Received: 8 January 2026 | Revised: 31 March 2026 | Accepted: 15 April 2026 | Published online: 6 May 2026

RESEARCH ARTICLE

Smart Wearable Technology
2026, Vol. 00(00) 1–10
DOI: 10.47852/bonviewSWT62029041

BON VIEW PUBLISHING

An Explainable Prediction Model for Cerebral Small Vessel Disease Based on Motor Function and Feature Interactions

Benben Wang[1], Yuefei Yan[1,2,*], Huizhen Lu[3], Baoqing Han[1], Yuangen Deng[1,2], Chengliang Zhang[3], Jianwei Xu[4], Jingtan Chen[1,2], Chuanliu Wang[3,*] and Congsi Wang[1,2,*]

[1] *State Key Laboratory of Electromechanical Integrated Manufacturing of High-performance Electronic Equipments, Xidian University, China*

[2] *Guangzhou Institute of Technology, Xidian University, China*

[3] *The Quzhou Affiliated Hospital of Wenzhou Medical University, China*

[4] *Quzhou Hospital of Traditional Chinese Medicine, China*

Abstract: Cerebral small vessel disease (CSVD) is a major condition with systemic and whole-brain effects, significantly impacting the daily lives of middle-aged and elderly people. Currently, predictions for CSVD are limited to traditional statistical methods and lack consideration of interactions between risk factors. Based on this, this study proposes an interpretable risk prediction method for CSVD based on combined motor function and interaction features. First, the proposed multidimensional feature selection method is applied for feature selection. Next, the RS+GWO+Voting method is proposed for model construction. Finally, the model constructed using SHapley Additive exPlanations (SHAP) is applied for interpretation, enhancing the clinical applicability of the model. The results show that the proposed model achieved a training set Area Under the Curve (AUC) of 0.829 and a validation set AUC of 0.824, demonstrating its strong predictive performance. The method provides theoretical support and technical backing for CSVD risk prediction under small-sample conditions and holds promise for integration with wearable devices to enable early warning and health management of CSVD.

Keywords: wearable applications, cerebral small vessel disease, motion function, explainable model, interaction features

1. Introduction

Cerebral small vessel disease (CSVD) has shown a continuous rising trend in recent decades. As a relatively common chronic cerebrovascular disease in the elderly, CSVD is closely associated with various neurological dysfunctions. Previous epidemiological studies have indicated that approximately 20–40% of cases of dementia in older adults are related to CSVD [1]. Some community-based studies from low- and middle-income countries have shown that among populations undergoing imaging, about 20% display typical imaging features of CSVD [2], resulting in significant social impact. CSVD refers to a series of clinical practice, imaging, and pathological syndromes caused by various factors affecting the small arteries in the brain and their distal branches, arterioles, capillaries, venules, and small veins. Clinical manifestations include cognitive impairment, emotional disturbances, and abnormal gait [3]. The clinical presentation of CSVD is highly heterogeneous, divided into acute ischemic CSVD and chronic insidious-onset clinical syndromes. Acute ischemic CSVD presents as specific lacunar syndromes, while chronic CSVD may be asymptomatic and is mostly diagnosed through imaging. As the burden of CSVD gradually increases, patients may develop symptoms such as cognitive impairment, motor disorders, and emotional disturbances.

CSVD has multiple causes, has unclear pathogenesis, and lacks clinically relevant animal models. Moreover, current imaging technologies cannot directly observe the brain's tiny blood vessels; they can only detect brain damage caused by CSVD through MRI, resulting in many patients missing early intervention opportunities [4]. Therefore, a rapid and reliable method for predicting CSVD is of great importance for early risk warning and personalized health management.

***Corresponding author:** Yuefei Yan, State Key Laboratory of Electromechanical Integrated Manufacturing of High-performance Electronic Equipments, Xidian University and Guangzhou Institute of Technology, Xidian University, China. Email: yuefeiyan@xidian.edu.cn, Chuanliu Wang, The Quzhou Affiliated Hospital of Wenzhou Medical University, China. Email: dzzhb@qzhospital.com and Congsi Wang, State Key Laboratory of Electromechanical Integrated Manufacturing of High-performance Electronic Equipments, Xidian University and Guangzhou Institute of Technology, Xidian University, China. Email: congsiwang@xidian.edu.cn

© The Author(s) 2026. Published by BON VIEW PUBLISHING. This is an open access article under the CC BY License (https://creativecommons.org/licenses/by/4.0/).

At the same time, the need for accurate and rapid CSVD assessment poses a significant challenge to clinical practice. Within a limited timeframe, doctors need to integrate and interpret a large number of physiological indicators and longitudinal follow-up data, which consumes a great deal of energy. Clinicians need reliable new technologies to reduce their workload. In this context, artificial intelligence (AI)-based data mining and risk prediction technologies have attracted increasing attention, with the potential to improve diagnostic efficiency and ensure the rational allocation of medical resources [5].

In recent years, some progress has been made in CSVD prediction based on electronic medical record data or imaging data. Cuadrado-Godia and colleagues systematically reviewed different types of CSVD and summarized the latest advances in the application of machine learning (ML) methods in this field [6]. Zhang et al. [7] developed a CSVD risk prediction model based on clinically relevant risk factors and further used a nomogram to facilitate clinical diagnosis and treatment decisions. In addition, Zhu et al. [8] constructed and validated a prognostic model to assess the severity of CSVD, which also utilized a nomogram-based approach and quantified disease severity through a comprehensive scoring system. Nomogram-based models are intuitive for clinical interpretation but are inherently tied to linear predictor assumptions and therefore less adaptable than many ML methods at capturing complex nonlinear relationships and interactions in high-dimensional or richly structured data. Furthermore, a recent meta-analysis systematically summarized ML models based on vascular imaging markers, reporting an overall C-index of approximately 0.84 for predicting cognitive impairment in CSVD patients [9], and this body of work primarily reflects aggregated performance for imaging-based models rather than clinical risk factor–based prediction.

Research in the field of CSVD comorbidities is also informative. Ali F integrated wearable sensor data with electronic medical records to develop a multi-stage heart disease prediction model [10]; Noah Hollmann and colleagues proposed the TabPFN method for small-sample data, combining a generative transformer with SHAP for prediction and interpretation [11]; Jia You employed ML to select predictors from a comprehensive variable space and develop a cardiovascular disease risk model [12]; and Qananwah applied ensemble algorithms for arrhythmia prediction [13].

In the field of CSVD prediction, on the one hand, existing methods mainly rely on imaging assessment results and lack risk prediction of CSVD based on motor function. Previous studies have shown that CSVD is closely associated with motor dysfunction. Sharma et al. [14] found that greater CSVD severity, especially increased white matter hyperintensity volume, was linked to slower gait, impaired performance, and higher fall risk. Su et al. [15] summarized that CSVD lesions disrupt motor-related neural networks, leading to gait disturbances and postural instability. Chen et al. [16] reported that CSVD exacerbates motor dysfunction in Parkinson's disease patients. These findings suggest that motor function measurements provide complementary information to traditional imaging markers in CSVD risk assessment.

