The Journal of Clinical Hypertension

WILEY

ORIGINAL ARTICLE OPEN ACCESS

Research on Prediction model of Carotid-Femoral Pulse Wave Velocity: Based on Machine Learning Algorithm

Minghui Chen[1] | Jing Xiong[2] | Moran Li[2] | Tao Hu[3] | Yi Zhang[2]

[1]School of Health Science and Engineering, University of Shanghai for Science and Technology, Shanghai, China | [2]Department of Cardiology, Shanghai Tenth People's Hospital, Tongji University School of Medicine, Shanghai, China | [3]Department of Cardiology, Xijing Hospital, Fourth Military Medical University, Xi 'an, Shanxi, China

Correspondence: Tao Hu (dr_ht556@163.com) | Yi Zhang (yizshcn@gmail.com)

Received: 2 September 2024 | **Revised:** 1 February 2025 | **Accepted:** 4 February 2025

Funding: This study was financially supported by the National Nature Science Foundation of China (82170388, 82370300), Shanghai Technology Research Leader Program (21XD1434700), the Cardiac rehabilitation fund by the International Medical Exchange Foundation (Z-2019-42-1908-3), and Grant for the construction of Innovative Flagship Hospital for Integrated Traditional Chinese and Western Medicine (No., ZY(2021-2023)-0205-05).

Keywords: cardiovascular patients | cervical and femoral pulse velocity | machine learning algorithm

ABSTRACT

Carotid-femoral pulse wave velocity (cf-PWV) is an important but difficult to obtain measure of arterial stiffness and an independent predictor of cardiovascular events and all-cause mortality. The objective of this study was to develop a predictive model for cf-PWV based on brachial-ankle pulse wave velocity (baPWV) and other the accessible clinical parameters.

This model aims to allow patients to estimate their cf-PWV in advance without the need for direct measurement. We selected participants of the Northern Shanghai community from 2013 to 2022 as the study object. The Pearson correlation coefficient was employed for correlation analysis in feature selection. The linear regression models demonstrated low root mean square error (RMSE), error term (ε), and R^2 values, indicating good predictive performance. A Cox proportional hazards model revealed a significant association between machine learning-predicted cf-PWV and mortality risk, supporting the validity of prediction model. Using a threshold of cf-PWV greater than 10 m/s as the criterion, a classification prediction model was developed. Shapley Additive Explanations (SHAP) analysis was then applied to the Gradient Boosting model to elucidate the predictive mechanism of the optimal model. Without precise instruments, doctors often cannot determine a patient's cf-PWV. When the cf-PWV value predicted by the machine learning algorithm is high, patients can be recommended for more precise measurements to confirm the prediction and emphasize the importance of follow-up health management and psychological support. It is feasible to use a machine learning algorithm based on baPWV and other readily available clinical parameters to predict cf-PWV.

1 | Introduction

Cardiovascular disease (CVD) is the leading cause of death worldwide [1]. More than 40% of all disease-related deaths in China are caused by CVD [2]. The higher the aortic stiffness, the higher the probability of cardiovascular risk events [3], if not timely intervention, the development of advanced atherosclerotic disease, the difficulty of treatment will be greatly increased. Arterial stiffness is an important predictor of cardiovascular events and all-cause mortality [4–8]. By monitoring arterial stiffness, changes in vascular function can be detected early, and the risk of cardiovascular disease can be predicted, so that appropriate prevention and treatment measures can be taken. Carotid-femoral pulse wave velocity (cf-PWV) is a key indicator of arterial stiffness and an independent predictor of cardiovascular events and all-cause mortality [9, 10]. Therefore,

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2025 The Author(s). *The Journal of Clinical Hypertension* published by Wiley Periodicals LLC.

accurate measurement and prediction of cf-PWV is essential for early detection and management of CVD. Traditionally, cf-PWV measurement techniques often require specialized equipment and trained professionals [11, 12]. However, these methods require specialized personnel and are complex to operate. As a result, there is growing interest in developing more accessible and cost-effective alternative methods for predicting cf-PWV. Currently, machine learning algorithms have become powerful tools for predicting various health outcomes [13–15]. These algorithms can learn from data sets and identify complex patterns and relationships that may not be found by traditional statistical methods [16–18]. In recent years, some studies have explored the use of machine learning algorithms to predict cardiovascular disease risk, and have achieved effective results [19–21].

The aim of this study was to explore the feasibility of using machine learning algorithms to predict cf-PWV in elderly cardiovascular patients. The Chinese Guidelines for the Prevention and Treatment of Hypertension (revised 2024 edition) state that when the cutoff value of cf-PWV exceeds 10 m/s, it indicates a high cardiovascular risk [12]. We hypothesized that by combining clinical characteristics such as blood pressure, sex, and age [22] with machine learning techniques, we could accurately predict cf-PWV and use machine learning classification methods to identify individuals at high risk for CVD events. We compare several commonly used machine learning algorithms to evaluate the model's performance based on its predictive accuracy.

By demonstrating the feasibility and potential utility of machine learning algorithms to predict cf-PWV, our research could lead to the development of more accessible and cost-effective cf-PWV prediction tools.

2 | Materials and Methods

2.1 | Study Design and Study Population

The design and sample size of this study, as well as the eligibility criteria for participants, are based on the previously published protocol of the Northern Shanghai Study [23]. The study is a prospective, community-based ongoing study approved by the Institutional Review Committee of Shanghai 10th People's Hospital and conducted with financial support from the Shanghai Municipal Government (grant number: 2013ZYJB0902 and 15GWZK1002). The preliminary expected sample size is 3000 to 4,000 participants.

Inclusion criteria for participants included: age 65 years or older, voluntarily sign informed consent, a local resident of a community in northern Shanghai, and a willingness to undergo long-term follow-up. Exclusion criteria included a diagnosis of severe heart disease (NYHA classification ≥IV) or end-stage renal disease (CKD stage ≥4); having cancer or a life expectancy of less than 5 years; had a stroke within the past 3 months; reluctance to participate in clinical research; must withdraw from the trial due to other medical conditions; breach of research protocol; or lose contact with laboratory staff, resulting in no long-term follow-up.

