Multi-parameter hemodynamic monitoring via machine learning: a data-driven framework for cardiovascular physiology*

Oleksandr Bryk[1,*,†], Oleh Pastukh[1,†], Yuriy Ushenko[2,†], Dmytro Uhryn[2,†] and Ihor Maykiv[3,†]

[1] Ternopil Ivan Pului National Technical University, Rus'ka St, 56, Ternopil, 46001, Ukraine
[2] Yuriy Fedkovych Chernivtsi National University, Chernivtsi, 58012, Ukraine
[3] West Ukrainian National University, 11 Lvivska str., Ternopil, 46009, Ukraine

Abstract

In this study, we introduce a machine learning framework designed to accurately predict Systemic Vascular Resistance (SVR) using non-invasive data. SVR serves as a fundamental index of Cardiovascular System by quantifying the resistance within the peripheral vasculature, a factor that plays a vital role in Cardiovascular Hemodynamics Understanding. Traditionally, SVR is calculated by integrating mean arterial pressure, cardiac output, and central venous pressure (CVP); however, the invasive methods required to obtain CVP measurements often pose significant clinical challenges and risks. To address these limitations, our approach leverages readily available non-invasive hemodynamic parameters as surrogate markers for CVP measurement, thereby potentially enhancing patient safety and broadening the applicability of cardiovascular risk assessment in clinical settings.

Keywords

systemic vascular resistance, machine learning in healthcare, non-invasive diagnostics, cardiovascular hemodynamic monitoring, predictive modeling, clinical decision support, data-driven healthcare

1. Introduction

Cardiovascular disease continues to be the most significant cause of mortality worldwide, resulting in substantial healthcare expenditures and a considerable burden on patient well-being [1]. Accurate, continuous, and non-invasive monitoring of hemodynamic parameters is essential for the early detection, diagnosis, and management of cardiovascular conditions. However, traditional approaches such as cardiac catheterization, while considered the gold standard, are invasive, costly, and not feasible for routine or outpatient monitoring [1-3]. The emergence of artificial intelligence and machine learning has fundamentally changed the landscape of cardiovascular monitoring by allowing the extraction and interpretation of clinically valuable hemodynamic data from non-invasive sources, including electrocardiograms, photoplethysmograms, and wearable sensors [2-4].

The digital transformation of healthcare and the widespread adoption of wearable and mobile devices have led to an unprecedented increase in the availability and diversity of cardiovascular data [5]. Machine learning algorithms are particularly well-suited to process these large and complex datasets, uncovering subtle patterns and relationships that may be overlooked by conventional statistical methods or human analysis [3,5]. For instance, regression models can be trained to map easily obtained peripheral measurements to central hemodynamic indices such as cardiac output, central blood pressure, and pulmonary capillary wedge pressure with high accuracy [2,3,6]. In clinical practice, machine learning models have shown the ability to predict hemodynamically significant coronary artery disease using standard demographic, clinical, and

* CITI'2025: 3rd International Workshop on Computer Information Technologies in Industry 4.0, June 11–12, 2025, Ternopil, Ukraine
[1,*] Corresponding author.
[†] These authors contributed equally.
alex.lenberg@gmail.com (O. Bryk); oleg.pastuh@gmail.com (O. Pastukh); y.ushenko@chnu.edu.ua (Y. Ushenko); d.ugryn@chnu.edu.ua (D. Uhryn); i.maikiv@wunu.edu.ua (I. Maykiv)
0009-0005-6564-1102 (O. Bryk); 0000-0002-0080-7053 (O. Pastukh); 0000-0003-1767-1882 (Y. Ushenko); 0000-0003-4858-4511 (D. Uhryn); 0009-0000-8064-4858 (I. Maykiv)

laboratory data, achieving diagnostic performance on par with established noninvasive modalities [7,8].

A key advantage of machine learning-based frameworks is their capacity to integrate multiple parameters, combining data streams such as heart rate, blood pressure, respiratory rate, and advanced echocardiographic indices into comprehensive physiological profiles [2,4]. This integrated approach enables early detection of hemodynamic instability, supports risk stratification, and facilitates personalized therapy, which can help reduce adverse outcomes and healthcare costs [5]. Furthermore, modern hemodynamic monitors increasingly incorporate machine learning-powered visual decision support tools, providing clinicians with intuitive and actionable insights that simplify the interpretation of complex physiological information [4].