On the other hand, existing methods mainly focus on the association between individual risk factors and CSVD, often overlooking the interactions between risk factors. Meanwhile, most existing ML prediction models primarily focus on conventional clinical and imaging features, without systematically incorporating motor function indicators or interaction terms among predictors. Although interpretable ML models developed from routine clinical data have been applied to CSVD risk assessment, to our knowledge, no prior study has explicitly combined motor function measurements with interaction features derived from baseline variables for risk prediction, which constitutes one of the motivations for the present study.

Disease risk assessment based on data collected from wearable technology has brought new surprises [17]. Wearable devices can collect vital signs and gait information in nonclinical settings, greatly facilitating the risk assessment and daily management of some chronic diseases. However, practical applications are still limited by issues such as data quality, privacy protection, and integration with clinical workflows [18]. Notably, cardiovascular signals, activity levels, and gait performance collected by wearable devices are closely related to vascular health and neurological function, which are also considered key factors in the development and progression of cerebrovascular diseases, particularly CSVD. In the cardiovascular field, studies have shown that wearable devices enable continuous cardiovascular monitoring [19], that activity trackers can detect conditions such as atrial fibrillation and COVID-19 [20], and that deep learning models based on continuous Photoplethysmography (PPG) measurements can provide longitudinal predictions [21]. In the cerebrovascular field, wearable-based stroke prediction models have achieved accuracies of up to 0.92 [22], while AI-driven continuous monitoring supports personalized stroke risk assessment [23], and wearable-based models can also predict bed rest duration, enabling a cost-effective and privacy-conscious approach [24].

Cardiovascular events, cerebrovascular diseases, and CSVD share overlapping risk factors and vascular mechanisms, making findings from related studies valuable for CSVD risk assessment. Some studies have explored CSVD-related tasks using wearable sensors and ML, such as a fall risk prediction model integrating gait, electromyography, and baseline data, which achieved an AUC of 0.986 on an independent dataset [25], and a gait pattern analysis combining gait parameters with CSVD imaging burden [26]. However, these studies mainly focus on single tasks and lack comprehensive CSVD risk prediction models systematically integrating wearable-derived motor function indicators with traditional clinical variables and multidimensional interaction features. The present study attempts to incorporate motor function information derived from wearable devices into a predictive model to explore its potential value in CSVD risk assessment.

Wearable devices can continuously and noninvasively acquire users' physiological and activity data. Compared with intermittent assessments in hospital clinical diagnoses, disease prediction based on wearable devices helps identify changes in risk factors in daily life and provides timely warnings. This study proposes a task-driven and interpretable framework for CSVD risk prediction. By integrating motor function features with expert-guided interaction modeling, the proposed approach effectively captures clinically meaningful complex relationships among risk factors. In addition, by incorporating a hybrid feature selection strategy and a multi-stage optimization and ensemble mechanism, the framework alleviates local optima issues and improves model generalization under high-dimensional, heterogeneous, small-sample, and complex parameter space conditions, thereby achieving a synergistic improvement in predictive performance and clinical interpretability.

2. Materials and Methods

2.1. Patient selection and data collection

In this study, data were collected from 419 patients with CSVD and 198 control subjects at Quzhou Hospital affiliated

with Wenzhou Medical University, covering the period from July 2022 to July 2024. CSVD data included laboratory test results, imaging examinations, and neuropsychological assessments, among other outcomes. MRI imaging results and comprehensive symptom assessments were used as the standard for CSVD diagnosis by physicians. The main variables of interest in the study included Age, TUG_abnormality, Years_of_schooling, SBP(MAX), Gender, BMI, DBP(MAX), Diabetes, Smoking, Drinking, and Atrial_fibrillation, Chronic Obstructive Pulmonary Disease (COPD), Obstructive Sleep Apnea (OSA). All data were screened according to professional inclusion and exclusion criteria.

2.2. Feature selection via multi-method consensus

Before screening for key risk factors, interaction terms of risk factors were constructed based on expert knowledge. During this process, particular attention was paid to the interaction between SBP(MAX) and exercise function and related risk factors. Specifically, seven interaction terms were created: SBP(MAX)_×_BMI, SBP(MAX)_×_Age, TUG_abnormality_×_Age, TUG_abnormality_×_Smoking, TUG_abnormality_×_Diabetes, SBP (MAX)_×_Diabetes, and BMI_×_Diabetes. These interaction features, combined with the original individual risk factors, form the candidate feature set for subsequent selection.

A feature selection method combining multiple ML techniques was proposed to identify key risk factors for CSVD and eliminate features with weak correlation to the disease. Four methods were used: Boruta, Lasso regression, XGBoost, and LightGBM. Each method was independently screened for key risk factors. The methods are described as follows:

Formally, let the candidate feature set be

$$R = \{\mathrm{r}_1, r_2, \ldots, r_p\}$$

where p denotes the total number of features. Denote the four selection methods as:

$$M_1 = Boruta, M_2 = Lasso, M_3 = XGBoost, M_4 = LightGBM$$

For each method M_k (k=1, 2, 3, 4), define the selection indicator for feature r_i as

$$S_{ik} = \begin{cases} 1, \text{ if } \mathrm{r}_i \text{ is selected by } method\, \mathrm{M_k} \\ 0, \text{ otherwise} \end{cases}$$

The consensus score for feature r_i is then

$$B_i = \sum_{k=1}^{4} S_{ik},\ i = 1, 2, ..., p$$

Finally, the selected feature set is defined as

$$R_{selected} = \{r_i \in R : B_i^{\tau} \geq \tau_{\omega}\}$$

Due to differences in model assumptions, regularization strategies, and nonlinear capturing capabilities, feature selection using a single model can be somewhat one-sided. Applying the proposed feature selection method that integrates multiple ML approaches to identify key features ensures that redundancy is reduced while critical omissions are avoided, which is a prerequisite for ensuring the performance of subsequent model construction.

2.3. Model development and optimization

As illustrated in Figure 1, the framework for developing the CSVD risk prediction model is presented. The dataset was randomly divided into training and test sets in a 7:3 ratio. We first constructed various ML models, including logistic regression (LR), multilayer perceptron (MLP), random forest (RF), and several boosting algorithms (gradient boosting decision trees [GBDT], AdaBoost, XGBoost, and CatBoost), and optimized their hyperparameters using random grid search. By comparing model performance and complementarity, a suitable model for predicting CSVD was selected. The gray wolf optimizer (GWO) was further applied to fine-tune the selected model, followed by a voting ensemble strategy. Multiple metrics, including accuracy, precision, recall, and F1-score, were used to evaluate model performance. Nested cross-validation was applied to the training set for model evaluation and hyperparameter optimization, and the test set was used for final performance assessment.