Through the Northern Shanghai Study data, the feature variables are selectively selected according to the inspection indicators, and the cf-PWV prediction model is established and its influencing factors are evaluated through a round of feature screening. The construction of prediction model includes the following steps: (1) data selection; (2) data preprocessing; (3) feature selection; (4) model training; and (5) model selection.

2.2 | Data Preprocessing

During the data cleansing process, we eliminate duplicates from the dataset to ensure the uniqueness of patient information. Additionally, we handle missing data to maintain data integrity and standardize data types into appropriate formats, such as dates and numbers. Furthermore, we identify and rectify outliers to ensure data quality and accurate analysis. In the data conversion phase, we transform the character data into numeric data to meet the requirements of subsequent feature processing. For model training, we utilize 80% of the dataset collected from 2013 to 2022, setting aside the remaining 20% as a test set. This partitioning strategy facilitates more efficient data processing and analysis.

2.3 | Feature Selection

For the correlation analysis, the Pearson correlation coefficient was employed to quantify the linear relationship between two variables, which ranges from −1 to 1. A coefficient of 1 signifies a perfect positive correlation, −1 denotes a perfect negative correlation, and 0 implies no correlation.

2.4 | Model Construction and Evaluation

To predict the value of cf-PWV, we evaluated five machine learning methods: linear regression (LR), support vector regression (SVR), gradient boosting (GB), random forest (RF), and K-Nearest Neighbor (KNN). Cross-validation was employed to optimize model parameters. The root mean square error (RMSE) was defined as follows:

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}\left(\widehat{\mathrm{PWV}} - \mathrm{PWV_i}\right)^2}{n}}$$

where n represents the size of the test dataset; $\widehat{\mathrm{PWV}}$ and $\mathrm{PWV_i}$ are the ith estimated and measured PWV, respectively. Then, a percentage error, ϵ, was calculated based on the RMSE:

$$\varepsilon = \frac{\mathrm{RMSE}}{\overline{\mathrm{PWV}}}$$

and R^2 were used as evaluation indexes for model effects.

For the classification task, individuals with cf-PWV greater than 10 m/s were classified as the high-risk group, whereas those with cf-PWV less than 10 m/s were classified as the low-risk group. To perform this classification, we utilized various machine learning models, including LR, Support Vector Classifier (SVC), gradient boosting decision tree (GB), RF, KNN, and eXtreme Gradient Boosting (XGBoost). To evaluate the effectiveness of these models, we employed metrics such as accuracy, recall, F1

score, receiver operating characteristic (ROC) curve, and area under the curve (AUC).

2.5 | Cox Proportional Risk Model

Following the construction of the machine learning regression model for predicting cf-PWV, we employed IBM SPSS Statistics 27 software to perform Cox proportional hazards regression analysis. The objective of this analysis was to assess the association between the predicted cf-PWV values and actual cardiovascular events, as well as to evaluate their relationship with patient survival time.

In the Cox survival analysis, we utilized the predicted cf-PWV values derived from five distinct machine learning models (namely, SVR_cf-PWV, RF_cf-PWV, LR_cf-PWV, KNN_cf-PWV, and GB_cf-PWV) along with the actual brachial-ankle pulse velocity (baPWV) and cf-PWV values measured using instruments as covariates. MACE4 served as the event state indicator, while MACE4_FU_Duration_d represented the follow-up time in days. The primary objective of this analysis was to assess the predictive capability of the predicted cf-PWV values in estimating cardiovascular risk, and thereby to validate the accuracy and reliability of our prediction results.

2.6 | SHapley Additive exPlanations

SHAP is a machine learning model explanation method proposed by Lundberg et al. [24]. SHAP has the additive consistency of output results. For each prediction sample, the model has a predicted value. SHAP value is the value assigned to each feature in the sample, and SHAP value was first proposed by Shapley based on cooperative game theory [25].

The SHAP method provides a quantitative measure of the contribution of each feature in each sample to the model's output by computing the SHAP value. Unlike traditional linear models that solely evaluate parameter magnitude or sign, the SHAP value captures the influence of each feature in an individual sample, indicating both the direction (positive or negative) and magnitude of its effect. A negative SHAP value signifies that the feature has a detrimental impact on the predicted value, whereas a positive value indicates a beneficial effect. The absolute value of the SHAP value corresponds to the feature's influence on the outcome, with higher absolute values indicating greater impact.

By employing the SHAP method, we can ascertain the degree of contribution of each indicator in each sample, thereby revealing its importance within the model. When an indicator demonstrates a consistent trend across the majority of samples, the model recognizes it as having a significant positive or negative effect. Consequently, the SHAP method serves as a powerful tool for elucidating the specific impact of each feature within the model on the prediction results.

3 | Result

A dataset comprising 2709 patients was collected and prepared for model training, construction, and validation, based on data gathered from a study conducted in the northern region of Shanghai between 2013 and 2022. To ensure the validity and generalization capability of the model, the dataset was partitioned into a training set (80%) and a test set (20%). During the model training phase in Figure 1, the Pearson correlation coefficient was employed for correlation analysis, and baPWV, age, sex, rbsbp (right brachial systolic blood pressure), and rbdbp (right brachial diastolic blood pressure) were identified as input features that exhibited high correlation with cf-PWV.

TABLE 1 | The results of the machine learning models in the North Shanghai Study.

Model	RMSE (m/s)	R^2	ε
SVR	1.415	0.464	15.384%
RF	1.467	0.424	15.948%
LR	1.383	0.507	15.049%
KNN	1.544	0.433	16.547%
GB	1.415	0.500	15.168%

Note: This table lists the effectiveness of predicting cf-PWV values using five machine learning algorithms, among which the LR model outperforms the other models in all prediction metrics.

