.3) was used to generate visual heatmaps [30].

TABLE 1: All datasets used in this study contain a total of 361 samples, among which there were 196 cases and 165 controls. All case samples were collected acute phase of ischemic stroke.

Datasets (GEO ID)	Data Case	Data Control	Sample type/ source	References	Category	GPL
GSE16561	39	24	Peripheral blood	(Barr et al., 2010; O'Connell et al., 2016; O'Connell et al., 2017) [11–13]	Train	GPL570
GSE58294	69	23	Peripheral blood	(Stamova et al., 2014) [14]	Train	GPL570
GSE37587	68	0	Peripheral blood	(Barr et al., 2015) [18]	Test	GPL6883
GSE46480	0	98	Peripheral blood	(Issa et al., 2016) [19]	Test	GPL570
GSE22255	20	20	PBMCs	(Krug et al., 2012) [17]	Test	GPL570

TABLE 2: Clinical characters of the merged data set.

	Total sample, N(%)	Stroke N = 62 (56.9%), N(%)	Control, N = 47 (43.1%), N(%)	Statistics/df	P value
Gender (% female)	58(53.2%)	33(53.2%)	25(53.2%)	$X^2$ 0.000013/1	0.9972
Age, y, mean $\pm$ SD	66.7 $\pm$ 16.86	72.6 $\pm$ 12.09	58.9 $\pm$ 7.51	$t$ -13.90302/107	<0.001
Race (% white)	97(89.0%)	54(87.1%)	43(91.5%)	$X^2$ 0.526507/1	0.4681
Hypertension	64(58.7%)	41(66.1%)	23(48.9%)	$X^2$ 3.26002/1	0.07099
Diabetes	22(20.2%)	15(24.2%)	7(14.9%)	$X^2$ 1.43527/1	0.2309
Dyslipidemia	40(36.7%)	24(38.7%)	16(34.0%)	$X^2$ 0.250672/1	0.6166

### 3. Result

#### 3.1. DEGs Enriched in Immune-Related Biological Process.

We first generated an independent dataset of 155 samples consisting of 108 IS patients and 47 matched controls, by merging two datasets: GSE16561 and GSE58294 (Tables 1 and 2, Materials and Methods). To ensure data consistency, batch-effect was controlled and the different subsets were normalized. The results showed that data preprocessing was effective and reliable (Figures 1(a)–1(d), Figure S1). Next, we performed differential analysis of gene expression by controlling age and 537 DEGs between IS patients and healthy controls were identified. (Figure 1(e)).

GO enrichment analysis showed that DEGs were mainly associated with immune receptor activation, pattern recognition receptor activation, and NAD<sup>+</sup> nucleotidase activation (Figure 1(f)), while DO enrichment analysis showed that the diseases enriched within DEGs included mainly arteriosclerosis, atherosclerosis, arteriosclerotic cardiovascular disease, myocardial infarction, obesity, nutritional diseases, and coronary artery disease (Figure S2A). These diseases are intimately associated with IS, indicating that the DEGs are involved in the occurrence and development of IS.

Furthermore, the GSEA results showed that the molecular pathways enriched with DEGs were complement and coagulation cascades, neutrophil extracellular trap formation (NETs), lipid and atherosclerosis, tumor necrosis factor (TNF) signaling pathway, Toll-like receptor signaling pathway, and NOD-like receptor signaling pathway (Figure S2B). Of these, the Toll-like receptor signaling pathway, NOD-like receptor signaling pathway, and TNF- $\alpha$  signaling pathways play

important roles in innate immunity (Figure S2C), consistent with the results of the GO enrichment analysis. These results provide evidence that immune-related biological processes might play important roles in IS.

#### 3.2. Four Immune-Related Genetic Biomarkers Perform Well in IS Diagnosis.

To further determine which kinds of immune genes are dramatically altered and associated with the occurrence of IS, we performed LASSO logistic regression analysis on immune-related DEGs generated from the intersection of IRGs and DEGs and found five genetic markers in peripheral blood (Figure 2(a), Materials and Methods). Among them, *VIM* was positively correlated with the occurrence of stroke, while *CARD11*, *ICAM2*, *CD19*, and *CCR7* were negatively correlated with the occurrence of stroke (Figure 2(b)). Meanwhile, ten genetic markers that reached significance were discovered using the RF algorithm (Figure 2(c), Materials and Methods). Surprisingly, four genes, *VIM*, *CARD11*, *ICAM2*, and *CD19*, were identified by both methods and we suggested them as immune-related genetic biomarkers (IRGBs) which might be used for auxiliary diagnosis of the disease. In order to eliminate the effect of age difference between the two groups, we controlled participants' age during analysis. The results demonstrated that gene expression of IRGBs showed no correlation with age and confirmed that our findings are truly between patients and healthy controls (Figure S3). To further verify the sensitivity and accuracy of these four genes in IS diagnosis, GSE37587 and GSE46480 were used as validation datasets, and a neural network model was used to examine the prediction performance (Figure S4). The results were excellent (AUC = 0.909; 95%CI = [0.861, 0.953], Figure 2(d)).

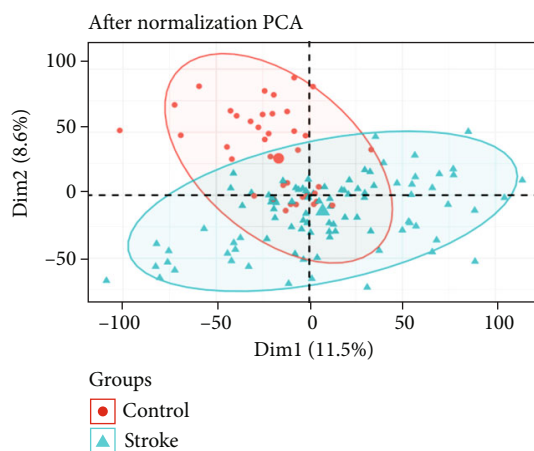
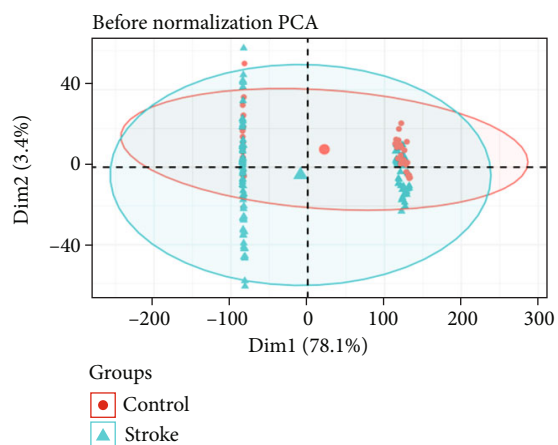
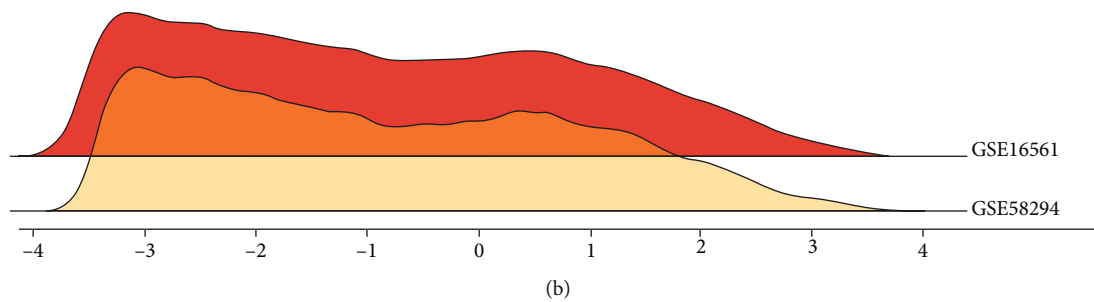
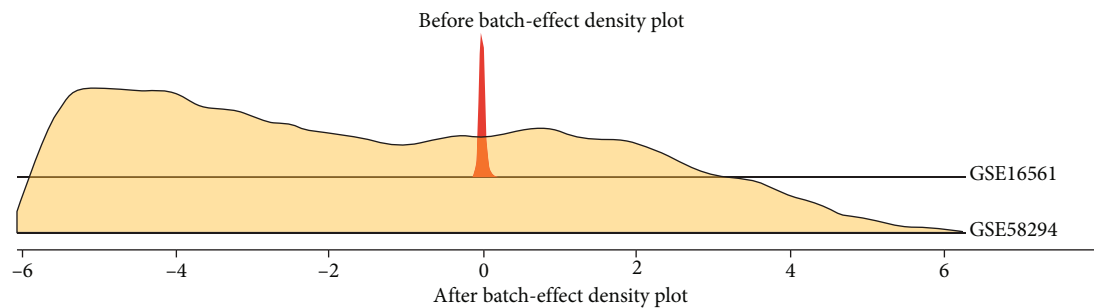
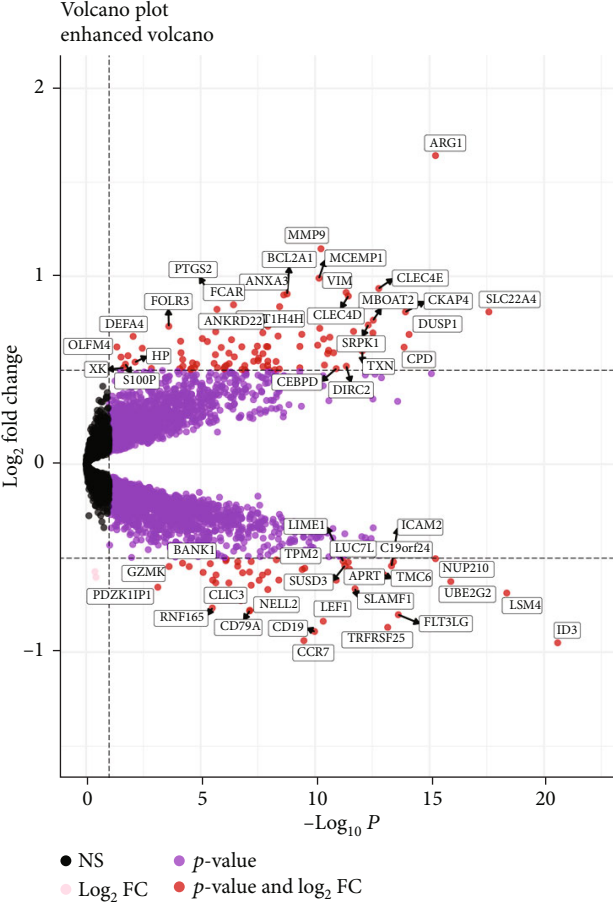