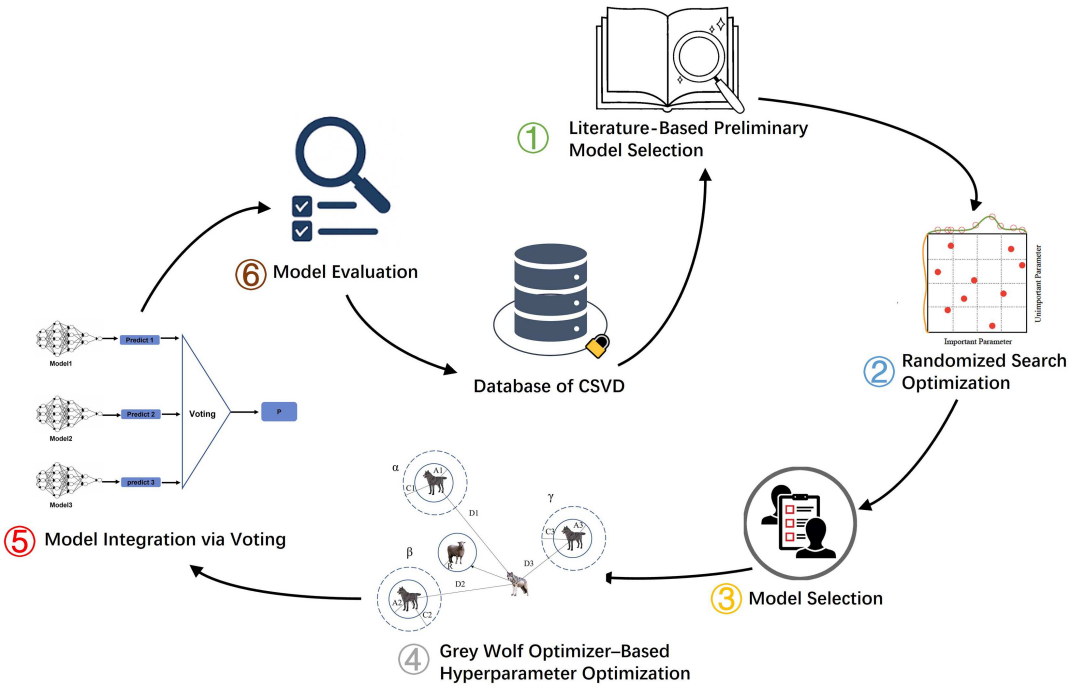
Random search (RS) enables relatively broad exploration of the hyperparameter space; however, under a limited number of iterations, it does not guarantee the identification of a near-optimal solution. Therefore, in this study, RS was first employed as an initial global exploration step to identify potentially promising hyperparameter regions and candidate base models. On this basis, the GWO algorithm was subsequently introduced to conduct further fine-grained search within the preselected parameter regions. Compared with purely random sampling, GWO incorporates a population-based cooperative mechanism with adaptive exploration and local exploitation capabilities, which can improve hyperparameter optimization efficiency and yield more robust performance outcomes. Finally, a voting ensemble strategy was adopted to integrate multiple well-performing models, thereby reducing model variance and enhancing generalization ability. In scenarios with limited sample sizes, where model performance is more susceptible to data partitioning, this staged optimization framework, combined with ensemble learning, helps improve overall model performance and reliability.

2.3.1. Model selection

By reviewing a large number of literature and monographs, various ML algorithms suitable for high-dimensional small-sample prediction tasks were selected, including LR, MLP, RF, GBDT, AdaBoost, XGBoost, and CatBoost. LR is simple and easy to understand, relatively easy to implement, and highly interpretable, performs well on small-sample data, and has been widely used in disease risk assessment and prediction in medical research. MLP can handle nonlinear problems and high-dimensional data, can deal with multi-class and regression problems, and has good generalization ability. RF can handle high-dimensional data without feature selection, identify correlations between features, has high prediction accuracy, and runs quickly. XGBoost has strong fitting ability, efficient learning algorithms, and the capability to handle high-dimensional sparse data. GBDT can handle nonlinear problems, manage missing and abnormal values, and is easy to understand and interpret. AdaBoost is easy to implement, highly adaptive, not prone to overfitting, and can maintain good generalization performance. CatBoost has unique advantages in handling categorical features, high computational efficiency, and good generalization.

RS was applied to optimize the hyperparameters of each individual model to ensure that each model is selected under relatively optimal parameters. Considering both the AUC and accuracy metrics, the models were selected.

Figure 1
Framework for CSVD risk prediction model development

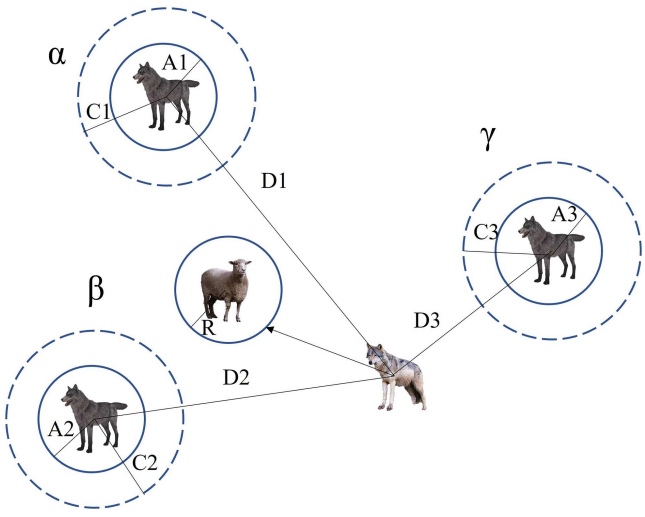


2.3.2. Model optimization

Use the GWO algorithm to fine-tune the hyperparameters of the selected model. As shown in Figure 2, the GWO algorithm operates according to the following principle. The GWO algorithm is inspired by the hunting behavior and social hierarchy of gray wolves in nature, efficiently searching for the optimal parameter combination by simulating the roles of alpha, beta, delta, and omega wolves.

A voting method is used to integrate the predictions of multiple base classifiers, aiming to combine the strengths of multiple models and improve model performance. This approach can effectively reduce overfitting and enhance the model's generalization ability under different sample conditions. Furthermore, using the voting method to integrate the predictions of multiple base classifiers aims to improve overall classification accuracy while reducing the risk of overfitting or underfitting associated with a single model.

Figure 2
Gray wolf optimizer principle diagram



2.3.3. Model evaluation metrics

We use metrics such as accuracy, precision, recall, and F1-score to evaluate the performance of the model and classify patients with CSVD as the positive group and patients without CSVD as the control group. Accuracy reflects the proportion of correct predictions among all samples; precision indicates the proportion of true positives among all samples predicted as positive, and recall measures the proportion of actual positive samples correctly identified by the model. The F1-score serves as the harmonic mean of precision and recall, balancing the two, and is particularly suitable for evaluating model performance in cases of class imbalance.