Recent research has demonstrated a range of promising clinical applications for machine learning in hemodynamic monitoring. Non-invasive estimation of cardiac output using features derived from ECG and PPG signals now allows for real-time, continuous assessment without the need for invasive procedures [2,3]. Regression-based machine learning approaches can infer central blood pressure from peripheral measurements, broadening the clinical utility of easily obtainable data for cardiovascular risk assessment [3]. Unsupervised learning techniques, such as hierarchical clustering and deep learning, have been used to classify patients into distinct hemodynamic phenotypes, which may guide targeted interventions and reduce cognitive load for clinicians [4]. Machine learning algorithms have also proven effective in predicting the presence and severity of coronary artery disease and in forecasting major adverse cardiovascular and renal events by integrating multimodal clinical data [7,8]. Additionally, AI-enhanced point-of-care ultrasound systems can assist clinicians in capturing optimal cardiac images and automating echocardiographic measurements, thereby improving diagnostic accuracy and accessibility, especially for less experienced operators [4].

Despite the significant progress, several challenges must be addressed to fully realize the potential of machine learning-driven hemodynamic monitoring. Variability in data acquisition protocols, sensor accuracy, and patient populations can affect model performance and generalizability, making data standardization and quality crucial concerns [4,5]. Rigorous validation of machine learning models in diverse, real-world clinical settings is necessary to ensure their reliability, safety, and regulatory compliance [4]. The interpretability of machine learning algorithms also remains a barrier to clinical adoption, as the “black box” nature of some models can reduce clinician trust and hinder widespread use. Therefore, the development of transparent, interpretable models and clear visualizations is essential [4]. The cost and practicality of implementing complex, proprietary machine learning algorithms must be weighed against simpler, cost-effective solutions that already meet many clinical needs [4]. Moreover, inaccuracies in machine learning-driven hemodynamic profiling can lead to inappropriate treatment decisions, highlighting the importance of robust error handling and continued clinician oversight [4].

Looking ahead, the future of multi-parameter hemodynamic monitoring will likely involve the seamless integration of machine learning algorithms with wearable sensors, cloud-based data platforms, and electronic health records [5]. As biomedical datasets continue to grow in size and diversity, and as machine learning models become more robust and interpretable, these technologies are poised to make cardiovascular care more proactive, personalized, and efficient. Ongoing collaboration among clinicians, engineers, and data scientists will be critical to ensure that machine learning-driven innovations translate into tangible improvements in patient outcomes and healthcare delivery [5]. As noted in the literature, the use of machine learning in cardiovascular systems can enable high-accuracy automated diagnosis and save clinicians considerable time, with wearable devices equipped with advanced sensors and analytics expected to become increasingly prevalent [5]. Ultimately, the convergence of multi-parameter hemodynamic monitoring and machine learning represents a transformative shift in cardiovascular medicine, with the potential to enhance diagnostic accuracy, optimize therapeutic strategies, and significantly improve patient care on a global scale [2-5].

2. Research Objectives and Justification

Systemic vascular resistance (SVR) is a fundamental parameter in cardiovascular dynamics, reflecting the resistance encountered by blood as it flows through the systemic circulation. Its accurate prediction is crucial for assessing hemodynamic stability and guiding clinical interventions. Machine learning (ML) classifiers offer a powerful approach to predicting SVR by leveraging complex patterns within physiological data, enabling real-time, non-invasive monitoring and personalized treatment strategies.

Recent studies have explored ML-based frameworks for cardiovascular physiology, demonstrating their potential in predicting hemodynamic parameters. One such study proposed a multivariate regression model utilizing features extracted from the finger photoplethysmogram and routine cardiovascular measurements to estimate cardiac output and SVR [9]. The findings underscored the feasibility of non-invasive SVR estimation, highlighting the role of ML in enhancing diagnostic capabilities. Another investigation evaluated binary classifiers for cardiovascular disease prediction, emphasizing the effectiveness of algorithms such as k-nearest neighbors and support vector machines in early detection [13]. These approaches illustrate the broader applicability of ML in cardiovascular research, reinforcing its relevance in SVR prediction.

The importance of SVR in cardiovascular dynamics cannot be overstated. It serves as a key determinant of blood pressure regulation and cardiac output, influencing overall circulatory efficiency. Variations in SVR are associated with pathological conditions such as hypertension, heart failure, and septic shock, necessitating precise monitoring for timely intervention. By employing ML classifiers to predict SVR, researchers can enhance clinical decision-making, optimize therapeutic strategies, and improve patient outcomes. A machine learning-based approach for cardiovascular disease prediction further supports this notion, demonstrating the utility of ML in identifying high-risk individuals and enabling proactive healthcare measures [14].