FIGURE 1: Continued.



Total = 13251 variables

(e)

FIGURE 1: Continued.

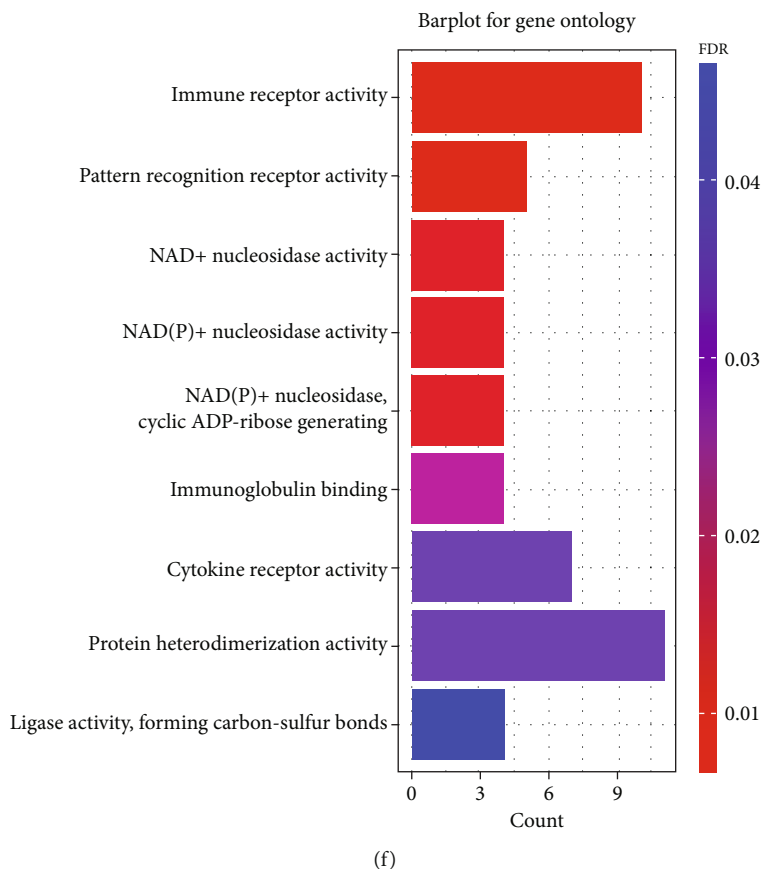


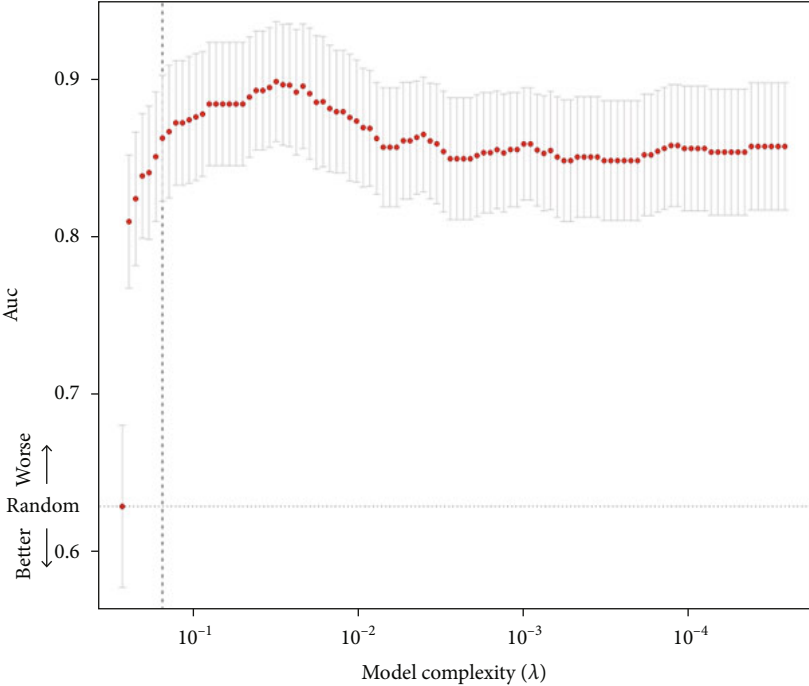
FIGURE 1: (a) The ridge plot of the merged dataset before eliminating the batch effect. (b) The ridge plot of the merged dataset after eliminating the batch effect. (c) The Principal Component Analysis (PCA) plot of the merged dataset before normalization. (d) The PCA plot of the stroke group and control group of the merged dataset after normalization. Note: In PCA analysis, all genes were used to observe the effect of normalization. (e) Volcano map of DEGs; red represents significant differential genes, grey represents no significant difference genes, pink represents genes with differential  $\log_2FC$ , and purple represents genes with differential  $P$  value. (f) GO enrichment analysis, where the horizontal axis represents the number of DEGs under the GO term.

We also randomly selected four DEGs and performed the same analysis. The prediction performance of these randomly selected DEGs was poor ( $AUC = 0.654$ ;  $95\%CI = [0.577, 0.739]$ , Figure 2(e)), demonstrating that the our IRGBs were significantly associated with the occurrence of IS. To further verify the conclusion, we select another dataset whose samples derived from PBMCs, and the diagnostic ability of the IRGBs is still good ( $AUC = 0.835$ ;  $95\%CI = [0.701, 0.951]$ , Figure S5).