2.4. Interpretation of model

To enhance the applicability of the CSVD prediction model in clinical practice, this study further integrates SHAP swarm plots to visualize the associations between various risk factors and CSVD outcomes. Through swarm plots, we can intuitively identify the key features that contribute most to the prediction results and, to some extent, reveal the relationship between these features and CSVD risk. In this study, single models use swarm plots to quantify and rank the relative importance of each feature in the model's predictions, while the voting method explains based on the importance ranking. Model interpretation helps to validate the rationality of the prediction results and can also enhance the credibility of this model in the clinical context of CSVD.

3. Results

3.1. Baseline data analysis

A total of 617 participants were included in this study, comprising 198 individuals without CSVD and 419 with CSVD. Age was significantly higher in the CSVD group compared with the non-CSVD group (69.99 ± 10.52 vs 62.17 ± 11.38 years, $p < 0.001$). The statistically significant features ($p < 0.05$) obtained from the baseline analysis are as follows: age, gender, DBP(MAX), SBP(MAX), years of schooling, smoking, drinking, TUG_abnormality, and diabetes. Overall, patients with CSVD are generally older, have lower educational levels and higher blood pressure, and also show higher rates of diabetes, smoking, drinking, and abnormal TUG results. Detailed clinical baseline characteristics are summarized in Tables 1 and 2.

3.2. Key feature selection

Using the proposed integrated feature selection method combining multiple ML techniques, 14 key features were selected from the initial 23 features to participate in constructing the CSVD prediction model. The selected features, as shown in Figure 3, include age, SBP(MAX), SBP(MAX)_×_Diabetes, years of schooling, smoking, TUG_abnormality_×_SBP(MAX), DBP(MAX), atrial fibrillation, SBP(MAX)_×_Age, BMI_×_Diabetes, coronary heart disease (CHD), BMI, and SBP(MAX) _×_BMI. In the figure, the bar chart represents the number of times the features were selected by various methods. Through feature selection and modeling analysis, we found that the proposed method requires only 10 original baseline features to achieve high-accuracy CSVD prediction. Among them, age and SBP(MAX) were identified as the most important features, contributing most to the model's

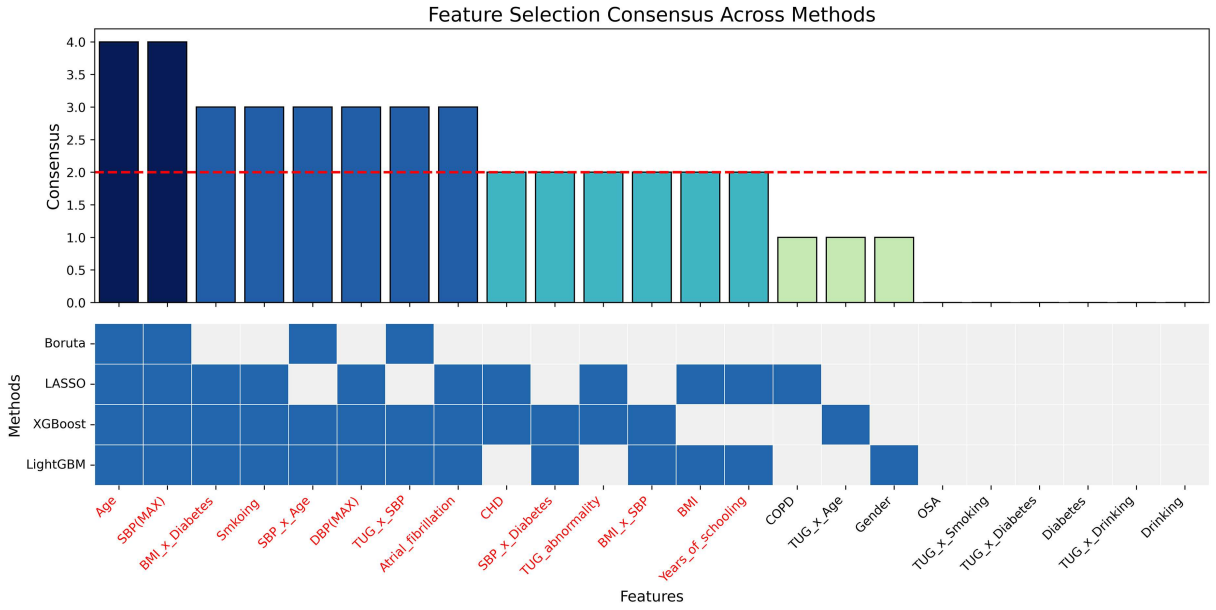
Table 1
Baseline distribution of continuous variables among participants

Clinical factors	No CSVD(n = 198)	With CSVD(n = 419)	p value
Age	62.17 ± 11.38	69.99 ± 10.52	<0.001
Years of schooling	7.98 ± 4.44	6.25 ± 4.42	<0.001
BMI	24.03 ± 3.32	23.59 ± 3.38	0.129
SBP(MAX)	147.16 ± 17.59	161.45 ± 17.73	<0.001
DBP(MAX)	80.52 ± 10.73	83.68 ± 11.37	0.001

Table 2
Baseline distribution of categorical variables among participants

Clinical factors	No CSVD(n = 198)	With CSVD(n = 419)	p value
Gender			0.038
Male	96(48.48%)	242(57.76%)	
Female	102(51.52%)	177(42.24%)	
Smoking			0.005
No	159(80.3%)	290(69.21%)	
Yes	39(19.7%)	129(30.79%)	
Drinking			0.023
No	164(82.83%)	311(74.22%)	
Yes	34(17.17%)	108(25.78%)	
Diabetes			<0.001
No	165(83.33%)	279(66.59%)	
Yes	33(5.56%)	140(33.41%)	
OSA			0.307
No	194(97.98%)	416(99.28%)	
Yes	4(2.02%)	3(0.72%)	
COPD			0.816
No	196(98.99%)	417(99.52%)	
Yes	2(1.01%)	2(0.48%)	
Atrial_fibrillation			0.111
No	179(90.40%)	395(94.27%)	
Yes	19(9.60%)	24(5.73%)	
TUG_abnormality			<0.001
No	160(80.81%)	160(38.19%)	
Yes	38(19.19%)	259(61.81%)	

Figure 3
Feature selection of CSVD prediction



prediction, representing the first-tier features. The second-tier features include DBP(MAX), smoking, and TUG abnormality, which also have a significant impact on model output. The third-tier features include CHD, SBP(MAX)_×_Diabetes, TUG_abnormality, BMI, BMI_×_SBP(MAX), and years of schooling, which have a relatively smaller influence on prediction but can still be included as key factors. COPD and gender form the fourth-tier features, with very limited roles in the model. Finally, alcohol consumption and diabetes are classified as fifth-tier features, as none of the four modeling methods used in this study identified them as key predictive features.