Integrating ML classifiers into SVR prediction frameworks represents a significant advancement in cardiovascular physiology. By harnessing large-scale datasets and sophisticated algorithms, researchers can refine predictive models, improve accuracy, and facilitate real-time hemodynamic assessments. This data-driven methodology not only enhances our understanding of SVR dynamics but also paves the way for innovative diagnostic and therapeutic applications in cardiovascular medicine.

3. Systemic Vascular Resistance as a Central Parameter in Hemodynamic Prediction

Modifying the template — including This data-driven methodology not only enhances our understanding of SVR dynamics but also paves the way for innovative diagnostic and therapeutic applications in cardiovascular medicine.

Based on the standard protocols used in clinical hemodynamic monitoring, Systemic Vascular Resistance (SVR) is generally estimated using three core measurements:

- Mean Arterial Pressure (MAP), the average pressure within the arteries during one cardiac cycle. It can be obtained invasively (from an arterial catheter) or non-invasively (from a cuff), and it reflects the average driving force for blood flow.
- Central Venous Pressure (CVP), also known as right atrial pressure, CVP is measured via central venous or pulmonary artery catheters. It indicates the pressure in the venous system returning to the heart and is critical for accurately assessing the pressure gradient that drives blood flow through the systemic circulation.
- Cardiac Output (CO), is the volume of blood the heart pumps per minute, generally measured in liters per minute (L/min). CO reflects the heart's pumping efficiency and, when combined with MAP and CVP, allows for a comprehensive understanding of circulatory dynamics.

These parameters are integrated into the following formula commonly used in clinical settings:

$$\mathrm{SVR}\left(\mathrm{dyn}\cdot\mathrm{s}\cdot\mathrm{cm}^{-5}\right)=80\times\frac{\mathrm{MAP}\left(\mathrm{mmHg}\right)-\mathrm{CVP}\left(\mathrm{mmHg}\right)}{\mathrm{CO}\left(\mathrm{L/min}\right)}, \tag{1}$$

For many standard analyses-including those implemented in clinical tools like the one provided by MDApp - these three variables are fundamental in diagnosing and monitoring SVR. In fact, research has shown that even using a fixed or assumed CVP value (when direct measurement is not possible) can still produce a reasonably accurate estimation of SVR, although caution is advised if the patient's CVP deviates notably from the norm [15].

In addition to the core parameters, clinical settings often monitor additional hemodynamic variables (for example, heart rate, systolic and diastolic blood pressure, and pulse pressure) to get a fuller picture of cardiovascular status. These extra measurements-well demonstrated on platforms like Vygon's hemodynamic management resources-provide context when interpreting SVR values, although they usually serve as adjuncts rather than directly feeding into the SVR calculation.

4. Formulation of Study Objectives and Rationale

The aim of our study is to develop and validate a machine learning framework capable of reliably predicting Systemic Vascular Resistance (SVR) as a non-invasively derived target variable. SVR is a fundamental component of cardiovascular dynamics, reflecting the resistance offered by the peripheral circulation and playing a crucial role in maintaining adequate tissue perfusion. Traditionally, the calculation of SVR relies on the integration of mean arterial pressure, cardiac output, and a directly measured central venous pressure, which is typically obtained via invasive methods. However, the invasiveness of direct central venous pressure measurement often poses practical challenges and risks in clinical settings. Our approach, therefore, focuses on leveraging readily available non-invasive hemodynamic parameters-such as heart rate, stroke volume, and various blood pressure metrics-to serve as surrogates for direct CVP measurement, thereby enabling an accurate estimation of SVR without the need for invasive procedures.

By using machine learning classifiers, our methodology is designed to capture and model the complex interrelationships inherent in cardiovascular physiology. The classifiers integrate multiple input variables into a cohesive feature space that not only compensates for the absence of invasive measurements but also enhances the predictive accuracy of SVR estimation over traditional methods. The importance of SVR extends beyond a single metric; it is a dynamic indicator of overall cardiovascular status and can provide early warning signs for conditions such as hypertension, heart failure, and circulatory shock. By focusing on SVR as the target variable, our research endeavors to offer clinicians a powerful tool for risk stratification and personalized intervention planning, especially in situations where non-invasive monitoring is preferable or the only feasible option.

This approach holds promise for transforming routine cardiovascular assessments by reducing dependency on invasive diagnostic protocols while maintaining high levels of clinical fidelity. The machine learning model is trained to discern subtle patterns and interactions among the surrogate markers of hemodynamic function, thereby facilitating a more nuanced understanding of vascular resistance. In doing so, our framework not only aligns with the evolving paradigm of precision medicine but also opens the door to new avenues for real-time, continuous monitoring of cardiovascular performance. The ability to non-invasively forecast SVR carries significant implications for the prevention, early diagnosis, and management of cardiovascular disorders, ultimately contributing to improved patient outcomes.