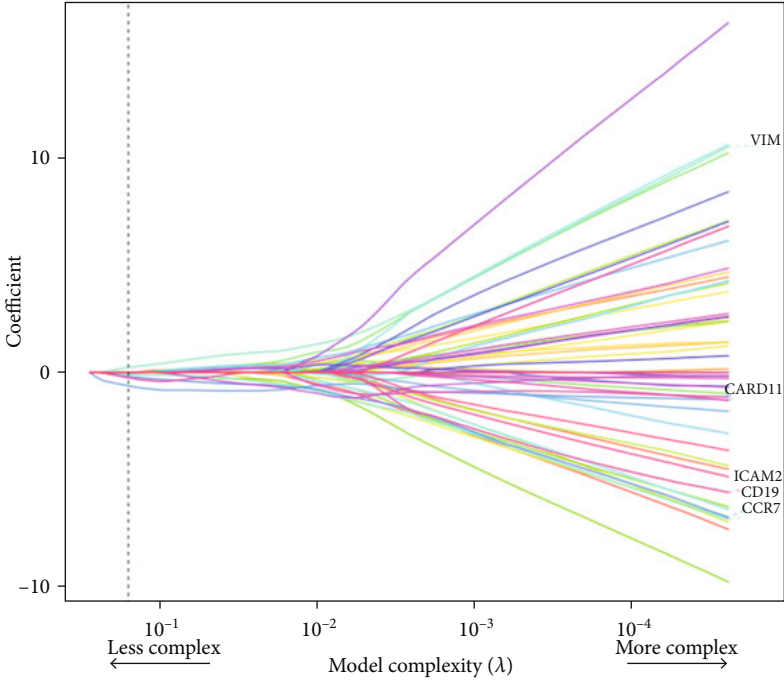
Considering that immune-related genes present significant changes after the onset of stroke (Figure 1(f)), we sought to explore whether there are drugs that can mitigate the process. We identified overlapping genes across three gene sets, DEGs, IRGs, and DTGs, and found six immune-related DEGs with known drug targets, including *PLCG1*, *TLR2*, *GSK3B*, *TLR4*, *ADM*, and *PTGS2* (Figure 3(a)). Drugs targeting these six genes include aspirin, pyrazolone (e.g., phenylbutazone), propionate derivatives (e.g., ibuprofen), diclofenac, paracetamol, and TNF- $\alpha$  inhibitors such as pravastatin (Figures 3(b) and 3(c)). Aspirin is already recommended for early treatment of stroke. In addition to acting as an antiplatelet, it plays an important role in modulating the post-IS immune environment [31]. TLR2 and TLR4 are

pattern recognition receptors (Figure 3(b)) and reveal the pattern recognition function of the Toll-like receptor signaling pathway in regulating the post-IS immune system.

**3.3. IRGBs Are Associated with the Change of Immune System following IS.** We applied the CIBERSORT classification algorithm to the merged dataset to demonstrate changes of the immune system after IS. Interestingly, the proportion of naive B cells, M0-type macrophages, monocytes, neutrophils,  $CD8^+$  T cells, and  $CD4^+$  naive T cells was significantly different between IS patients and healthy controls (Figure 4(a)). M0-type macrophages and monocytes were significantly enriched in the peripheral blood of IS patients, while naive B cells,  $CD8^+$  T cells, and naive  $CD4^+$  T cells were significantly decreased. In addition, using the xCell deconvolution algorithm to compute enrichment scores for 67 immune-related factors, we found that 12 immune cells represented significantly different constituent components between IS patients and healthy controls (Figure 4(b)). More specifically, neutrophils, macrophages, and monocytes were significantly enriched in the peripheral blood of IS patients compared with healthy controls, consistent with the results of the CIBERSORT classification algorithm. Additionally,

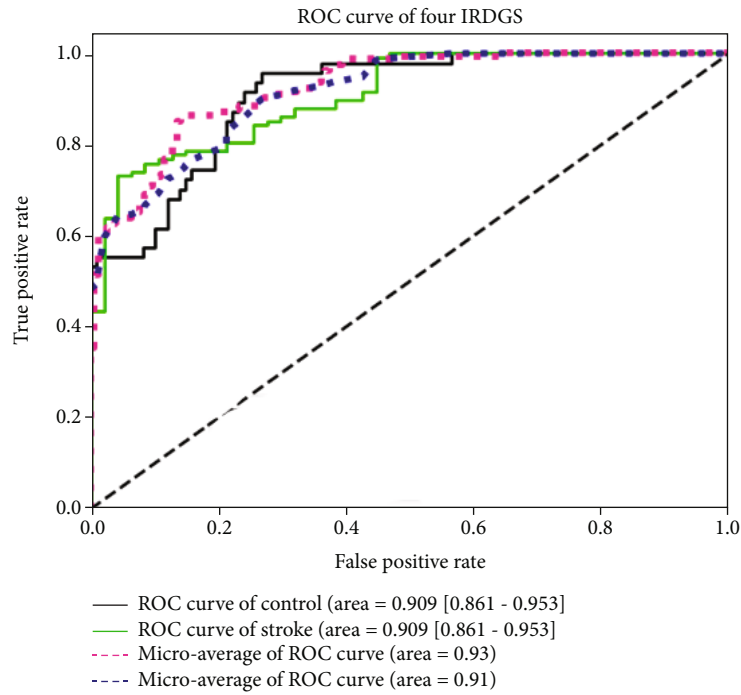


(a)

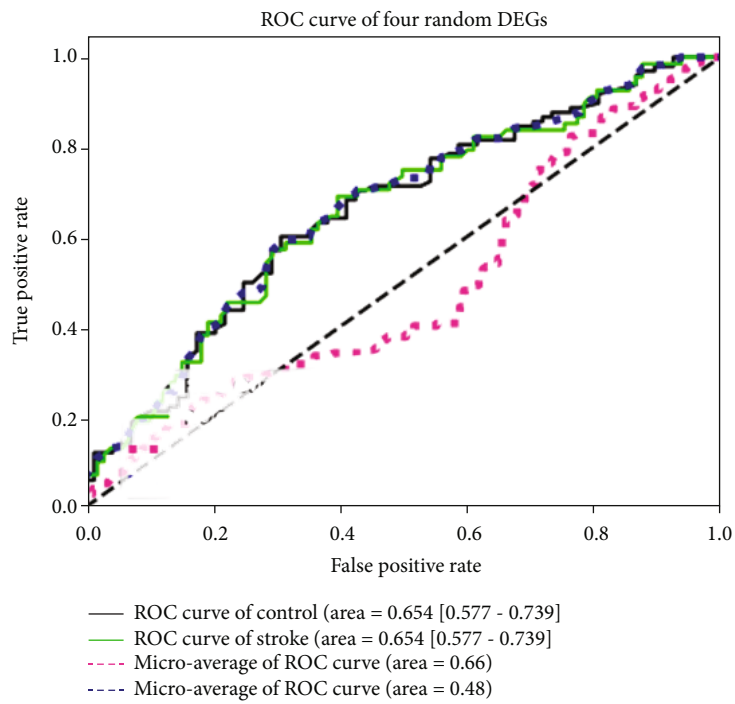


(b)

FIGURE 2: Continued.



(c)



(d)

FIGURE 2: Continued.



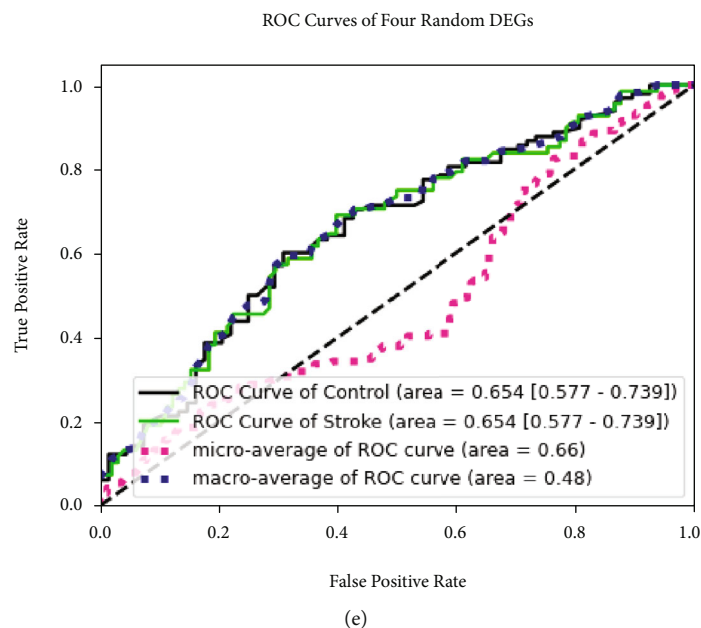


FIGURE 2: (a) and (b) Construction of the IRG diagnostic classifier by the LASSO logistic regression algorithm shows the process of dimension reduction and coefficient decomposition. The horizontal axis represents the complexity of models, and the vertical axis represents the AUC value of models and model coefficient of labeled gene expression, respectively. (c) Construction of the IRG diagnostic classifier by the random forest regression algorithm. (d) The ROC curve of the neural network diagnosis model using four selected diagnostic markers. (e) The ROC curve of the neural network diagnosis model using four random DEGs.

there were 9 immune cells with enrichment fractions significantly lower than in healthy controls, including multiple types of B cells and T cells.