3.3. Classifiers' performance

We applied an RS strategy to optimize seven single models, with prediction performance shown in Figure 4. In terms of AUC, the model performance ranking is RF, LR, CatBoost, GBDT, XGBoost, AdaBoost, and MLP. Among these, CatBoost, GBDT, and XGBoost have similar performance, whereas AdaBoost and MLP have relatively weaker discriminative capabilities.

In terms of accuracy, the models are ranked as follows: LR, RF, XGBoost, MLP, CatBoost, GBDT, and AdaBoost. It is noteworthy that RF, XGBoost, and MLP have comparable performance, whereas GBDT and AdaBoost perform relatively poorly. Considering both AUC and accuracy, RF, LR, and XGBoost demonstrate the best balance between discriminative ability and classification stability. Therefore, these three models were selected to construct an ensemble learning model.

The performance comparison of RF, LR, and XGBoost models before and after optimization using GWO is shown in Figure 5. For the XGBoost model, both AUC and F1-scores improved significantly after optimization. However, accuracy also

Figure 4
Results for the seven single ML models

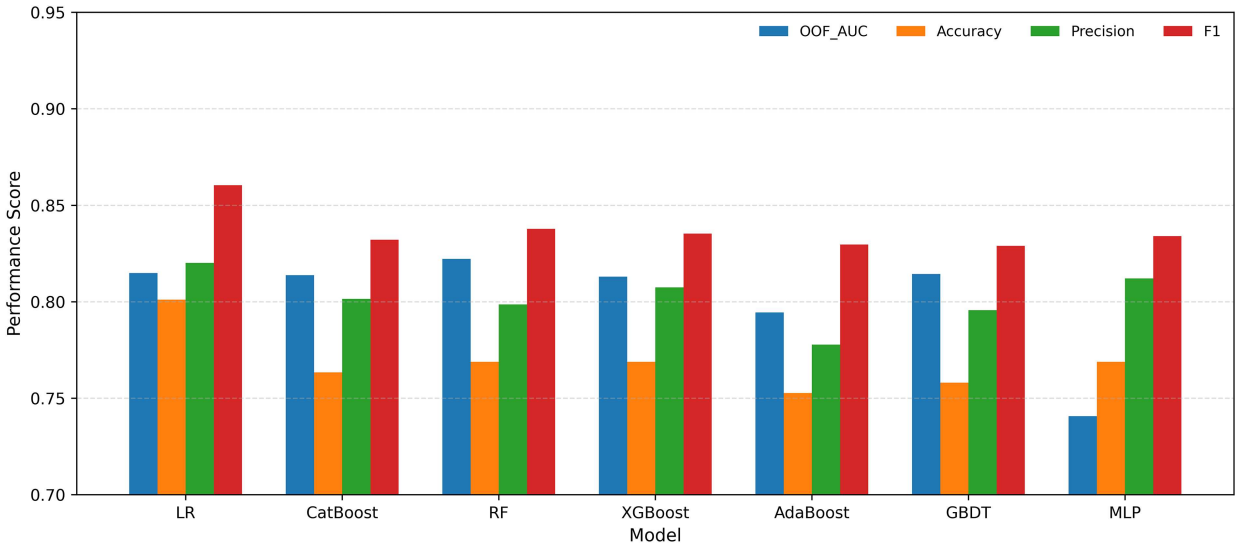


Figure 5
Results for the single ML models

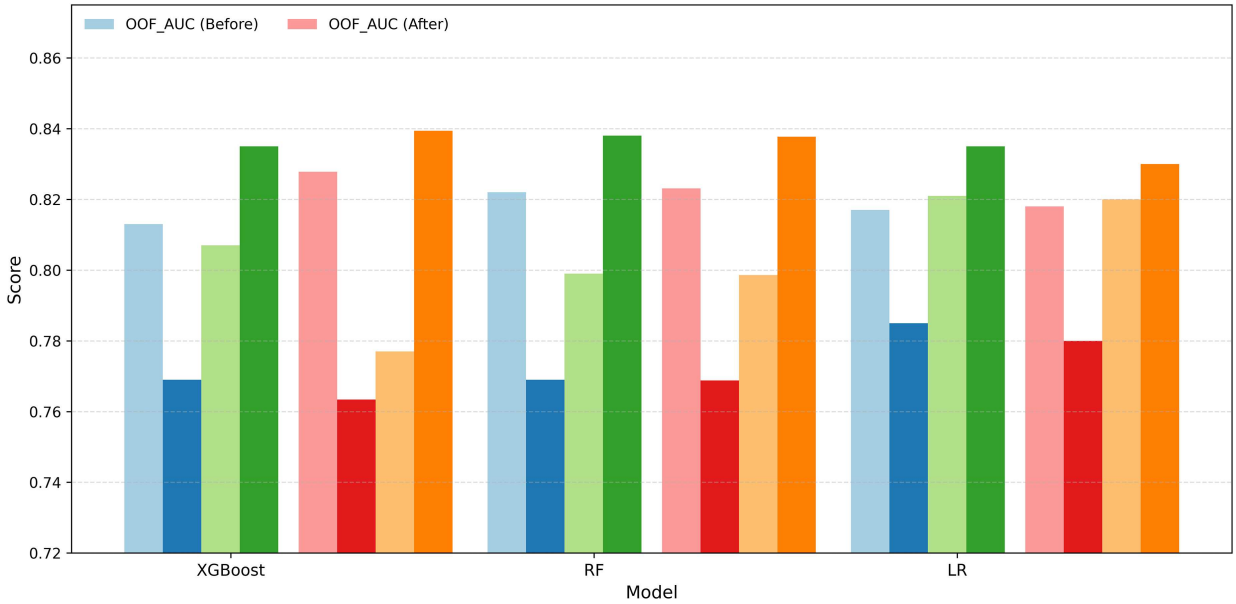


Table 3
Comparison of model performance before and after ensemble learning

	OOF-AUC	Test-AUC	Precision	Accuracy	F1-score
LR	0.817	0.833	0.866	0.779	0.830
RF	0.823	0.812	0.799	0.769	0.838
XGBoost	0.828	0.808	0.777	0.763	0.839
Proposed method	0.829	0.824	0.804	0.774	0.841

decreased noticeably. This phenomenon aligns with the requirement for maximum sensitivity in screening for cerebrovascular small vessel disease risk. GWO had little effect on improving the performance of the RF and LR models, indicating that they were already near their optimal performance during the initial RS phase. XGBoost achieved the highest training set AUC (0.828), while the LR model had the highest test set AUC (0.833). A summary of performance comparisons before and after the ensemble learning methods is shown in Table 3. It can be seen that the proposed voting ensemble method performed well, with an Out-of-Fold Area Under the Curve (OOF-AUC) of 0.829 and a Test-AUC of 0.824, demonstrating higher predictive performance and generalization capability.