The results of the machine learning regression models are presented in Table 1. The SVR model achieved an RMSE value of 1.415, an R^2 value of 0.464, and a percentage error (ε) of 15.384%, indicating its stability in predictions and effective control of the difference between predicted and actual values. Among the evaluated models, LR performed optimally, with an RMSE of 1.383, an R^2 of 0.507, and the lowest percentage error of 15.049%. These results underscore LR's robust predictive and fitting abilities for this dataset, making it the preferred choice for predicting cf-PWV values. While KNN and RF models exhibited slightly lower R^2 values, their RMSE and percentage errors remained within acceptable ranges, suggesting that these models can maintain a degree of predictive accuracy for specific data types.

Table 2 presents the baseline characteristics of the study population. The analyses were conducted using R version 4.2.3, with the gtsummary package version 1.7.2 serving as the primary tool for data summarization. MACE4, a composite indicator, encompasses non-fatal myocardial infarction, nonfatal stroke, revascularization (including stent placement or coronary artery bypass grafting, CABG), and cardiovascular death. The value of MACE4 indicates whether these events have occurred (occurrence of MACE4) or not (nonoccurrence of MACE4). Additionally, MACE4_FU_Duration_d represents the follow-up time in days for MACE4 events.

In the Cox survival analysis, we incorporated the predicted cf-PWV values derived from five distinct machine learning models—SVR_cf-PWV, RF_cf-PWV, LR_cf-PWV, KNN_cf-PWV, and GB_cf-PWV—along with the actual cf-PWV and baPWV values measured using specialized instruments, as covariates. MACE4 served as the event state variable, while MACE4_FU_Duration_d represented the follow-up duration in days.

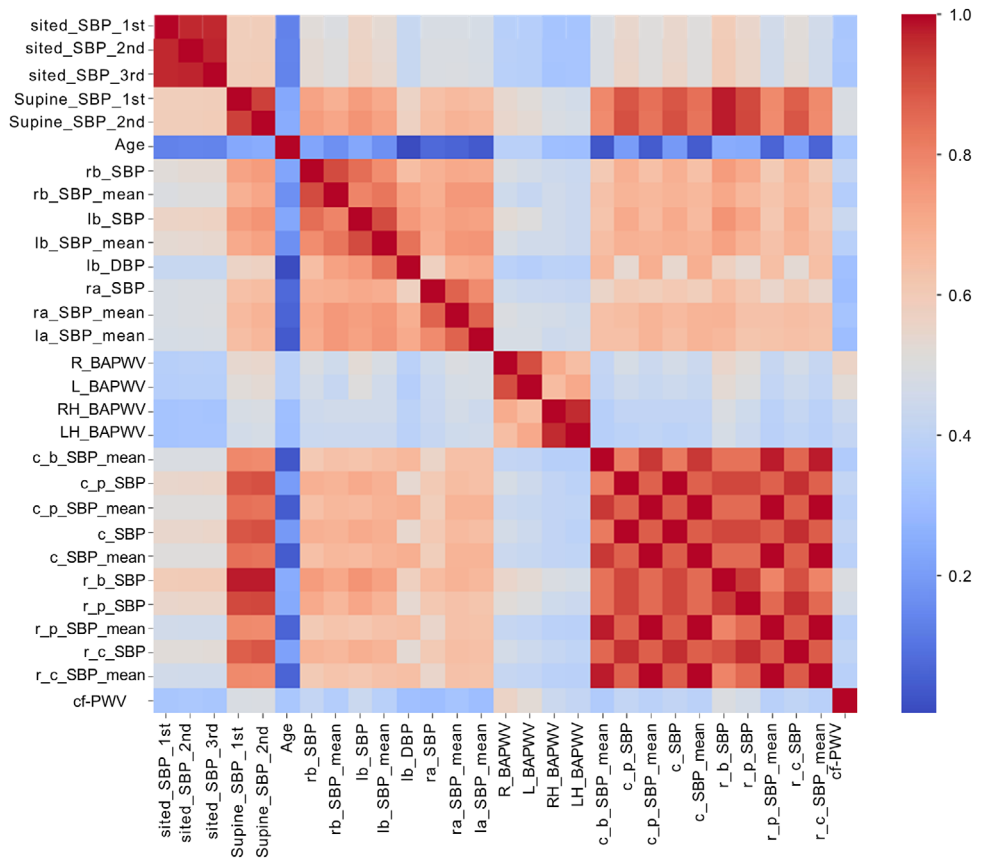

FIGURE 1 | Correlation heatmap of significant features with target. This image displays the correlation magnitude between various features, with values ranging from 0 to 1 represented by color, where the color becomes darker as the value increases. The features include age: Age; sited_SBP_1st: sited Systolic Blood Pressure at 1st time; sited_SBP_2nd: sited Systolic Blood Pressure at 2nd time; sited_SBP_3rd: sited Systolic Blood Pressure at 3rd time; Supine_SBP_1st: Supine Systolic Blood Pressure at 1st time; Supine_SBP_2nd: Supine Systolic Blood Pressure at 2nd time; rb_SBP: Right Brachial Systolic Blood Pressure; rb_SBP_mean: the mean value of 3 times Right Brachial Systolic Blood Pressure; lb_SBP: Left Brachial Systolic Blood Pressure; lb_SBP_mean: the mean value of 3 times Left Brachial Systolic Blood Pressure; lb_DBP: Left Brachial Diastolic Blood Pressure; ra_SBP: Right Arm Systolic Blood Pressure; ra_SBP_mean: the mean value of 3 times Right Arm Systolic Blood Pressure; la_SBP_mean: the mean value of 3 times Left Arm Systolic Blood Pressure; R_BAPWV: Right brachial-ankle pulse wave velocity; L_BAPWV: left brachial-ankle pulse wave velocity; RH_BAPWV: Right Heart to Brachial Artery Pulse Wave Velocity; LH_BAPWV: Left Heart to Brachial Artery Pulse Wave Velocity; CF_PWV: Carotid-Femoral Pulse Wave Velocity; c_b_SBP_mean: the mean value of 3 times central Beside Systolic Blood Pressure; c_p_SBP: central Position Systolic Blood Pressure; c_p_SBP_mean: the mean value of 3 times central Position Systolic Blood Pressure; c_SBP: central_SBP; c_mbp: the mean value of 3 times central_SBP_mean; r_b_SBP: Right Bedside Systolic Blood Pressure; r_p_SBP: Right Position Systolic Blood Pressure; r_p_SBP_mean: the mean value of 3 times Right Position Systolic Blood Pressure; r_c_SBP: Right Central Systolic Blood Pressure; r_c_SBP_mean: the mean value of 3 times Right Central Systolic Blood Pressure.