Correlation analysis indicated a close relationship between IRGBs and 18 differentially distributed immune cells (Figure 4(c)). *CARD11* was positively correlated with CD8<sup>+</sup> effective memory T cells ( $r = 0.47$ ,  $P < 0.001$ , Figure 4(c)) and negatively correlated with macrophages ( $r = -0.48$ ,  $P < 0.001$ ); *ICAM2* was positively correlated with CD8<sup>+</sup> T cells ( $r = 0.46$ ,  $P < 0.001$ ) and negatively correlated with neutrophils ( $r = 0.61$ ,  $P < 0.001$ ); *VIM* was slightly correlated with neutrophils ( $r = 0.32$ ,  $P < 0.001$ ) and macrophages ( $r = 0.3$ ,  $P < 0.001$ ); and *CD19* was positively correlated with B cells ( $r = 0.84$ ,  $P < 0.001$ ), naive B cells ( $r = 0.8$ ,  $P < 0.001$ ), and memory B cells ( $r = 0.78$ ,  $P < 0.001$ ). Our analysis showed that IRGBs are significantly associated with changes in the immune environment after IS, demonstrating an important role in IS diagnosis and therapy.

#### 4. Discussion

Acute stroke (mainly IS) is followed by severe inflammatory episodes. With the blood-brain barrier being disrupted, immune cells in the peripheral circulation invade the CNS. Many studies have demonstrated that neutrophils, monocytes, macrophages, and various types of T cells in the peripheral blood infiltrate the ischemic area and induce systemic inflammatory responses [32–35]. However, this early activation of the immune system is present only for a short period and is generally replaced by a state of systemic immunosuppression that predisposes one to poststroke infection

rapidly after stroke. Both morbidity and mortality due to infectious diseases are much higher in IS patients than in the normal population, and the alteration in the immune system is closely associated with a high recurrence rate in stroke survivors [7, 8].

In this study, 537 DEGs were found between IS patients and healthy controls. Four significantly differential expressed immune-related genes, namely, *CARD11*, *ICAM2*, *VIM*, and *CD19*, were identified by both LASSO logistic regression and RF algorithm analysis. Furthermore, these four genes were used to construct a neural network prediction model and presented excellent predictive value and accuracy in validation datasets.

*CARD11* is a scaffold protein composed of 1154 amino acids and belongs to the membrane-associated guanylate kinase (MAGUK) superfamily [36]. *CARD11* is expressed in peripheral leukocytes and is an important component of the antigen inducible NF- $\kappa$ B signaling pathway in T cells [37]. Intercellular adhesion molecule-2 (*ICAM2*) belongs to the ICAM family of adhesion proteins [38]. It is widely expressed in vascular endothelial cells and peripheral blood cells and plays a key role in cell-cell interactions, promoting neutrophil binding to, and transmigration across, the vascular endothelium [39]. *ICAM2* is also a key gene in platelet leukocyte interaction and participates in IS by regulating the platelet leukocyte aggregation process [40]. *VIM* protein (vimentin) is part of the intermediate membrane system, which plays an important role in maintaining cell shape, resisting mechanical stress, transmitting cytoskeletal cross-talk, and organizing signaling molecules [41]. Vimentin also has an important role in maintaining plasticity in the nervous system, and several studies have demonstrated that its