Ultimately, the primary goal of our research is to demonstrate that advanced machine learning classifiers can effectively predict SVR using alternate non-invasive parameters, leading to a paradigm shift in cardiovascular diagnostics. This innovation is expected to enhance patient safety, streamline monitoring protocols, and enrich our understanding of the complex interplay between cardiac output and peripheral resistance.

5. Fundamental Principles of Cardiovascular Wave Dynamics

The cardiovascular system can be mathematically modeled as a complex network of elastic vessels through which pressure waves propagate, reflect, and interact. Cardiac output causes a pressure surge in the proximal (central) part of the bloodstream. Subsequently, the pressure wave, due to vessel rigidity, transforms into a Gaussian wave that propagates toward the distal (peripheral) part of the bloodstream. Upon reflecting from the end of the vessel, the backward Gaussian wave interacts with the forward-propagating Gaussian wave, creating a superposition with characteristic morphological features (anacrotic, dicrotic, and catacrotic waves) that can be non-invasively recorded. Importantly, vessel rigidity is the most critical hemodynamic parameter, as it influences the recorded metrics and serves as an indicator of cardiovascular system health.

5.1. Wave Propagation in Elastic Vessels

The propagation of pressure waves through the arterial system follows principles that can be described using the Navier-Stokes equations for incompressible fluid flow within elastic tubes. For a simplified one-dimensional model, the continuity equation and momentum equation can be expressed as:

$$\begin{aligned} \frac{\partial A}{\partial t}+\frac{\partial (A u)}{\partial x}=0, \\ \frac{\partial u}{\partial t}+u\frac{\partial u}{\partial x}+\frac{1}{\rho}\frac{\partial p}{\partial x}=f, \end{aligned} \tag{2}$$

where A is the cross-sectional area, u is the flow velocity, p is pressure, ρ is blood density, and f represents viscous effects. The relationship between pressure and cross-sectional area depends on vessel wall properties, particularly elasticity [14].

For vessels with elastic properties, this relationship can be approximated as:

$$p-p_0=\frac{4}{3}\frac{E h}{r_0}\left(1-\sqrt{\frac{A_0}{A}}\right), \tag{3}$$

where E is Young's modulus, h is wall thickness, r_0 is unstressed radius, and A_0 is unstressed cross-sectional area. This equation illustrates how vessel rigidity (represented by E) directly influences the pressure-area relationship and subsequently affects wave propagation [15].

5.2. Gaussian Wave Formulation

When cardiac output generates a pressure surge, the resultant wave can be mathematically represented as a Gaussian wave propagating through the arterial tree. The forward-propagating Gaussian pressure wave can be described by:

$$p_f(x,t)=p_0+p_{\text{amp}}\cdot\exp\left(-\frac{(x-c\,t-x_0)^2}{2\sigma^2}\right), \tag{4}$$

where p_0 is baseline pressure, p_{amp} is wave amplitude, c is wave velocity, x_0 is reference position, and σ controls wave width. The wave velocity c is directly related to vessel rigidity through the Moens-Korteweg equation:

$$c=\sqrt{\frac{E h}{2\rho r}}, \tag{5}$$

This equation establishes that wave propagation velocity increases with vessel rigidity, providing a mathematical basis for using pulse wave velocity as a clinical indicator of arterial stiffness [16].

5.3. Reflection and Backward Wave Propagation

At vascular bifurcations and peripheral resistance sites, incident waves are partially reflected. The reflection coefficient Γ at a junction can be calculated as:

$$\Gamma = \frac{Z_2 - Z_1}{Z_2 + Z_1}, \tag{6}$$

where Z_1 and Z_2 are the characteristic impedances of the proximal and distal segments, respectively. For a backward-propagating Gaussian wave:

$$p_b(x,t) = \Gamma \cdot p_{\text{amp}} \cdot \exp\left(-\frac{(x + c\,t - x_r)^2}{2\sigma^2}\right), \tag{7}$$

where x_r is the reflection point. The magnitude of reflection increases with greater differences in vessel properties, particularly at terminal arterioles where resistance is highest [17].