3.4. Motor function contribution/ablation analysis

To further evaluate the contribution of motor function features, an ablation analysis was conducted by comparing model performance with and without the TUG feature. In this analysis, Model A represents the model constructed without the TUG feature. As shown in Table 4, the inclusion of motor function features improved the predictive performance of the model.

3.5. Interpretation of model

The visualization results of the model explanation are shown in Figures 6 and 7. Among them, LR, RF, and XGB models are presented as SHAP bee swarm plots, while the voting ensemble model is interpreted using feature importance plots. From the figure, it can be seen that the key features and their contributions to the prediction results are generally consistent across different models. Taking XGBoost as an example, features such as TUG_abnormality _×_SBP(MAX), SBP(MAX)_×_Age, TUG_abnormality, Age, SBP(MAX), SBP(MAX)_×_Diabetes, DBP(MAX), BMI_×_Diabetes, BMI_×_SBP(MAX), and smoking exhibit predominantly positive SHAP values, indicating that they tend to push the predicted probability toward a higher CSVD risk within the model framework. In contrast, the SHAP values of BMI, years of education, and CHD are relatively small, suggesting a comparatively smaller contribution to the model output in most instances.

Table 4
Impact of motor function features on model performance

Model	OOF-AUC	Test-AUC	Precision	Accuracy	F1-score
Model A	0.773	0.789	0.790	0.768	0.840
Model B	0.829	0.824	0.804	0.774	0.841

Figure 6
Model interpretability visualizations. Panels (a), (b), and (c) represent the SHAP plots of the selected LR, RF, and XGBoost models, respectively

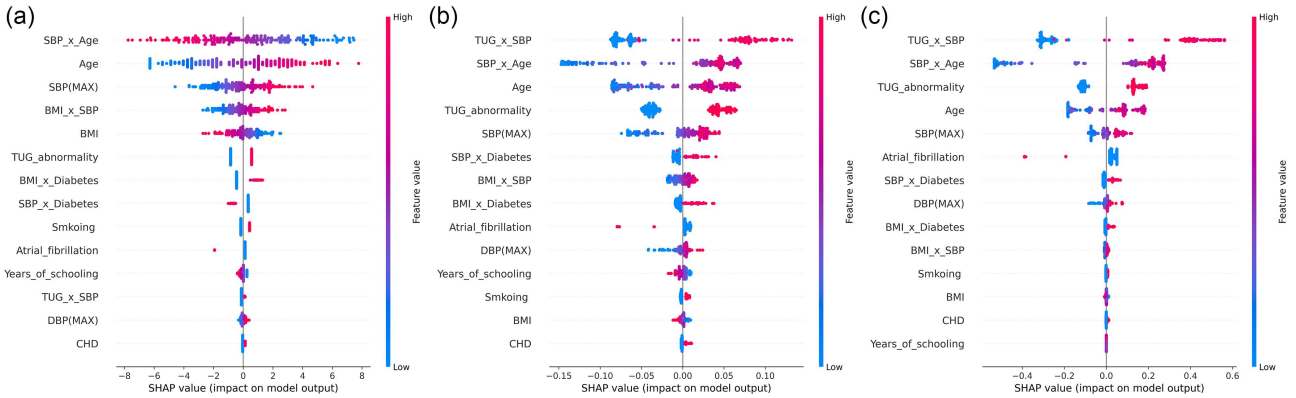
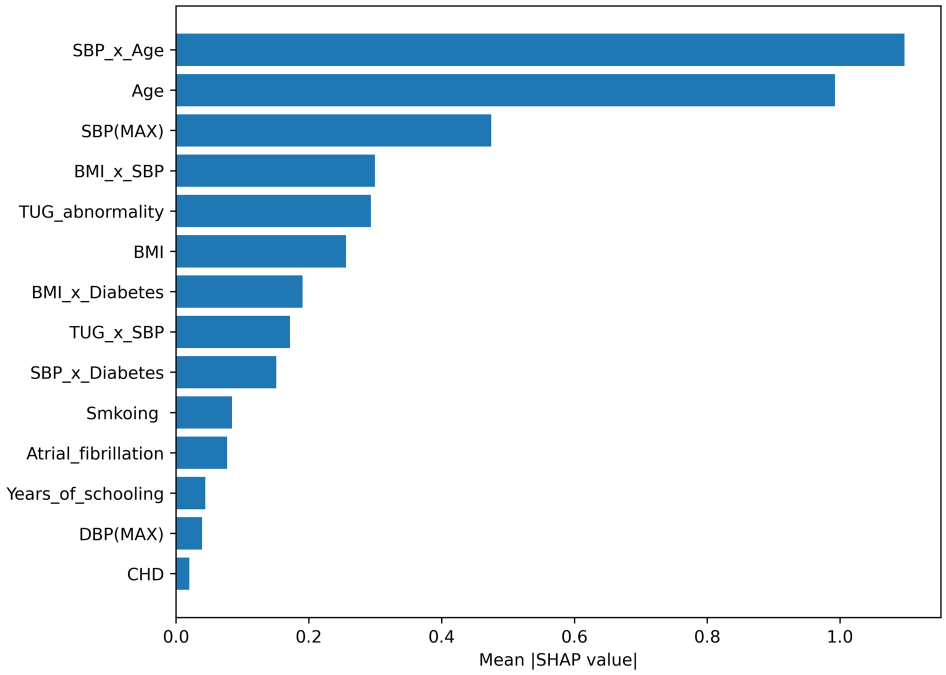


Figure 7
Voting ensemble feature importance



4. Discussion

This paper proposes a wearable-oriented, interpretable risk prediction model for CSVD. With only 10 independent features and a small number of interactions, the model achieved an AUC of 0.829 on the training set and 0.824 on the test set, enabling stable prediction of CSVD using noninvasive features. After introducing interaction features into the model, it was found that the impact of interaction features on CSVD was greater than that of individual features, suggesting a potential coupling amplification effect among risk factors of CSVD. Among these, Age × SBP(MAX) shows the highest contribution, highlighting the synergistic effect of aging and hypertension. After incorporating physical function, it was observed that abnormalities in physical function were the second most important factor in the model interpretation, only after peak systolic blood pressure, highlighting the significant role of physical function in predicting CSVD. A detailed discussion is provided below.

This study utilizes 10 noninvasive indicators—age, blood pressure, BMI, smoking, TUG abnormality, atrial fibrillation, CHD, diabetes, and years of schooling—to enable risk assessment without the need for laboratory tests, demonstrating good accessibility and practicality. This framework provides a theoretical basis for wearable device applications, in which users can input baseline information and integrate it with real-time monitoring data to achieve early CSVD risk warning and health management.