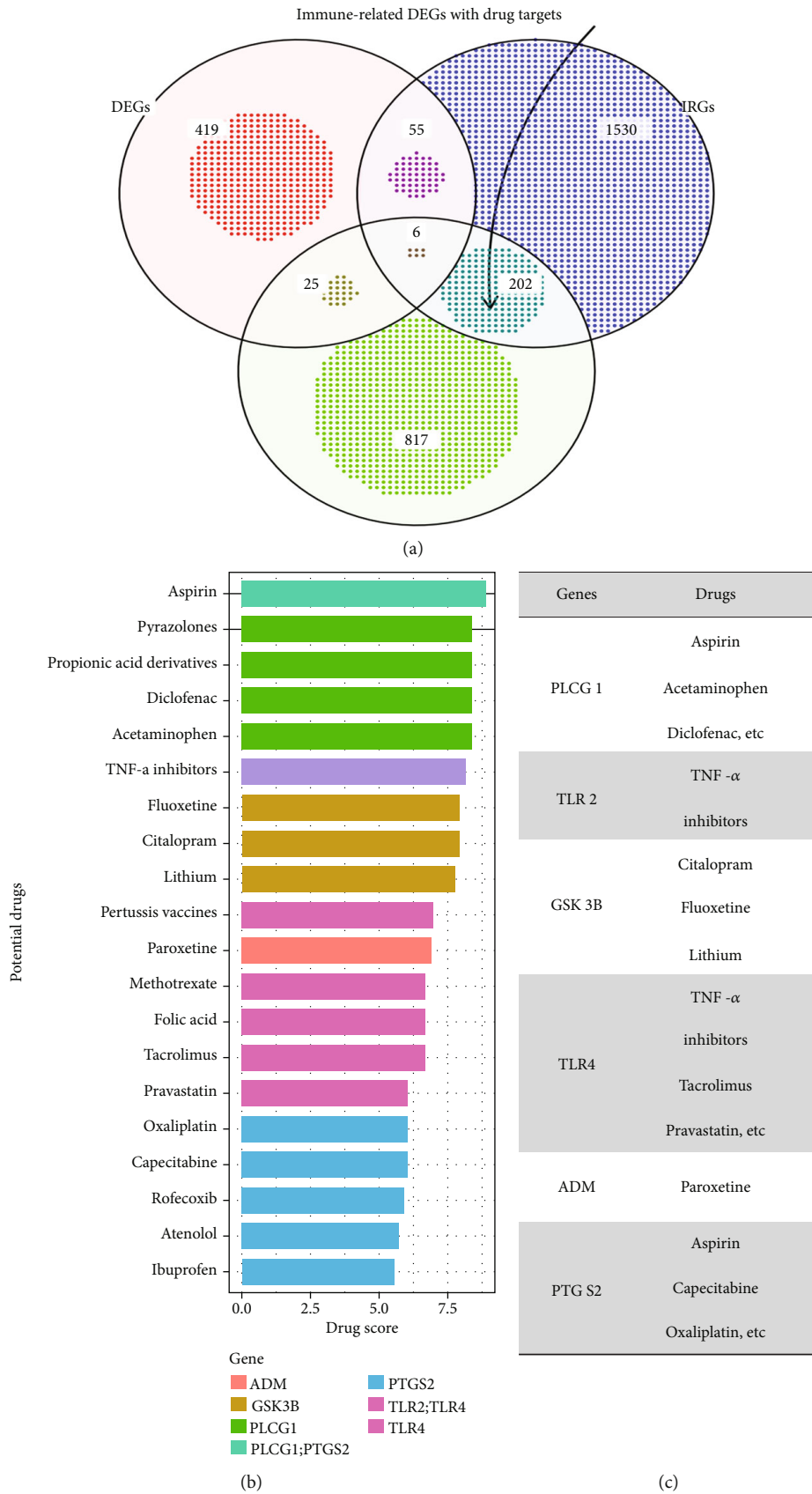
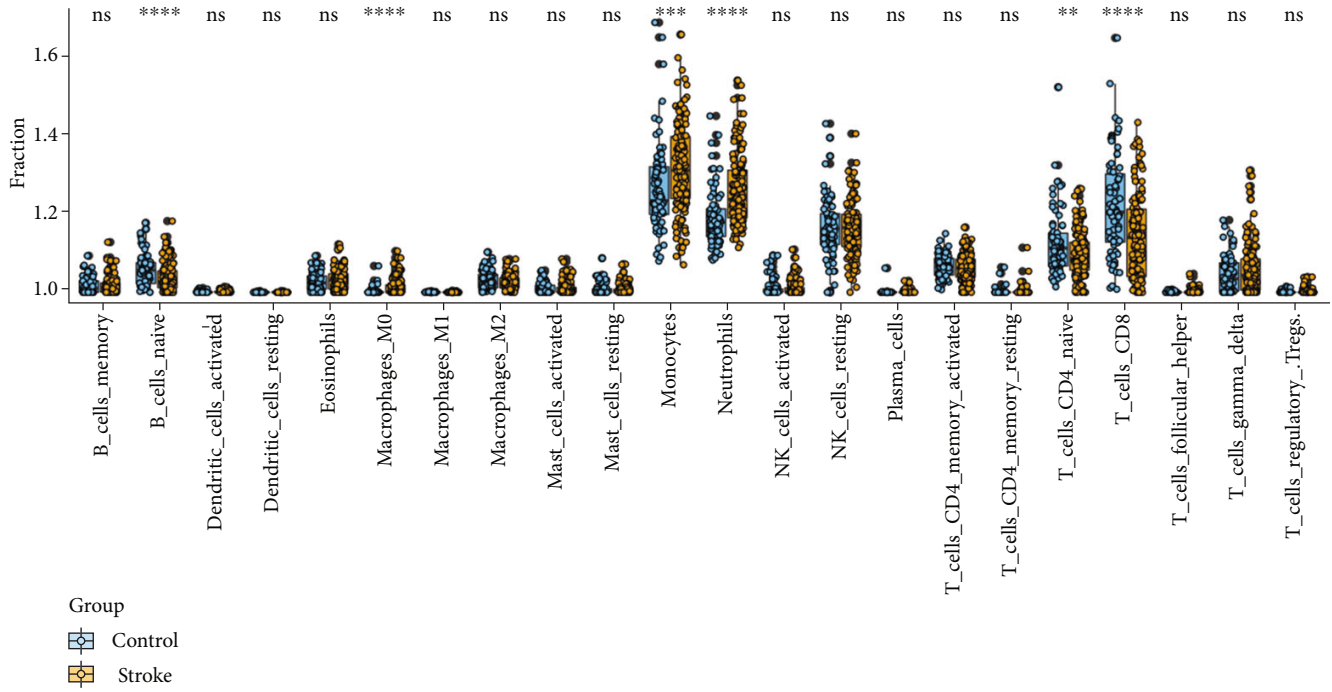
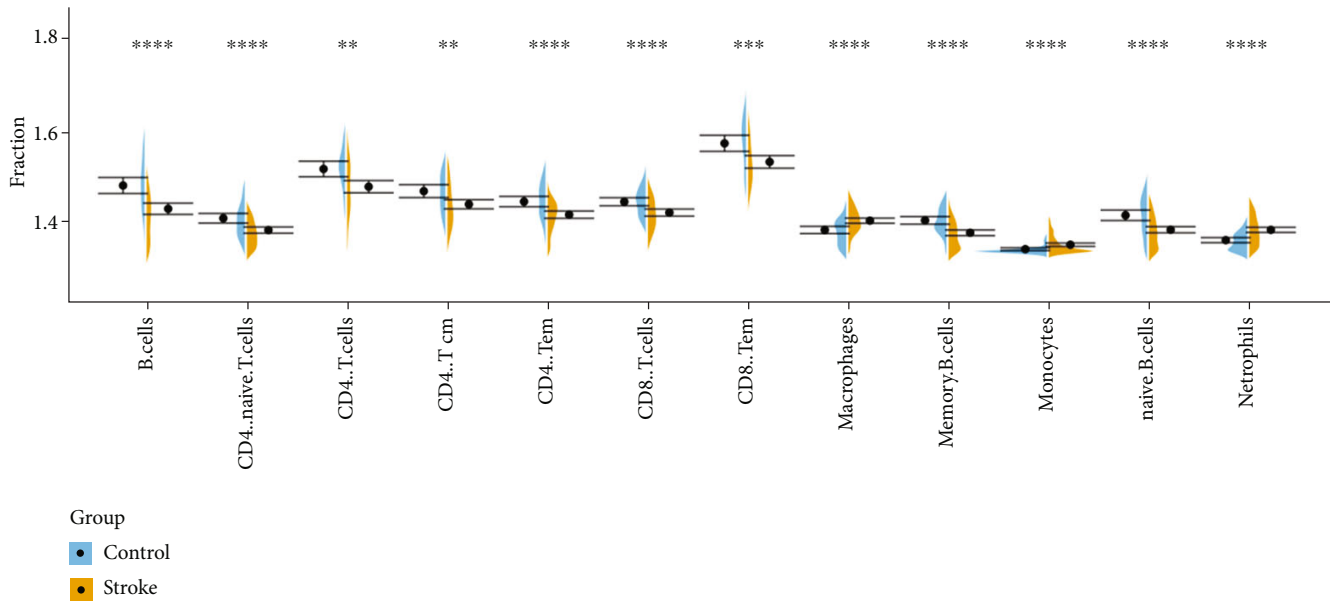


FIGURE 3: (a) Venn diagram shows the target genes of potential drugs obtained by the their gene sets. (b) and (c) Barplot and table show potential drugs and their targeted genes.



(a)



(b)

FIGURE 4: Continued.

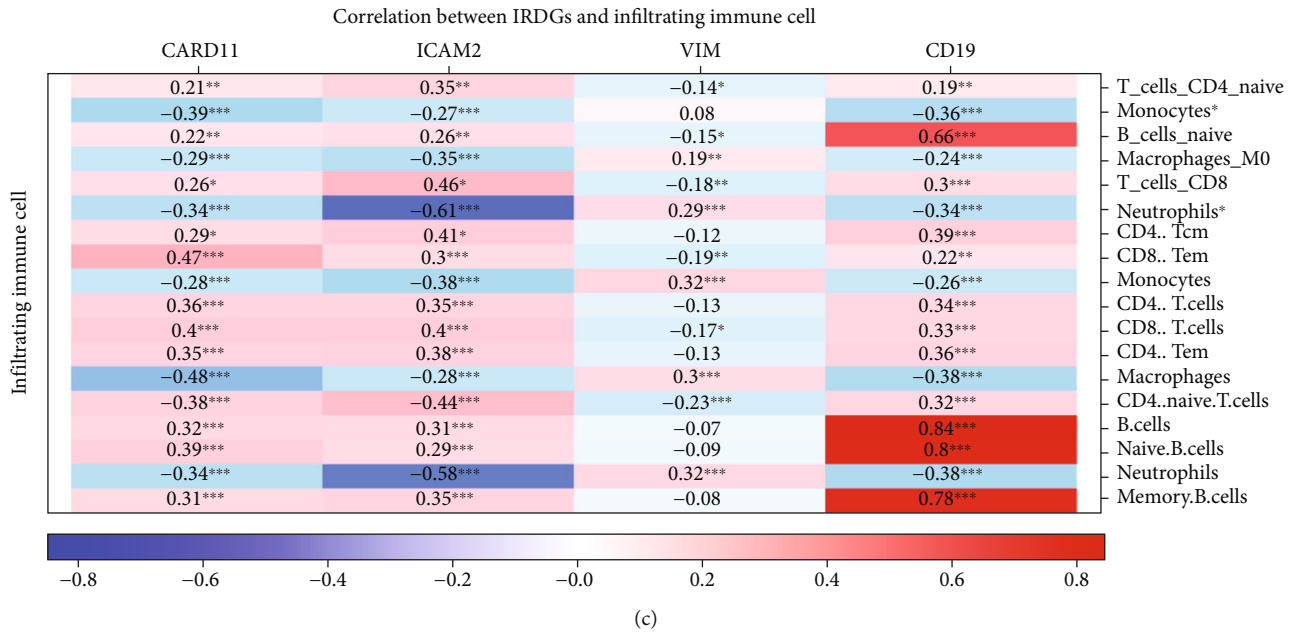


FIGURE 4: (a) and (b) Violin diagram of the proportion of immune cells obtained by CIBERSORT and xcell tool. The yellow marks represent the difference in immune cell proportion between the two groups of samples. \* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$ ; \*\*\*\* $P < 0.0001$ . (c) Correlation between *CARD11*, *ICAM2*, *VIM*, *CD19*, and significant differential immune cells. \*: Monocytes\* and neutrophils\* were generated by CIBERSORT tool. \* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$ ; \*\*\*\* $P < 0.0001$ .