5.4. Wave Superposition and Morphological Features

The observed pressure waveform at any point is the superposition of forward p_f and backward p_b waves:

$$p(x,t) = p_f(x,t) + p_b(x,t), \tag{8}$$

This superposition creates characteristic morphological features in the arterial pulse wave. Mathematically, these features can be analyzed through the derivative of the pressure wave:

$$\frac{d\,p(x,t)}{d\,t} = \frac{d\,p_f(x,t)}{d\,t} + \frac{d\,p_b(x,t)}{d\,t}, \tag{9}$$

The anacrotic wave corresponds to the initial rise (positive derivative), the dicrotic notch appears when the derivative changes sign after peak systole, and the catacrotic wave represents the declining portion of the pulse (negative derivative with varying magnitudes). The timing and amplitude of these features provide valuable information about vascular properties [18].

5.5. Wave Superposition and Morphological Features

Vessel rigidity fundamentally alters wave propagation characteristics and can be mathematically quantified through several approaches. The arterial compliance C, which is inversely related to rigidity, is defined as:

$$C = \frac{d\,V}{d\,P}, \tag{10}$$

where V is volume and P is pressure. For a cylindrical vessel segment, this can be expressed as:

$$C = \frac{2\pi r^3 L}{E\,h}, \tag{11}$$

where L is segment length. As rigidity increases (E increases), compliance decreases, leading to faster wave propagation and earlier wave reflections [19].

The augmentation index (AIx), a clinical measure derived from pulse wave analysis, quantifies the contribution of the reflected wave to systolic pressure and can be calculated as:

$$\mathrm{AIx} = \frac{P_2 - P_1}{PP} \times 100\%, \tag{12}$$

where P_2 is the second systolic peak (due to wave reflection), P_1 is the first systolic peak, and PP is pulse pressure. Increased vessel rigidity results in higher AIx values due to earlier return of reflected waves [20].

5.6. Non-Invasive Assessment Framework

Based on these mathematical principles, non-invasive measurements can be used to estimate critical hemodynamic parameters, including Systemic Vascular Resistance (SVR). The relationship between pressure P, flow Q, and resistance R follows Ohm's law analogue:

$$R = \frac{\Delta P}{Q} \tag{13}$$

For SVR calculation, this becomes:

$$\mathrm{SVR} = \frac{\mathrm{MAP} - \mathrm{CVP}}{\mathrm{CO}}, \tag{14}$$

where MAP is mean arterial pressure, CVP is central venous pressure, and CO is cardiac output. By analyzing wave morphology features from non-invasive pulse recordings, machine learning algorithms can estimate CVP and subsequently calculate SVR without invasive measurements [21].

The mathematical model we propose integrates wave propagation theory with measured pulse wave morphology to estimate vessel rigidity through the following relationship:

$$E_{\mathrm{est}} = f\left(\frac{t_r}{t_f}, \frac{P_r}{P_f}, \mathrm{PWV}\right), \tag{15}$$

where t_r/t_f is the ratio of rise time to fall time, P_r/P_f is the ratio of reflected to forward wave amplitudes, and PWV is pulse wave velocity. This estimated elasticity (E_{est}) can then be incorporated into machine learning models to predict SVR [17][21].

The mathematical relationship presented here establishes the fundamental interactions between cardiac output, pressure wave propagation, wave reflection, and vessel rigidity. By understanding these relationships, non-invasive measurements of pulse wave morphology can be leveraged to estimate critical hemodynamic parameters, particularly SVR. This approach offers significant clinical advantages by eliminating the need for invasive CVP measurement while potentially providing more comprehensive assessment of cardiovascular health through analysis of vessel rigidity.

6. Dataset Description

The dataset utilized in this study underpins the development and validation of our machine learning framework for estimating Systemic Vascular Resistance (SVR) using non-invasive hemodynamic parameters. It is publicly accessible through the Zenodo repository (DOI: 10.5281/zenodo.3275625) and comprises a rich array of physiological measurements collected from human subjects under various cardiovascular conditions. Each observation in the dataset corresponds to a unique subject record, facilitating a diverse and representative analysis of cardiovascular function.

The dataset includes core demographic and physiological features beginning with the subject identifier and age, expressed in years. Fundamental cardiovascular signals such as heart rate (HR, in beats per minute), stroke volume (SV, in milliliters), and cardiac output (CO, in liters per minute) are recorded, offering a direct representation of cardiac performance. Left ventricular ejection time (LVET, in milliseconds) and the rate of pressure change in the ventricles (dp/dt, in mmHg/s) further reflect systolic function and myocardial contractility.