Table 3 presents the results of the survival analysis and outcome event prediction using the Cox proportional hazards model, incorporating both the predicted cf-PWV values from various machine learning models and the actual cf-PWV values. The table includes key statistical indicators such as -2 Log Likelihood, Chi-square, degrees of freedom (df), and the significance level (Sig.). A smaller -2 Log Likelihood value indicates a better fit of the model to the data. When the Chi-square value is large, it signifies a substantial difference between the actual observed values and the expected frequencies, indicating a high degree of deviation between the observed data and theoretical expectations. Conversely, a smaller sig value (typically < 0.05) suggests that a particular parameter is significant within the model, thereby making a notable contribution to its overall fit.

A univariate Cox regression analysis on the entire dataset revealed a statistically significant association between baPWV and survival time or survival risk ($p < 0.05$), further validating previous research that baPWV is a crucial factor in predicting survival risk. However, when analyzing the 20% validation set data used for the machine learning model, the association between baPWV and MACE (major adverse cardiovascular events) became non-significant ($p = 0.402$), suggesting that baPWV's predictive ability diminished in smaller datasets. In contrast, the machine learning models, trained on 80% of the data and tested on the 20% validation set, demonstrated a statistically significant relationship between the predicted cf-PWV values and MACE in all models ($p < 0.05$). This indicates that, with limited data (20%), the machine learning predicted cf-PWV can effectively predict MACE events.

TABLE 2 | The statistical indicators of the selected features.

Variable	Group			*p* value[b]
	Overall, *N* = 2709[a]	Nonoccurrence of MACE4, *N* = 2443[a]	Occurrence of MACE4, *N* = 266[a]	
Age				<0.001
Median (IQR)	70.0 (67.0, 74.0)	69.0 (67.0, 74.0)	71.0 (67.0, 77.0)	
Mean (SD)	71.2 (5.7)	71.0 (5.5)	73.0 (6.7)	
Range	65.0, 94.0	65.0, 94.0	65.0, 94.0	
cf-PWV				<0.001
Median (IQR)	8.9 (7.8, 10.4)	8.9 (7.8, 10.4)	9.4 (8.2, 10.8)	
Mean (SD)	9.2 (2.0)	9.2 (2.0)	9.6 (1.9)	
Range	4.3, 22.2	4.3, 22.2	5.1, 16.0	
baPWV				<0.001
Median (IQR)	1781.0 (1565.0, 2044.0)	1772.0 (1556.0, 2037.5)	1853.0 (1652.5, 2099.5)	
Mean (SD)	1824.8 (359.8)	1817.3 (358.2)	1893.5 (367.9)	
Range	960.0, 4517.0	960.0, 4517.0	1003.0, 3171.0	
rbsbp				<0.001
Median (IQR)	141.0 (128.0, 155.0)	140.0 (127.0, 154.0)	146.0 (133.0, 160.0)	
Mean (SD)	143.0 (21.3)	142.4 (21.1)	147.8 (22.6)	
Range	92.0, 241.0	92.0, 241.0	104.0, 226.0	
rbdbp				0.036
Median (IQR)	79.0 (72.0, 85.0)	79.0 (72.0, 85.0)	80.0 (73.0, 87.0)	
Mean (SD)	79.2 (10.9)	79.1 (10.8)	80.7 (11.5)	
Range	48.0, 138.0	48.0, 138.0	50.0, 127.0	
				<0.001
MACE4_FU_Duration_d				
Median (IQR)	1973.0 (1595.0, 2658.0)	2322.0 (1602.0, 2660.0)	1181.0 (560.5, 1792.3)	
Mean (SD)	2021.8 (721.5)	2111.4 (658.4)	1198.8 (755.9)	
Range	1.0, 3015.0	70.0, 3015.0	1.0, 2881.0	
Sex, *n* (%)				0.037
Female	1528 (56.4)	1394 (57.1)	134 (50.4)	
Male	1181 (43.6)	1049 (42.9)	132 (49.6)	

[a]Median (IQR) or frequency (%).
[b]Wilcoxon rank sum test; Pearson's Chi-squared test.

Figure 2 depicts the classification outcomes of various machine learning models, including SVR, RF, LR, KNN, GB, and XGBoost. Furthermore, Figure 3 presents the ROC curve for these models, offering a visual assessment of their performance on the dataset. Among the evaluated models, the GB model demonstrated the best performance, with an AUC of 0.8449, an accuracy of 0.7856, a precision of 0.7067, a recall of 0.5856, and a high F1 score. These metrics collectively underscore the GB model's superiority in accurately identifying individuals with carotid-femoral pulse wave velocity (cf-PWV) values exceeding 10 m/s.

Figure 4 presents a ranking of feature importance from top to bottom along the Y-axis, offering insights into which features contribute most significantly to the model's predictions. Furthermore, Figure 5 specifically shows the feature importance ranking within the GB model. Each point in this figure represents a feature and its corresponding Shapley value, which quantifies the feature's contribution to the model's output. The size of the points is distinguished by color, providing a visual indication of the feature's importance.