expression is upregulated after ischemic stroke [42–45]. *CD19* is widely expressed during B cell development and is a key coreceptor for B cell antigen receptor signal transduction [46]. Courties et al. showed that cerebral ischemic injury leads to an organismal stress response that activates the HPA axis and mediates B lymphopoiesis deficiency, which could explain the downregulated expression of *CD19* in the peripheral blood of IS patients [9]. The neural network prediction model exhibited excellent accuracy and sensitivity using these four IRGBs.

GO enrichment analysis showed that DEGs were mainly associated with immune receptor activation, pattern recognition receptor activation, and  $\text{NAD}^+$  nucleotidase activation. The diseases enriched by DO were mainly atherosclerosis, myocardial infarction, obesity, nutritional diseases, and coronary artery disease. The results of GO enrichment analysis indicated that the immune response plays an important role in IS, and the diseases enriched via DO are highly correlated with IS. Moreover, the pathways enriched in GSEA mainly involved complement and lectin cascades, NETs, lipid and atherosclerosis, TNF signaling pathway, Toll-like receptor signaling pathway, and NOD-like receptor signaling pathway. Previous studies have demonstrated that the signaling pathways of complement and coagulation cascades, lipids, and atherosclerosis are the core factors of IS [47–49]. Laridan et al. showed that neutrophils and NETs are important components of cerebral thrombi and are more abundant in thrombi for longer, suggesting that NETs play an important role in the progression of IS [50]. TNF- $\alpha$  is one of the typical proinflammatory cytokines. Cui et al. suggested that it plays an important role in the development of IS, and the activation of microglia induces TNF- $\alpha$  expression and activation and

recruitment of circulating neutrophils, monocytes, and lymphocytes into the CNS [51]. Toll-like receptors, key pattern recognition receptors in innate immunity, are also crucial for NET formation by neutrophils as well as infiltration by monocytes, contributing to thrombus formation [52]. NOD-like receptor proteins containing nucleotide-binding oligomerization domains can recognize damage-associated molecular patterns in sterile inflammation [53]. Yang et al. found that NADPH oxidase-mediated NLRP3 signaling contributed to cerebral ischemic injury by aggravating inflammation and neurovascular injury [54]. The above findings were consistent with our analysis results.

In addition to differentially expressed genes, six immune-related DEGs were found to be targeted by drugs: *PLCG1*, *TLR2*, *GSK3B*, *TLR4*, *ADM*, and *PTGS2*. Drugs targeting these six genes include aspirin, pyrazolone (e.g., phenylbutazone), propionate derivatives (e.g., ibuprofen), diclofenac, paracetamol, and the TNF- $\alpha$  inhibitor, pravastatin. Aspirin and statins, alone or in combination, have been shown to modulate the secretome profile of peripheral blood monocyte macrophages after stroke [55]. Toll-like receptor signaling pathway is important in mediating the formation of immune system after stroke, and previous animal experiments have shown that TLR2 or TLR4 deficient mice have less brain tissue damage after stroke than wild-type mice [56].

To further explore immune cell changes after IS, we performed immune cell enrichment analysis on the merged dataset using CIBERSORT and xCell tools. The results showed that the enrichment fractions of neutrophils, macrophages, and monocytes increased while the enrichment fractions of various types of B cells and T cells decreased, which may be due to the occurrence and development of IS.

Previous studies have shown that neutrophils are one of the first blood-derived immune cells to enter and be stationed in brain tissue in most experimental stroke models [32, 33]. Weston et al. reported that neutrophil infiltration was elevated on the first day, peaked on the third day, and began to decline, but was still present at 7 and 15 days after cerebral ischemia, with the degree of increase positively correlated with infarct volume and functional deficits [34]. Kaito et al. found an increase in the total monocyte-macrophage numbers in the human peripheral blood circulation early after brain injury [32]. In the chronic phase of IS, 3–7 days after the onset of ischemia, monocyte-derived macrophages from the peripheral blood circulation peak at the injured site [4]. GöKhan et al. found that, in contrast to neutrophils, monocytes, and macrophages, patients with ischemic stroke have a reduced number of lymphocytes in the peripheral blood and consequently an increased neutrophil/lymphocyte ratio [33]. Previous studies have shown that after IS, B cell development in the bone marrow is severely impaired at the pro-B cell stage and that this impairment leads to peripheral lymphopenia [9]. During this period, there is evidence of a shift in CD4<sup>+</sup> T cells from a Th1 type response mediated by cellular immunity to a Th2 type response mediated by humoral immunity, and this shift protects the brain from further damage [35]. However, the immune system is suppressed, and the number of T lymphocytes and B lymphocytes in the peripheral blood is reduced. The evidence from these studies combined with the results from our analysis suggests that neutrophils, macrophages, monocytes, various types of B cells, and T cells play important roles in IS and should be the focus of further research. There is currently a lack of in-depth research exploring the relationship between B cells and T cells in each category and the occurrence of IS, and further experiments are needed to reveal more details. Furthermore, our results revealed correlations between the four IRGBs and differentially expressed immune cells in IS. *CARD11* was significantly positively correlated with CD8<sup>+</sup> effector memory T cells and negatively correlated with macrophages; *ICAM2* was significantly positively correlated with CD8<sup>+</sup> T cells and negatively correlated with neutrophils; *VIM* was positively correlated with neutrophils and macrophages; and *CD19* was significantly positively correlated with B cells, naive B cells, and memory B cells. *CARD11* is ubiquitously expressed in peripheral leukocytes and is important in inducing the activation of NF- $\kappa$ B signaling pathway in T cells, but its relationship with macrophages is currently unclear. CD8<sup>+</sup> T cell-mediated immunosurveillance depends on the LFA-1: ICAM adhesion pathway for recognition, whereas ICAM2 is involved in neutrophil-mediated plasma leakage, perhaps explaining the association between *ICAM2* and CD8<sup>+</sup> T cells and neutrophils [38, 39]. Multiple studies have shown that vimentin can activate ERK1/2 signaling to recruit macrophages and neutrophils [41–43]. *CD19* is widely expressed on almost all B cells, and their significant positive relationship justifies the analysis results. Research is still needed to clarify the relationship between IRGBs and immune cells.

In this study, RF algorithm and LASSO logistic regression algorithm were used to search for IS genetic biomark-

ers, and a neural network was built to verify the accuracy. RF is a nonparametric tree-based machine learning method that searches for optimal variables with minimum depth statistics [57]. LASSO logistic regression belongs to the shrinkage estimation, and during the reduction process of regression coefficients, some insignificant regression coefficients can be directly reduced to 0, that is, to the function of variable screening [58]. These two algorithms were used to screen the feature variables and construct the best classification model. In this study, *CARD11*, *ICAM2*, *VIM*, and *CD19* were identified as genetic biomarkers for IS by combining the RF and LASSO logistic regression methods. The biomarkers selected by integration performed excellently in the neural network prediction model.