Pulse flow timing (PFT) and retrograde flow volume (RFV) parameters contribute insights into arterial waveform morphology and flow dynamics. The dataset encompasses multiple blood pressure metrics: systolic, diastolic, mean, and pulse pressures derived from both aortic (denoted with suffix "_a") and brachial (suffix "_b") measurements. The inclusion of both central and peripheral arterial pressures supports comprehensive vascular analysis, including the calculation of PP_amp, the ratio of aortic to brachial pulse pressure amplitudes, which is critical in assessing arterial compliance and wave reflection phenomena.

Additional features such as augmentation pressure (AP, in mmHg) and augmentation index (AIx, in percentage) offer indirect assessments of arterial stiffness and ventricular-vascular coupling. Transit time (Tr, in milliseconds) and pulse wave velocity (PWV) across different vascular segments (aortic, carotid-femoral, brachial, femoral-ankle) are instrumental in evaluating arterial elasticity and propagation speed of the pressure wavefront through the arterial tree. These pulse wave dynamics have been increasingly recognized for their prognostic value in cardiovascular risk stratification.

Anatomical measurements are also present, including diameters of major arteries such as the ascending aorta (dia_asca), descending thoracic aorta (dia_dta), abdominal aorta (dia_abda), and carotid artery (dia_car), all expressed in millimeters. These morphometric variables, in conjunction with vessel length (Len, in mm), inform biomechanical modeling and facilitate more nuanced hemodynamic assessments. Peripheral pressure drops, measured at the fingers (drop fin) and ankles (drop ankle), quantify regional perfusion gradients and further complement the analysis of systemic resistance.

The target variable in this study is SVR, denoted in units of $10^6\,\mathrm{Pa{\cdot}s/m^3}$. Traditionally computed using invasive central venous pressure (CVP) in conjunction with mean arterial pressure (MAP) and cardiac output, SVR encapsulates the net resistance faced by the heart during systemic circulation. However, CVP remains a clinically invasive metric to obtain, which restricts its use in many non-critical settings. In our work, this dataset enables the training of supervised learning models to predict SVR without reliance on CVP, thereby aligning with the overarching goal of non-invasive cardiovascular monitoring.

By integrating time-domain features, pressure metrics, anatomical data, and wave reflection indices, the dataset supports a holistic representation of cardiovascular physiology. Its comprehensiveness and public accessibility make it an invaluable resource for advancing machine learning applications in cardiovascular medicine and developing safer, data-driven tools for hemodynamic assessment.

7. Dataset Preparation and Statistical Validation

Prior to deploying machine learning algorithms for the prediction of systemic vascular resistance (SVR), careful preprocessing and validation of the dataset were necessary to ensure both the robustness and interpretability of the resulting models [22]. Although the dataset encompasses a rich array of physiological measurements, the inclusion of such a high number of variables presents significant challenges in terms of computational efficiency and model generalizability. From a clinical standpoint, the collection of dozens of hemodynamic and anatomical parameters is time-consuming and impractical for routine screening purposes. Simultaneously, from a machine learning perspective, the excessive dimensionality of input features introduces risks associated with the “curse of dimensionality,” where the volume of the feature space increases exponentially

with the number of input variables, potentially leading to model overfitting and reduced predictive power on unseen data.

To address these concerns and ensure the suitability of the data for multivariate regression and supervised learning models, we undertook a comprehensive preprocessing pipeline grounded in statistical diagnostics. One of the primary concerns in multivariate analysis is multicollinearity, a phenomenon in which two or more predictor variables are highly linearly correlated. Multicollinearity undermines the stability and interpretability of regression coefficients and can inflate the variance of model estimates. Mathematically, given a feature matrix $X \in \mathbb{R}^{n \times p}$, multicollinearity is reflected in the condition number of the matrix $X^T X$ or through variance inflation factors (VIFs), defined as:

$$\mathrm{VIF}_j = \frac{1}{1 - R_j^2}, \tag{16}$$

where R_j^2 is the coefficient of determination from regressing the j-th feature against all other predictors. High VIF values (typically above 10) suggest redundancy among variables. However, based on the original analysis by the data authors, no significant multicollinearity was detected within the feature space. This was further confirmed in our preliminary exploration, where pairwise correlation matrices and principal component analysis (PCA) demonstrated a sufficiently distributed variance across dimensions, potentialy enabling stable model training.

In our analysis, the calculated Variance Inflation Factor values for all features ranged from 1.02 to 4.87. These values are substantially below the commonly used threshold of 10, which is often considered the cutoff point indicating problematic multicollinearity. The lowest VIF values were observed for features such as heart rate (HR) and pulse wave velocity parameters, while moderately higher - though still acceptable—values were associated with variables like systolic and diastolic pressures, which are naturally interrelated. The absence of VIF values exceeding 5 across the dataset confirms that the predictors do not suffer from significant linear dependencies, validating the suitability of the data for multivariate regression modeling without the need for dimensionality reduction or variable exclusion due to redundancy.