The SHAP analysis revealed that baPWV was the most critical feature in predicting cf-PWV, acting as a positive predictor where higher baPWV values are associated with higher predicted cf-PWV in patients. The second most influential feature, following baPWV, was rbsbp, which also served as a positive predictor of cf-PWV. These findings underscore the pivotal role of baPWV in predicting cf-PWV and highlight the importance of considering this feature in clinical assessments.

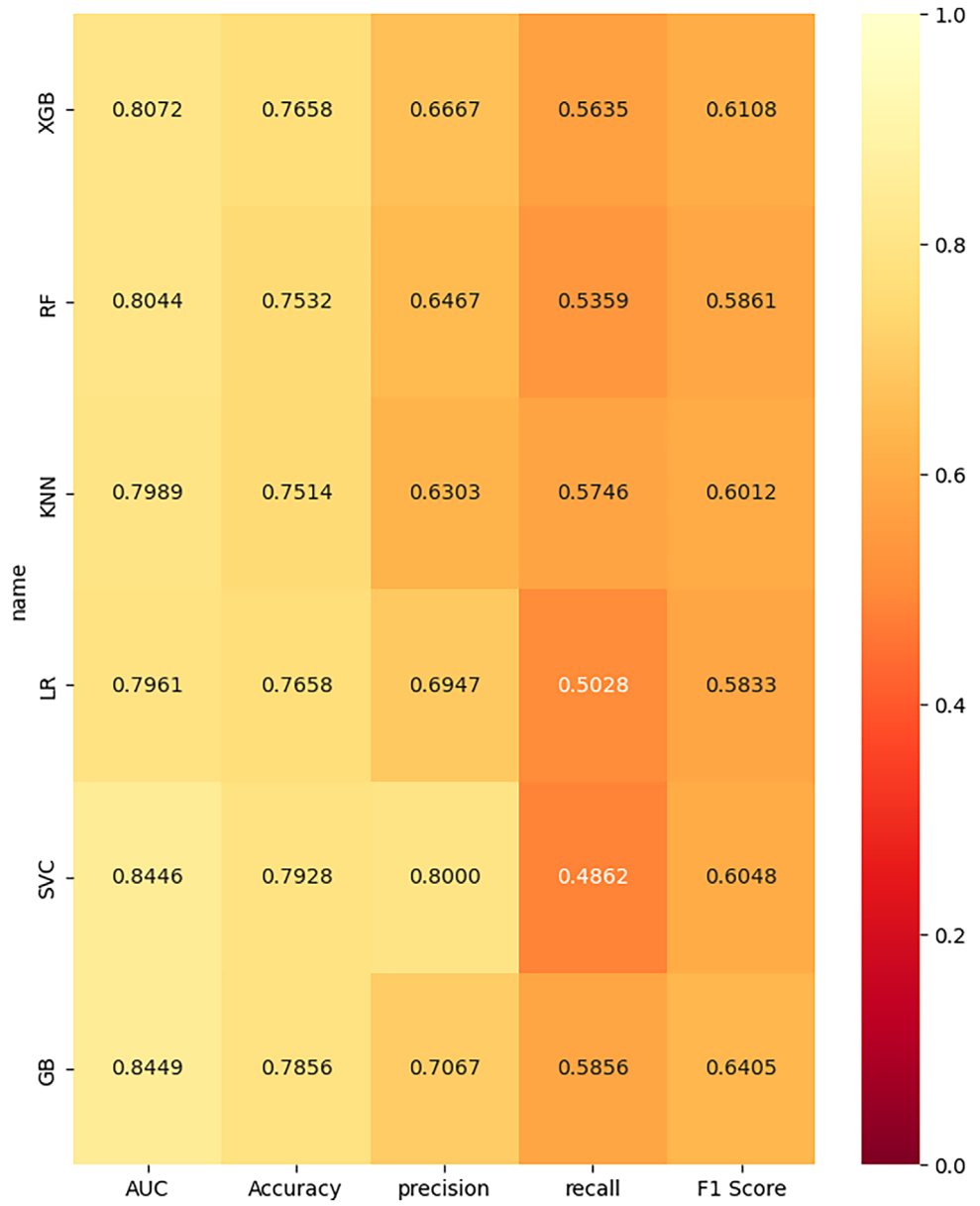

FIGURE 2 | Results of machine learning classification model in the North Shanghai Study. This figure displays the performance of various classification models, including Gradient Boosting Decision Tree (GB), eXtreme Gradient Boosting (XGBoost), K-nearest Neighbor (KNN), Random Forest (RF), Linear Regression (LR), and Support Vector Classifier (SVC), all evaluated based on their Area Under the Curve (AUC) for distinguishing cases with cfPWV greater than 10 m/s.

4 | Discussion

The aim of this study was to develop and validate a predictive model using baPWV and other clinical features to assess the cardiovascular health of patients. cf-PWV is an important indicator of arterial stiffness and serves as an independent predictor of cardiovascular events and mortality. However, traditional measurement methods require specialized equipment and trained personnel and are not readily available. Therefore, our study aimed to explore the potential of machine learning algorithms for accurately predicting cf-PWV, enabling physicians to anticipate their patients' cf-PWV and take further measurements accordingly.

Our results demonstrate the efficacy of machine learning models, particularly linear regression (LR), in predicting cf-PWV values with high accuracy. Among the evaluated models, LR exhibited the lowest RMSE and the highest R^2 value, indicating its strong predictive and fitting capabilities for this dataset. These findings suggest that LR can be a reliable tool for predicting cf-PWV in clinical settings, facilitating timely interventions to reduce CVD risk.