The ImmPort and PharmGBK databases were applied, based on a comprehensive analysis of gene expression and drug activity, to select immune-related DEGs targeted by drugs and new drug targets that could modulate the post-IS immune system. To our knowledge, it is the first time that CIBERSORT and xCell have been used together, in the analysis of post-IS immune system changes in peripheral blood, and their performance has been verified to each other. CIBERSORT and xCell were both based on limited genetic data that may deviate from heterotypic interactions of cells, disease-induced disorders, or phenotypic plasticity. The information on immune genes and drug targets included in the ImmPort and PharmGBK databases was also incomplete. Furthermore, our study represents secondary mining and analysis of previously published datasets. Although several previous findings are consistent with our analysis, the reliability of the results of this study requires further experimental validation.

## Data Availability

The microarray data used to support the findings of this study have been deposited in the GEO database (GSE16561: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE16561>; GSE58294: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE58294>; GSE37587: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE37587>; GSE46480: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE46480>; GSE22255: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE22255>).

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Authors' Contributions

Wenhao Liu and, Xin-Zhuang Yang, and contributed equally to this study.

## Acknowledgments

This work was supported by the National Basic Research Program of China (973 Program) [No. 2016YFC0901004]

and National Natural Science Foundation of China [31900481].

## Supplementary Materials

Supplement Figure 1: quantile-quantile (Q-Q) plots show that the interbatch differences among four selected datasets are eliminated. Gray shading represents the 95% confidence intervals around the normal distribution. Supplement Figure 2: (A) Disease ontology (DO) enrichment analysis. The size of the circle corresponds to the number of differentially expressed genes (DEGs) under the DO term. (B) Gene set enrichment analysis, where the horizontal axis represents up-and-down regulation and the curve area represents the number of DEGs under the gene set enrichment analysis (GSEA) term. (C) Gene set enrichment analysis. *P* values were determined using the Kolmogorov-Smirnov test. Supplement Figure 3: correlation analysis between gene expression of IRGBs and sample age. Supplement Figure 4: structure of the neural network. Supplement Figure 5: the ROC curve of the neural network diagnosis model using four selected genetic biomarkers in GSE22255 dataset. (*Supplementary Materials*)

## References

- [1] S. S. Virani, A. Alonso, E. J. Benjamin et al., “Heart disease and stroke Statistics-2020 update: a report from the American Heart Association,” *Circulation*, vol. 141, no. 9, pp. e139–e596, 2020.
- [2] J. Anrather and C. Iadecola, “Inflammation and stroke: an overview,” *Neurotherapeutics*, vol. 13, no. 4, pp. 661–670, 2016.
- [3] D. Levard, I. Buendia, A. Lanquetin, M. Glavan, D. Vivien, and M. Rubio, “Filling the gaps on stroke research: focus on inflammation and immunity,” *Brain, Behavior, and Immunity*, vol. 91, pp. 649–667, 2021.
- [4] R. L. Jayaraj, S. Azimullah, R. Beiram, F. Y. Jalal, and G. A. Rosenberg, “Neuroinflammation: friend and foe for ischemic stroke,” *Journal of Neuroinflammation*, vol. 16, no. 1, p. 142, 2019.
- [5] M. Hagen, J. A. Sembill, M. I. Sprugel et al., “Systemic inflammatory response syndrome and long-term outcome after intracerebral hemorrhage,” *Neurology: Neuroimmunology & Neuroinflammation*, vol. 6, no. 5, article e588, 2019.
- [6] U. Dirnagl, J. Klehmet, J. S. Braun et al., “Stroke-induced immunodepression: experimental evidence and clinical relevance,” *Stroke*, vol. 38, no. 2, pp. 770–773, 2007.
- [7] K. Shi, K. Wood, F. D. Shi, X. Wang, and Q. Liu, “Stroke-induced immunosuppression and poststroke infection,” *Stroke and Vascular Neurology*, vol. 3, no. 1, pp. 34–41, 2018.
- [8] K. Malone, S. Amu, A. C. Moore, and C. Waeber, “The immune system and stroke: from current targets to future therapy,” *Immunology and Cell Biology*, vol. 97, no. 1, pp. 5–16, 2019.
- [9] G. Courties, V. Frodermann, L. Honold et al., “Glucocorticoids regulate bone marrow B lymphopoiesis after stroke,” *Circulation Research*, vol. 124, no. 9, pp. 1372–1385, 2019.
- [10] F. E. Mercier, C. Ragu, and D. T. Scadden, “The bone marrow at the crossroads of blood and immunity,” *Nature Reviews. Immunology*, vol. 12, pp. 49–60, 2011.
- [11] T. L. Barr, Y. Conley, J. Ding et al., “Genomic biomarkers and cellular pathways of ischemic stroke by RNA gene expression profiling,” *Neurology*, vol. 75, no. 11, pp. 1009–1014, 2010.
- [12] G. C. O’Connell, A. B. Petrone, M. B. Treadway et al., “Machine-learning approach identifies a pattern of gene expression in peripheral blood that can accurately detect ischaemic stroke,” *NPJ Genomic Medicine*, vol. 1, no. 1, p. 16038, 2016.
- [13] G. C. O’Connell, M. B. Treadway, A. B. Petrone et al., “Peripheral blood AKAP7 expression as an early marker for lymphocyte-mediated post-stroke blood brain barrier disruption,” *Scientific Reports*, vol. 7, no. 1, p. 1172, 2017.
- [14] B. Stamova, G. C. Jickling, B. P. Ander et al., “Gene expression in peripheral immune cells following cardioembolic stroke is sexually dimorphic,” *PLoS One*, vol. 9, no. 7, article e102550, 2014.
- [15] J. T. Leek, W. E. Johnson, H. S. Parker, A. E. Jaffe, and J. D. Storey, “The sva package for removing batch effects and other unwanted variation in high-throughput experiments,” *Bioinformatics*, vol. 28, no. 6, pp. 882–883, 2012.
- [16] M. E. Ritchie, B. Phipson, D. Wu et al., “Limma powers differential expression analyses for RNA-sequencing and microarray studies,” *Nucleic Acids Research*, vol. 43, no. 7, article e47, 2015.
- [17] T. Krug, J. P. Gabriel, R. Taipa et al., “TTC7B Emerges as a novel risk factor for ischemic stroke through the convergence of several genome-wide approaches,” *Journal of Cerebral Blood Flow and Metabolism*, vol. 32, no. 6, pp. 1061–1072, 2012.
- [18] T. L. Barr, R. VanGilder, S. Rellick et al., “A genomic profile of the immune response to stroke with implications for stroke recovery,” *Biological Research for Nursing*, vol. 17, no. 3, pp. 248–256, 2015.
- [19] A. N. Issa, N. M. Herman, R. J. Wentz, B. J. Taylor, D. C. Summerfield, and B. D. Johnson, “Association of cognitive performance with time at altitude, sleep quality, and acute mountain sickness symptoms,” *Wilderness & Environmental Medicine*, vol. 27, no. 3, pp. 371–378, 2016.
- [20] G. Yu, L. G. Wang, Y. Han, and Q. Y. He, “clusterProfiler: an R package for comparing biological themes among gene clusters,” *OMICS*, vol. 16, no. 5, pp. 284–287, 2012.
- [21] M. Kanehisa, M. Furumichi, M. Tanabe, Y. Sato, and K. Morishima, “KEGG: new perspectives on genomes, pathways, diseases and drugs,” *Nucleic Acids Research*, vol. 45, no. D1, pp. D353–D361, 2017.
- [22] A. Subramanian, P. Tamayo, V. K. Mootha et al., “Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 43, pp. 15545–15550, 2005.
- [23] R. B. Altman, “PharmGKB: a logical home for knowledge relating genotype to drug response phenotype,” *Nature Genetics*, vol. 39, no. 4, p. 426, 2007.
- [24] S. Bhattacharya, P. Dunn, C. G. Thomas et al., “ImmPort, toward repurposing of open access immunological assay data for translational and clinical research,” *Scientific Data*, vol. 5, no. 1, p. 180015, 2018.
- [25] S. Engebretsen and J. Bohlin, “Statistical predictions with *glmnet*,” *Epigenetics*, vol. 11, no. 1, p. 123, 2019.
- [26] P. Sarang, *Artificial neural networks with TensorFlow 2*, Apress, Berkeley, CA, 2021.
- [27] O. G. Yalçın, *Applied neural networks with TensorFlow 2*, Apress, Berkeley, CA, 2021.