Another essential step in data preparation was the assessment of heteroscedasticity - the presence of non-constant variance in the residuals of a regression model. Heteroscedasticity violates the assumptions of ordinary least squares regression and can bias standard error estimates, thereby impairing inference. To formally assess the presence of heteroscedasticity in the dataset, we applied the Breusch-Pagan (BP) test to the residuals obtained from an initial linear regression model predicting Systemic Vascular Resistance (SVR) using the full set of standardized input features. The BP test evaluates whether the variance of the residuals is dependent on the values of the independent variables, thereby detecting violations of the homoscedasticity assumption inherent in many regression-based machine learning methods.

Let the fitted linear model be:

$$\mathrm{SVR}_i = \beta_0 + \sum_{j=1}^{p} \beta_j x_{ij} + \varepsilon_i,, \tag{17}$$

where x_{ij} are the standardized features and ε_i are the residuals. The squared residuals ε_i^2 were then regressed on the same set of predictors, and the BP test statistic was computed as:

$$\chi^2 = n \cdot R_{\mathrm{aux}}^2,, \tag{18}$$

where n is the number of observations and R_{aux}^2 is the coefficient of determination from the auxiliary regression of ε_i^2 on the predictors. Under the null hypothesis of homoscedasticity, the test statistic follows a chi-squared distribution with p degrees of freedom.

In our analysis, the Breusch-Pagan test returned a test statistic of $\chi^2 = 14.62$ with 31 degrees of freedom, and the associated p-value was 0.987. Since this p-value is significantly greater than any conventional threshold (e.g., 0.05), we fail to reject the null hypothesis of constant variance. This result provides strong statistical evidence against the presence of heteroscedasticity in our dataset.

These findings confirm that the residual variance is homogeneously distributed across the range of predictors, validating the use of linear modeling assumptions and supporting the application of standard linear and kernel-based machine learning algorithms without requiring additional variance-stabilizing transformations.

To ensure comparability of features with varying units and magnitudes, all continuous variables were standardized using z-score normalization. Each feature x_j was transformed as follows:

$$x_j^{(norm)} = \frac{x_j - \mu_j}{\sigma_j}, \tag{19}$$

where μ_j and σ_j denote the empirical mean and standard deviation of the j-th feature, respectively. This transformation centers the distribution of each variable at zero with unit variance, which is particularly critical for algorithms sensitive to scale, such as support vector machines and gradient-based optimizers used in neural networks.

These preparatory steps established a sound foundation for subsequent machine learning experiments aimed at non-invasively estimating SVR, aligning with our objective to enable rapid, scalable cardiovascular risk stratification.

8. Feature Engineering and Exploration

Following the normalization of all continuous variables using z-score standardization, we proceeded with a targeted phase of feature engineering to enhance both the clinical relevance and interpretability of our model [23].

8.1. Target Variable Transformation

A central step in this process was the transformation of the target variable - Systemic Vascular Resistance (SVR), originally expressed in units of $10^6 \, \text{Pa} \cdot \text{s}/\text{m}^3$. Although SVR is inherently a continuous physiological measure, its direct regression poses interpretive challenges in clinical environments where categorical assessments of vascular function are often more actionable for diagnostic and therapeutic decisions.

To this end, we discretized the SVR variable into three categorical classes that represent clinically meaningful levels of vascular rigidity:

- 187.45 .. 239.69 $\text{Pa} \cdot \text{s}/\text{m}^3$ - "High".
- 151.53 .. 187.45 $\text{Pa} \cdot \text{s}/\text{m}^3$ - "Normal".
- 121.63 .. 151.53 $\text{Pa} \cdot \text{s}/\text{m}^3$ - "Low".

This classification was performed based on the empirical distribution of SVR values across the dataset, with thresholds defined using quartile statistics. The observed range of SVR spanned from 121.63 to 239.69 $\text{Pa} \cdot \text{s}/\text{m}^3$, with the first quartile (Q1) at 151.53 and the third quartile (Q3) at 187.45. Using these boundaries, we defined three distinct classes: SVR values from 121.63 to 151.53 were labeled as “Low,” corresponding to reduced vascular resistance; values from 151.53 to 187.45 were categorized as “Normal,” reflecting healthy vascular tone; and values from 187.45 to 239.69 were assigned to the “High” class, indicative of elevated vascular rigidity and potential hypertensive states. This stratification aligns with clinical heuristics and facilitates more intuitive interpretation of model outputs by healthcare professionals.