In the classification task, the GB model demonstrated superior performance, characterized by high AUC, accuracy, and recall. These indicators emphasize the model's proficiency in identifying individuals with cf-PWV exceeding 10 m/s, a threshold

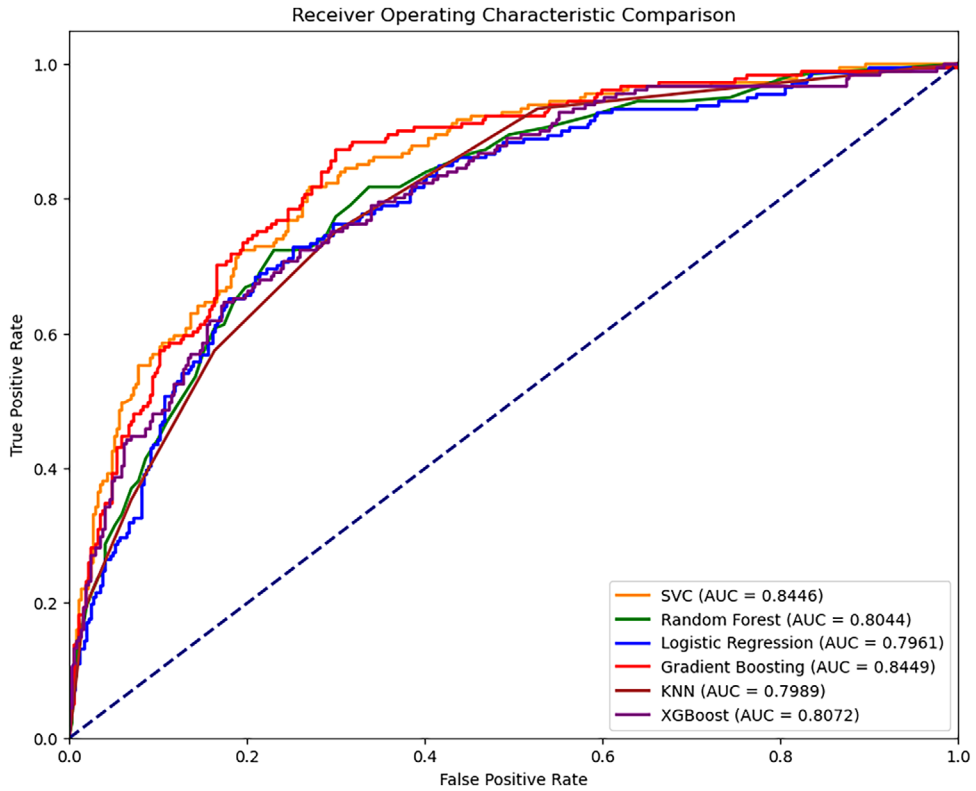

FIGURE 3 | ROC curve of machine learning classification model in North Shanghai Study. Different colored curves are used to represent the receiver operating characteristic (ROC) Curves of various models, including the Support Vector Classifier (SVC), with the Area Under the Curve (AUC) as a performance metric, and eXtreme Gradient Boosting (XGBoost), all plotted with the False Positive Rate (FPR) on the x-axis and the True Positive Rate (TPR) on the y-axis.

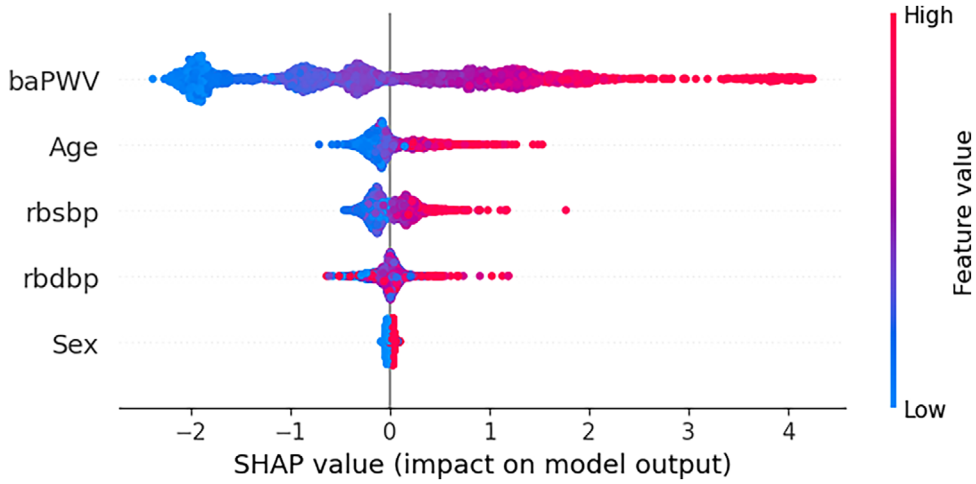

FIGURE 4 | SHAP summary plot. The SHAP summary plot illustrates the impact of various features, namely baPWV (brachial-ankle pulse wave velocity), rbsbp (right brachial systolic blood pressure), and rbdbp (right brachial diastolic blood pressure), on the model's output.

indicative of heightened cardiovascular risk. This capability underscores the GB model's potential for widespread clinical application, enabling physicians to more precisely assess patients' cardiovascular health and guide those requiring further accurate measurements. For patients, knowledge of whether their cf-PWV exceeds 10 m/s is crucial for assessing their cardiovascular health status. This information can motivate them to adopt more proactive health management measures, such as lifestyle modifications and blood pressure control, ultimately reducing their risk of cardiovascular disease.

Furthermore, SHAP analysis highlighted the pivotal role of baPWV and systolic blood pressure (rbsbp) in predicting cf-PWV. This finding not only reinforces the physiological basis of our model but also provides clinicians with valuable insights into the key factors influencing arterial stiffness. By integrating