- [28] D. Aran, Z. Hu, and A. J. Butte, "xCell: digitally portraying the tissue cellular heterogeneity landscape," *Genome Biology*, vol. 18, no. 1, p. 220, 2017.
- [29] A. M. Newman, C. L. Liu, M. R. Green et al., "Robust enumeration of cell subsets from tissue expression profiles," *Nature Methods*, vol. 12, no. 5, pp. 453–457, 2015.
- [30] M. Postma and J. Goedhart, "PlotsOfData—a web app for visualizing data together with their summaries," *PLoS Biology*, vol. 17, no. 3, article e3000202, 2019.
- [31] N. Satani, K. Giridhar, C. Cai et al., "Aspirin in stroke patients modifies the immunomodulatory interactions of marrow stromal cells and monocytes," *Brain Research*, vol. 1720, p. 146298, 2019.
- [32] M. Kaito, S. Araya, Y. Gondo et al., "Relevance of distinct monocyte subsets to clinical course of ischemic stroke patients," *PLoS One*, vol. 8, no. 8, article e69409, 2013.
- [33] S. Gökhan, A. Ozhasenekler, H. Mansur Durgun, E. Akil, M. Ustündag, and M. Orak, "Neutrophil lymphocyte ratios in stroke subtypes and transient ischemic attack," *European Review for Medical and Pharmacological Sciences*, vol. 17, no. 5, pp. 653–657, 2013.
- [34] R. M. Weston, N. M. Jones, B. Jarrott, and J. K. Callaway, "Inflammatory cell infiltration after endothelin-1-induced cerebral ischemia: histochemical and myeloperoxidase correlation with temporal changes in brain injury," *Journal of Cerebral Blood Flow and Metabolism*, vol. 27, no. 1, pp. 100–114, 2007.
- [35] W. N. Jin, R. Gonzales, Y. Feng et al., "Brain ischemia induces diversified Neuroantigen-specific T-cell responses that exacerbate brain injury," *Stroke*, vol. 49, no. 6, pp. 1471–1478, 2018.
- [36] J. K. Yamamoto-Furusako, G. Fonseca-Camarillo, J. Furuzawa-Carballeda et al., "Caspase recruitment domain (CARD) family (CARD9, CARD10, CARD11, CARD14 and CARD15) are increased during active inflammation in patients with inflammatory bowel disease," *Journal of Inflammation*, vol. 15, no. 1, p. 13, 2018.
- [37] L. Liu, H. Guo, A. Song et al., "Progranulin inhibits LPS-induced macrophage M1 polarization via NF- $\kappa$ B and MAPK pathways," *BMC Immunology*, vol. 21, no. 1, p. 32, 2020.
- [38] K. Halai, J. Whiteford, B. Ma, S. Nourshargh, and A. Woodfin, "ICAM-2 facilitates luminal interactions between neutrophils and endothelial cells in vivo," *Journal of Cell Science*, vol. 127, pp. 620–629, 2014.
- [39] R. Lyck and G. Enzmann, "The physiological roles of ICAM-1 and ICAM-2 in neutrophil migration into tissues," *Current Opinion in Hematology*, vol. 22, no. 1, pp. 53–59, 2015.
- [40] B. Schmalbach, O. Stepanow, A. Jochens, C. Riedel, G. Deuschl, and G. Kühlenbaumer, "Determinants of platelet-leukocyte aggregation and platelet activation in stroke," *Cerebrovascular Diseases*, vol. 39, no. 3–4, pp. 176–180, 2015.
- [41] S. X. Jiang, J. Slinn, A. Aylsworth, and S. T. Hou, "Vimentin participates in microglia activation and neurotoxicity in cerebral ischemia," *Journal of Neurochemistry*, vol. 122, no. 4, pp. 764–774, 2012.
- [42] T. N. Mak and H. Bruggemann, "Vimentin in Bacterial Infections," *Cell*, vol. 5, no. 2, p. 5, 2016.
- [43] A. Datta, Q. Jingru, T. H. Khor, M. T. Teo, K. Heese, and S. K. Sze, "Quantitative neuroproteomics of an in vivo rodent model of focal cerebral ischemia/reperfusion injury reveals a temporal regulation of novel pathophysiological molecular markers," *Journal of Proteome Research*, vol. 10, no. 11, pp. 5199–5213, 2011.
- [44] Z. Liu, Y. Li, Y. Cui et al., "Beneficial effects of gfap/vimentin reactive astrocytes for axonal remodeling and motor behavioral recovery in mice after stroke," *Glia*, vol. 62, no. 12, pp. 2022–2033, 2014.
- [45] U. Wilhelmsson, L. Li, M. Pekna et al., "Absence of glial fibrillary acidic protein and vimentin prevents hypertrophy of astrocytic processes and improves post-traumatic regeneration," *The Journal of Neuroscience*, vol. 24, no. 21, pp. 5016–5021, 2004.
- [46] X. Li, Y. Ding, M. Zi et al., "CD19, from bench to bedside," *Immunology Letters*, vol. 183, pp. 86–95, 2017.
- [47] Y. Ma, Y. Liu, Z. Zhang, and G. Y. Yang, "Significance of complement system in ischemic stroke: a comprehensive review," *Aging and Disease*, vol. 10, no. 2, pp. 429–462, 2019.
- [48] W. Herrington, B. Lacey, P. Sherliker, J. Armitage, and S. Lewington, "Epidemiology of atherosclerosis and the potential to reduce the global burden of Atherothrombotic disease," *Circulation Research*, vol. 118, no. 4, pp. 535–546, 2016.
- [49] H. J. A. van Os, M. J. H. Wermer, F. R. Rosendaal, J. W. Govers-Riemslog, A. Algra, and B. S. Siegerink, "Intrinsic coagulation pathway, history of headache, and risk of ischemic stroke," *Stroke*, vol. 50, no. 8, pp. 2181–2186, 2019.
- [50] E. Laridan, F. Denorme, L. Desender et al., "Neutrophil extracellular traps in ischemic stroke thrombi," *Annals of Neurology*, vol. 82, no. 2, pp. 223–232, 2017.
- [51] G. Cui, H. Wang, R. Li et al., "Polymorphism of tumor necrosis factor alpha (TNF-alpha) gene promoter, circulating TNF-alpha level, and cardiovascular risk factor for ischemic stroke," *Journal of Neuroinflammation*, vol. 9, no. 1, p. 235, 2012.
- [52] A. Lasek-Bal, H. Jedrzejowska-Szypulka, S. Student et al., "The importance of selected markers of inflammation and blood-brain barrier damage for short-term ischemic stroke prognosis," *Journal of Physiology and Pharmacology*, vol. 70, no. 2, 2019.
- [53] T. Gong, L. Liu, W. Jiang, and R. Zhou, "DAMP-sensing receptors in sterile inflammation and inflammatory diseases," *Nature Reviews. Immunology*, vol. 20, no. 2, pp. 95–112, 2020.
- [54] F. Yang, Z. Wang, X. Wei et al., "NLRP3 deficiency ameliorates neurovascular damage in experimental ischemic stroke," *Journal of Cerebral Blood Flow and Metabolism*, vol. 34, no. 4, pp. 660–667, 2014.
- [55] N. Satani, X. Zhang, K. Giridhar et al., "A combination of atorvastatin and aspirin enhances the pro-regenerative interactions of marrow stromal cells and stroke-derived monocytes in vitro," *Frontiers in Pharmacology*, vol. 12, article 589418, 2021.
- [56] A. Chamorro, A. Meisel, A. M. Planas, X. Urra, D. van de Beek, and R. Veltkamp, "The immunology of acute stroke," *Nature Reviews. Neurology*, vol. 8, no. 7, pp. 401–410, 2012.
- [57] C. Lindner, P. A. Bromiley, M. C. Ionita, and T. F. Cootes, "Robust and accurate shape model matching using random forest regression-voting," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 9, pp. 1862–1874, 2015.
- [58] S. Ren, S. Huang, J. Ye, and X. Qian, "Safe feature screening for generalized LASSO," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 12, pp. 2992–3006, 2018.