Regarding interaction features, in addition to Age × SBP(MAX), BMI × SBP(MAX), BMI × Diabetes, and SBP(MAX) × Diabetes also exhibit high importance, further emphasizing the critical role of comprehensive management of blood pressure, blood glucose, and body weight in CSVD prevention. Moreover, the interaction between TUG abnormality and

SBP suggests that motor dysfunction may contribute to disease progression by influencing blood pressure variability.

This study developed a CSVD risk prediction model that integrates motor function indicators with multidimensional interaction features. Previous studies have shown that CSVD imaging burden is closely associated with slower gait, impaired balance, and increased fall risk; however, most have remained at the level of correlation analysis or single-feature evaluation, with few attempts to systematically incorporate motor function and interactions with other clinical or structural features into a comprehensive risk prediction model. Our results indicate that motor function variables retain independent contributions within a multivariable framework, providing incremental information partially non-overlapping with traditional indicators, which may reflect the integrative status of brain networks and offer complementary value for risk stratification. Compared with studies focusing mainly on single indicators, this study highlights the importance of multidimensional feature integration and demonstrates the synergistic role of motor function and interaction features in CSVD risk prediction, thereby expanding existing research paradigms and providing new insights for developing more comprehensive and interpretable predictive models.

From a clinical perspective, the proposed model, while still at the prototype stage, provides a noninvasive approach for early screening and risk stratification of CSVD using readily available clinical and motor function features. Its robust preliminary performance indicates potential utility for identifying high-risk individuals and guiding early intervention and longitudinal monitoring. Notably, the model is intended to complement, rather than replace, imaging-based diagnosis, particularly in resource-limited settings or large-scale screening scenarios. In addition, the use of explainable ML enhances model transparency and supports clinical decision-making. Looking forward, the incorporation of real-world wearable data, such as continuous gait and activity monitoring, could further refine predictive accuracy, enable dynamic risk tracking, and facilitate personalized health management, paving the way toward real-time, patient-specific CSVD risk assessment once the prototype undergoes further development and empirical validation.

This study also has several limitations. First, motor function assessment is limited to TUG_abnormality, lacking precise data. In the future, based on this study, monitoring factors such as stride length and step length could be added to refine motor function assessment. Second, this study is conducted as a single-center, small-sample investigation and has not undergone external validation, leaving its applicability across different regions and populations to be rigorously assessed. Future work should leverage multicenter data through joint modeling to systematically evaluate and enhance the model's generalizability across diverse populations and clinical settings. Finally, this study lacks real-time data from wearable devices, and monitoring of parameters such as heart rate and sleep quality is missing. Future work could gradually collect relevant data based on preliminary applications and even consider replacing maximum systolic pressure with blood pressure variability for monitoring, further improving the predictive accuracy for CSVD.

5. Conclusion

This study proposes a noninvasive, interpretable model for predicting CSVD. It is the first risk prediction of CSVD combining motor function and interaction features. First, we propose a key risk factor selection method based on the co-selection of multiple ML algorithms. Second, we apply the proposed RS+GWO+Voting method to construct the CSVD prediction model. Finally, we use the SHAP method to interpret the model. In the future, data collected from wearable devices, such as stride, step length, and sleep, will be further utilized to achieve precise prediction and daily intervention for CSVD.

Funding Support

This research was supported by the National Natural Science Foundation of China (Grant Nos. 52275268 and U25A20300), the China Postdoctoral Science Foundation (Grant No. 2024M762525), the Natural Science Basic Research Program of Shaanxi Province (Grant No. 2025JC-QYXQ-026), the "Tianchi Talent" Introduction Program, and Xidian University Specially Funded Project for Interdisciplinary Exploration (Grant No. TZJH2024025).

Ethical Statement

This retrospective study was approved by the Ethics Committee of the Quzhou Affiliated Hospital of Wenzhou Medical University ([2024] No. 040). Given the retrospective nature of the study and the use of anonymized data, the requirement for informed consent was waived by the ethics committee in accordance with national legislation and institutional requirements.

Conflicts of Interest

Congsi Wang is an editorial board member for *Smart Wearable Technology* and was not involved in the editorial review or the decision to publish this article. The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

Data are available from the corresponding author upon reasonable request.

Author Contribution Statement

Benben Wang: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Writing – original draft, Visualization. **Yuefei Yan:** Methodology, Writing – review & editing, Project administration. **Huizhen Lu:** Validation, Investigation, Data curation. **Baoqing Han:** Writing – review & editing, Project administration. **Yuangen Deng:** Software. **Chengliang Zhang:** Data curation. **Jianwei Xu:** Data curation. **Jingtan Chen:** Supervision. **Chuanliu Wang:** Resources, Supervision. **Congsi Wang:** Conceptualization, Formal analysis, Resources, Funding acquisition.

References

[1] S, Markus, H., & Joutel, A. (2025). The pathogenesis of cerebral small vessel disease and vascular cognitive impairment. *Physiological Reviews*, *105*(3), 1075–1171. https://doi.org/10.1152/physrev.00028.2024

[2] Lam, B. Y. K., Cai, Y., Akinyemi, R., Biessels, G. J., van den Brink, H., Chen, C., . . . , & Mok, V. C. T. (2023). The global burden of cerebral small vessel disease in low-and middle-income countries: A systematic review and meta-analysis. *International Journal of Stroke*, *18*(1), 15–27. https://doi.org/10.1177/17474930221137019

[3] Li, Q., Yang, Y., Reis, C., Tao, T., Li, W., Li, X., & Zhang, J. H. (2018). Cerebral small vessel disease. *Cell Transplantation*, *27*(12), 1711–1722. https://doi.org/10.1177/0963689718795148

[4] Pantoni, L. (2010). Cerebral small vessel disease: From pathogenesis and clinical characteristics to therapeutic challenges. *The Lancet Neurology*, *9*(7), 689–701. https://doi.org/10.1016/S1474-4422(10)70104-6

[5] Al Askar, A., Buchireddygari, D., Middi, B. V. S. R., Ravishankar, S., Namidis, I., Garcés, M., . . . , & Chamayi, L. S. (2025). The role of artificial intelligence and machine learning in the assessment, diagnosis, and prediction of cerebral small vessel disease. *Cureus*, *17*(9), e93376. https://doi.org/10.7759/cureus.93376

[6] Cuadrado-Godia, E., Dwivedi, P., Sharma, S., Santiago, A. O., Gonzalez, J. R., Balcells, M., . . . , & Suri, J. S. (2018). Cerebral small vessel disease: A review focusing on pathophysiology, biomarkers, and machine learning strategies. *Journal of Stroke*, *20*(3), 302–320. https://doi.org/10.5853/jos.2017.02922