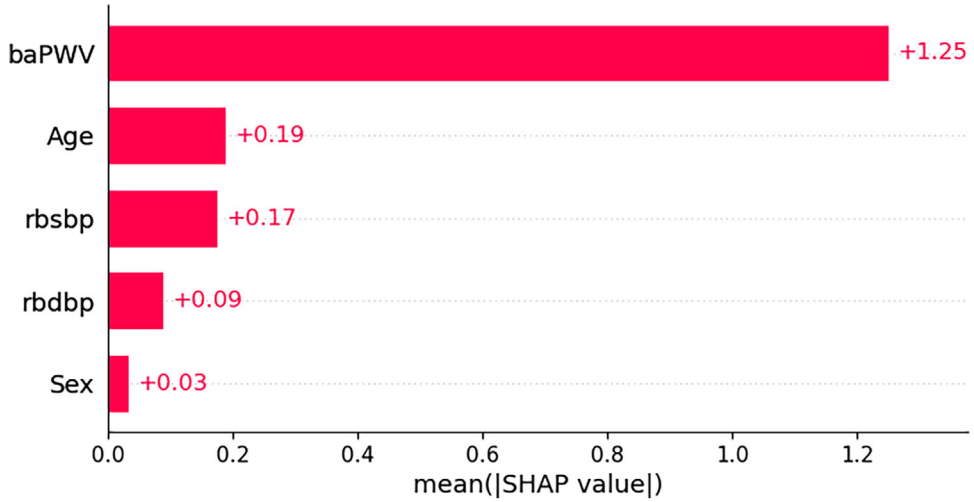

FIGURE 5 | SHAP dependence plot. This graph illustrates the magnitude and direction of the impact of various features, including baPWV (brachial-ankle pulse wave velocity), rbsbp (right brachial systolic blood pressure), and rbdbp (right brachial diastolic blood pressure), on the model's predictions, with positive SHAP values indicating an increase in the model's prediction probability or output value.

TABLE 3 | Comparison of the predictive value of various machine learning models and actual cf-PWV and baPWV for survival time and outcome events using the Cox proportional hazards model.

	−2 Log Likelihood	Chi-square	df	Sig.
Actual_baPWV	4379.680	9.918	1	0.003
Actual_cf-PWV	5069.810	17.882	1	0.000
SVR_cf-PWV	906.595	4.278	1	0.039
RF_cf-PWV	917.286	5.585	1	0.018
LR_cf-PWV	520.956	8.206	1	0.004
KNN_cf-PWV	1027.156	4.387	1	0.036
GB_cf-PWV	869.209	3.965	1	0.046
20%_Actual_baPWV	768.779	0.730	1	0.402
20%_Actual_cfPWV	520.671	8.625	1	0.003

Note: SVR_cf-PWV, RF_cf-PWV, LR_cf-PWVKNN_cf-PWV, and GB_cf-PWV represent cf-PWV prediction by different machine learning models, Actual_cf-PWV and Actual_baPWV are the value of cf-PWV and baPWV measured by an instrument. The dataset used for this comparison includes 20% of the actual baPWV values (20%_Actual_baPWV) and 20% of the actual cf-PWV values (20%_Actual_cfPWV). This table summarizes the performance of five machine learning models in predicting cf-PWV values, assessed using the Cox proportional hazards model. It lists statistical indicators such as -2 Log Likelihood, Chi-square, df, and Sig.

these factors into consideration, physicians can devise more personalized and effective prevention and treatment strategies, ultimately enhancing patients' overall cardiovascular health.

5 | Conclusion

This study not only demonstrates the effectiveness and practicality of machine learning algorithms in predicting cf-PWV but also provides a new assessment tool for early detection and risk assessment for cardiovascular disease. In the future, we will continue to optimize the performance of the model and explore its wide application in the clinic, with a view to contributing to improving the effectiveness of cardiovascular disease management.

Author Contributions

Conceptualization: Tao Hu, Yi Zhang, and Minghui Chen. Methodology: Minghui Chen and Jing Xiong. Validation: Minghui Chen. Formal analysis: Minghui Chen. Investigation: Minghui Chen and Jing Xiong. Resources: Tao Hu and Yi Zhang. Data Curation: Minghui Chen and Jing Xiong. Writin–original draft: Minghui Chen. Visualization: Minghui Chen and Moran Li. Supervision: Tao Hu and Yi Zhang. Project administration: Tao Hu and Yi Zhang. Funding acquisition: Yi Zhang and Tao Hu.

Acknowledgments

This study was financially supported by the National Nature Science Foundation of China (82170388,82370300), Shanghai Technology Research Leader Program (21XD1434700), the Cardiac rehabilitation fund by the International Medical Exchange Foundation (Z-2019-42-1908-3), and Grant for the construction of Innovative Flagship Hospital for Integrated Traditional Chinese and Western Medicine (No., ZY(2021-2023)-0205-05)

Ethics Statement

Shanghai Tenth People's Hospital Ethics Committee approved this study (ethical approval number: 22K148). The research followed the Declaration of Helsinki established by the World Medical Association.

Consent

All participants provided written informed consent.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data underlying the results presented in this study are available from the corresponding author upon reasonable request. Due to privacy and confidentiality concerns, the data are not publicly available. Interested researchers may contact the corresponding author at yizshcn@gmail.com to discuss potential collaboration or data access arrangements.

References

1. World Health Organization, *Global Report on Hypertension: The Race Against a Silent Killer* (World Health Organization, 2023), 291 p. ISBN: 978-92-4-008106-2.

2. C. Weiwei, G. Runlin, L. Lisheng, et al., "Outline of the Report on Cardiovascular Diseases in China, 2014," *European Heart Journal Supplements* 18, no. suppl F (2016): F2–F11, https://doi.org/10.1093/eurheartj/suw030.

3. B. M. Kaess, J. Rong, M. G. Larson, et al., "Aortic Stiffness, Blood Pressure Progression, and Incident Hypertension," *JAMA* 308, no. 9 (2012): 875–881.

4. P. Boutouyrie, A. I. Tropeano, R. Asmar, et al., "Aortic Stiffness Is an Independent Predictor of Primary Coronary Events in Hypertensive Patients: A Longitudinal Study," *Hypertension* 39, no. 1 (2002): 10–15.

5. T. Willum-Hansen, J. A. Staessen, C. Torp-Pedersen, et al., "Prognostic Value of Aortic Pulse Wave Velocity as Index of Arterial Stiffness in the General Population," *Circulation* 113, no. 5 (2006): 664–670, https://doi.org/10.1161/CIRCULATIONAHA.105.579342.

6. G. F. Mitchell, S. J. Hwang, R. S. Vasan, et al., "Arterial Stiffness and Cardiovascular Events: The Framingham Heart Study," *Circulation* 121, no. 4 (2010): 505–511, https://doi.org/10.1161/CIRCULATIONAHA.109.886655.

7. C. Vlachopoulos, K. Aznaouridis, and C. Stefanadis, "Prediction of Cardiovascular Events and All-Cause Mortality With Arterial Stiffness: A Systematic Review and Meta-analysis," *Journal of the American College of Cardiology* 55, no. 13 (2010): 1318–1327, https://doi.org/10.1016/j.jacc.2009.10.061.

8. Y. Ben-Shlomo, M. Spears, C. Boustred, et al., "Aortic Pulse Wave Velocity Improves Cardiovascular Event Prediction: An Individual Participant Meta-Analysis of Prospective Observational Data From 17,635 Subjects," *Journal of the American College of Cardiology* 63, no. 7 (2014): 636–646, https://doi.org/10.1016/j.jacc.2013.09.063.

9. J. C. Bramwell and A. V. Hill, "Velocity of Transmission of the Pulse-wave," *Lancet* 199, no. 5149 (1922): 891–892.

10. E. D. Kim, S. H. Ballew, H. Tanaka, G. Heiss, J. Coresh, and K. Matsushita, "Short-Term Prognostic Impact of Arterial Stiffness in Older Adults Without Prevalent Cardiovascular Disease," *Hypertension* 74, no. 6 (2019): 1373–1382, https://doi.org/10.1161/HYPERTENSIONAHA.119.13496.

11. H. Tanaka, F. A. Dinenno, K. D. Monahan, C. M. Clevenger, C. A. DeSouza, and D. R. Seals, "Aging, Habitual Exercise, and Dynamic Arterial Compliance," *Circulation* 102, no. 11 (2000): 1270–1275, https://doi.org/10.1161/01.cir.102.11.1270.

12. R. R. Townsend, I. B. Wilkinson, E. L. Schiffrin, et al., "Recommendations for Improving and Standardizing Vascular Research on Arterial Stiffness: A Scientific Statement From the American Heart Association," *Hypertension* 66, no. 3 (2015): 698–722, https://doi.org/10.1161/HYP.0000000000000033.

13. J. Dong, T. Feng, B. Thapa-Chhetry, et al., "Machine Learning Model for Early Prediction of Acute Kidney Injury (AKI) in Pediatric Critical Care," *Critical Care (London, England)* 25, no. 1 (2021): 288, https://doi.org/10.1186/s13054-021-03724-0.

14. R. O. Alabi, A. A. Mäkitie, M. Pirinen, M. Elmusrati, I. Leivo, and A. Almangush, "Comparison of Nomogram With Machine Learning Techniques for Prediction of Overall Survival in Patients With Tongue Cancer," *International Journal of Medical Informatics* 145 (2021): 104313, https://doi.org/10.1016/j.ijmedinf.2020.104313.

15. A. Rahman, Y. Chang, J. Dong, et al., "Early Prediction of Hemodynamic Interventions in the Intensive Care Unit Using Machine Learning," *Critical Care (London, England)* 25, no. 1 (2021): 388, https://doi.org/10.1186/s13054-021-03808-x.

16. L. N. Grendas, L. Chiapella, D. E. Rodante, and F. M. Daray, "Comparison of Traditional Model-Based Statistical Methods With Machine Learning for the Prediction of Suicide Behaviour," *Journal of Psychiatric Research* 145 (2021): 85–91, https://doi.org/10.1016/j.jpsychires.2021.11.029.

17. S. G. Choi, M. Oh, D. H. Park, et al., "Comparisons of the Prediction Models for Undiagnosed Diabetes Between Machine Learning Versus Traditional Statistical Methods," *Scientific Reports* 13, no. 1 (2023): 13101, https://doi.org/10.1038/s41598-023-40170-0.

18. Z. Zhang, K. M. Ho, and Y. Hong, "Machine Learning for the Prediction of Volume Responsiveness in Patients With Oliguric Acute Kidney Injury in Critical Care," *Critical Care (London, England)* 23, no. 1 (2019): 112, https://doi.org/10.1186/s13054-019-2411-z.

19. K. Inoue, T. E. Seeman, T. Horwich, M. J. Budoff, and K. E. Watson, "Heterogeneity in the Association between the Presence of Coronary Artery Calcium and Cardiovascular Events: A Machine-Learning Approach in the MESA Study," *Circulation* 147, no. 2 (2023): 132–141, https://doi.org/10.1161/CIRCULATIONAHA.122.062626.

20. A. Ward, A. Sarraju, S. Chung, et al., "Machine Learning and Atherosclerotic Cardiovascular Disease Risk Prediction in a Multi-Ethnic Population," *NPJ Digital Medicine* 3 (2020): 125, https://doi.org/10.1038/s41746-020-00331-1.

21. F. Sánchez-Cabo, X. Rossello, V. Fuster, et al., "Machine Learning Improves Cardiovascular Risk Definition for Young, Asymptomatic Individuals," *Journal of the American College of Cardiology* 76, no. 14 (2020): 1674–1685, https://doi.org/10.1016/j.jacc.2020.08.017.

22. B. M. Kaess, J. Rong, M. G. Larson, et al., "Aortic Stiffness, Blood Pressure Progression, and Incident Hypertension," *JAMA* 308, no. 9 (2012): 875–881, https://doi.org/10.1001/2012.jama.10503.

23. H. Ji, J. Xiong, S. Yu, et al., "Northern Shanghai Study: Cardiovascular Risk and Its Associated Factors in the Chinese Elderly – A Study Protocol of a Prospective Study Design," *BMJ Open* 7, no. 3 (2017): e013880, https://doi.org/10.1136/bmjopen-2016-013880.

24. S. M. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," in *Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17)* (Curran Associates Inc., 2017), 4768–4777.

25. L. S. Shapley, "A Value for n-Person Games," *Contributions to the Theory of Games* (1953).