[7] Zhang, L., Gao, F., Zhang, Y., Hu, P., Yao, Y., Zhang, Q., . . . , & Zhang, Y. (2022). Analysis of risk factors for the development of cognitive dysfunction in patients with cerebral small vessel disease and the construction of a predictive model. *Frontiers in Neurology*, *13*, 944205. https://doi.org/10.3389/fneur.2022.944205

[8] Zhu, F., Yao, J., Feng, M., & Sun, Z. (2024). Establishment and evaluation of a clinical prediction model for cognitive impairment in patients with cerebral small vessel disease. *BMC Neuroscience*, *25*(1), 35. https://doi.org/10.1186/s12868-024-00883-y

[9] Zheng, W., Wu, Q., Mu, R., Kuang, J., Yang, P., Lv, J., . . . , & Zhu, X. (2025). Development and validation of radiomics and deep transfer learning models to assess cognitive impairment in patients with cerebral small vessel disease. *Neuroscience*, *572*, 145–154. https://doi.org/10.1016/j.neuroscience.2025.03.012

[10] Ali, F., El-Sappagh, S., Islam, S. R., Kwak, D., Ali, A., Imran, M., & Kwak, K. S. (2020). A smart healthcare monitoring system for heart disease prediction based on ensemble deep learning and feature fusion. *Information Fusion*, *63*, 208–222. https://doi.org/10.1016/j.inffus.2020.06.008

[11] Hollmann, N., Müller, S., Purucker, L., Krishnakumar, A., Körfer, M., Hoo, S. B., . . . , & Hutter, F. (2025). Accurate predictions on small data with a tabular foundation model. *Nature*, *637*(8045), 319–326. https://doi.org/10.1038/s41586-024-08328-6

[12] You, J., Guo, Y., Kang, J. J., Wang, H. F., Yang, M., Feng, J. F., . . . , & Cheng, W. (2023). Development of machine learning-based models to predict 10-year risk of cardiovascular disease: A prospective cohort study. *Stroke and Vascular Neurology*, *8*(6). https://doi.org/10.1136/svn-2023-002332

[13] Qananwah, Q., Ababneh, M., & Dagamseh, A. (2024). Cardiac arrhythmias classification using photoplethysmography database. *Scientific Reports*, *14*(1), 3355. https://doi.org/10.1038/s41598-024-53142-9

[14] Sharma, B., Wang, M., McCreary, C. R., Camicioli, R., & Smith, E. E. (2023). Gait and falls in cerebral small vessel disease: a systematic review and meta-analysis. *Age and Ageing*, *52*(3), afad011. https://doi.org/10.1093/ageing/afad011

[15] Su, C., Yang, X., Wei, S., & Zhao, R. (2022). Association of cerebral small vessel disease with gait and balance disorders. *Frontiers in Aging Neuroscience*, *14*, 834496. https://doi.org/10.3389/fnagi.2022.834496

[16] Chen, H., Wan, H., Zhang, M., Liu, G., Wang, X., Wang, Z., . . . , & Wang, Y. (2021). Cerebral small vessel disease may worsen motor function, cognition, and mood in Parkinson's disease. *Parkinsonism & Related Disorders*, *83*, 86–92. https://doi.org/10.1016/j.parkreldis.2020.12.025

[17] Qadrie, Z., Ashraf, H., Maqbool, M., & Farooq, S. (2025). Wearable technology in healthcare: Opportunities, challenges, and future directions. *Smart Wearable Technology*, *1*, A12. https://doi.org/10.47852/bonviewSWT52025578

[18] Adepoju, V. A., Jamil, S., Biswas, M. S., & Chowdhury, A. A. (2025). Wearable technology in the management of chronic diseases: A growing concern. *Chronic Diseases and Translational Medicine*, *11*(02), 117–121. https://doi.org/10.1002/cdt3.156

[19] Xie, H., Yang, L., Jiang, B., Huang, Z., & Lin, Y. (2025). State-of-the-art wearable sensors for cardiovascular health: A review. *npj Cardiovascular Health*, *2*(1), 53. https://doi.org/10.1038/s44325-025-00090-6

[20] Singh, B., Chastin, S., Miatke, A., Curtis, R., Dumuid, D., Brinsley, J., . . . , & Maher, C. (2024). Real-world accuracy of wearable activity trackers for detecting medical conditions: Systematic review and meta-analysis. *JMIR mHealth and uHealth*, *12*, e56972. https://doi.org/10.2196/56972

[21] Nie, G., Zhao, Q., Tang, G., Li, Y., & Hong, S. (2025). Artificial intelligence-derived photoplethysmography age as a digital biomarker for cardiovascular health. *Communications Medicine*, *5*(1), 481. https://doi.org/10.1038/s43856-025-01188-9

[22] Lu, Y., Perumal, T., & Liu, C. (2024). Improved model for stroke prediction on wearable devices with CNN_LSTM_KAN. In *Proceedings of the 2024 the 12th International Conference on Information Technology (ICIT)*, 225–230. https://doi.org/10.1145/3718391.3718430

[23] Olawade, D. B., Aderinto, N., David-Olawade, A. C., Egbon, E., Adereni, T., Popoola, M. R., & Tiwari, R. (2025). Integrating AI-driven wearable devices and biometric data into stroke risk assessment: A review of opportunities and challenges. *Clinical Neurology and Neurosurgery*, *249*, 108689. https://doi.org/10.1016/j.clineuro.2024.108689

[24] Ogasawara, T., Mukaino, M., Matsunaga, K., Wada, Y., Suzuki, T., Aoshima, Y., . . . , & Tsukada, S. (2024). Prediction of stroke patients' bedroom-stay duration: Machine-learning approach using wearable sensor data. *Frontiers in Bioengineering and Biotechnology*, *11*, 1285945. https://doi.org/10.3389/fbioe.2023.1285945

[25] Zhou, Y., Zhang, D., Ji, Y., Bu, S., Hu, X., Zhao, C., . . . , & Li, L. (2025). Wearable sensors and machine learning fusion-based fall risk prediction in covert cerebral small vessel disease. *Frontiers in Neuroscience*, *19*, 1493988. https://doi.org/10.3389/fnins.2025.1493988

[26] Xu, K., Wang, Y., Jiang, Y., Wang, Y., Li, P., Lu, H., . . . , & Chen, X. (2024). Analysis of gait pattern related to high cerebral small vessel disease burden using quantitative gait data from wearable sensors. *Computer Methods and Programs in Biomedicine*, *250*, 108162. https://doi.org/10.1016/j.cmpb.2024.108162

How to Cite: Wang, B., Yan, Y., Lu, H., Han, B., Deng, Y., Zhang, C., …, & Wang, C. (2026). An Explainable Prediction Model for Cerebral Small Vessel Disease Based on Motor Function and Feature Interactions. *Smart Wearable Technology*. https://doi.org/10.47852/bonviewSWT62029041