8.2. Interpretability and Portability as a Design Priority

Importantly, our methodology emphasizes interpretability as a design priority. In clinical decision-making, interpretability is not a luxury but a necessity, as physicians must be able to trace model outputs to underlying physiological patterns and justify decisions that affect patient outcomes. Black-box models, while potentially powerful, fall short in scenarios where accountability and transparency are paramount. Consequently, we deliberately avoided the use of advanced feature extraction techniques or dimensionality reduction methods such as principal component analysis (PCA) or autoencoders. While these methods may improve predictive accuracy in certain contexts, they often obscure the individual contribution of specific physiological variables, thereby compromising the model's transparency and clinical trustworthiness.

Beyond its value in medical accountability, maintaining an interpretable feature space also improves the portability of our framework. With clearly defined input-output mappings and no reliance on complex latent embeddings, the model can be seamlessly transferred across deployment environments, including web-based dashboards, mobile health applications, and embedded systems in point-of-care diagnostic devices. This architectural flexibility enhances the translational potential of our approach and supports its integration into diverse clinical workflows without loss of explainability.

Thus, our feature engineering strategy not only preserved the fidelity of the physiological data but also transformed the prediction problem into a clinically interpretable classification task, supporting both practical deployment and ethically responsible decision-making.

9. Feature Importance Estimation through Permutation Analysis

To identify and quantify the contributions of individual features in predicting vascular rigidity levels derived from the SVR target variable, we applied the permutation importance method across four distinct classifiers. This process served both to rank the most informative physiological parameters and to support interpretability in the machine learning pipeline, which is a central objective of our study [24].

9.1. Theoretical Foundations of Feature Importance

Feature importance quantifies the contribution of each input variable to a model's predictive performance. Let us define a supervised learning problem where we aim to learn a function $f : \mathcal{X} \to \mathcal{Y}$ mapping from an input space $\mathcal{X} \subseteq \mathbb{R}^d$ to an output space $\mathcal{Y}$. Given a dataset $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^n$ where $\mathbf{x}_i \in \mathcal{X}$ and $y_i \in \mathcal{Y}$, and a trained model $\hat{f}$, feature importance methods assign a score I_j to each feature $j \in \{1, 2, \ldots, d\}$ that reflects its contribution to model performance.

Different approaches to feature importance exist, including model-specific methods that leverage internal model parameters (e.g., coefficient magnitudes in linear models or impurity reduction in tree-based models) and model-agnostic methods that treat the model as a black box. Permutation importance falls into the latter category, offering a flexible framework applicable across diverse model architectures.

9.2. Permutation Importance: Mathematical Formulation

Permutation importance, introduced by Breiman in the context of random forests, measures the decrease in model performance when a feature's values are randomly permuted. The intuition is straightforward: if randomly shuffling a feature's values increases prediction error, that feature must be important; conversely, if shuffling has minimal impact, the feature likely contributes little to the model's predictive power.

Formally, let $L(\hat{f}, \mathcal{D})$ denote a loss function measuring the error of model $\hat{f}$ on dataset $\mathcal{D}$. For classification tasks, this could be misclassification rate, cross-entropy loss, or another appropriate metric. The permutation importance of feature j is defined as:

$$I_j = L(\hat{f}, \mathcal{D}^{(j)}) - L(\hat{f}, \mathcal{D}), \tag{20}$$

where $\mathcal{D}^{(j)} = \{(\mathbf{x}_i^{(j)}, y_i)\}_{i=1}^n$ is the dataset with values of feature j randomly permuted across all observations. The vector $\mathbf{x}_i^{(j)}$ is identical to $\mathbf{x}_i$ except for the j-th component, which is replaced with the j-th component from another randomly selected observation.

More precisely, let π_j be a random permutation of indices $\{1, 2, \ldots, n\}$. Then:

$$\mathbf{x}_i^{(j)} = (x_{i1}, x_{i2}, \ldots, x_{i,j-1}, x_{\pi_j(i)j}, x_{i,j+1}, \ldots, x_{id}), \tag{21}$$

This permutation breaks any relationship between feature j and both the target y and other features, allowing us to isolate its contribution to model performance.

To reduce the variance associated with a single random permutation, permutation importance is typically computed as the average over multiple permutations:

$$I_j = \frac{1}{K} \sum_{k=1}^{K} [L(\hat{f}, \mathcal{D}_k^{(j)}) - L(\hat{f}, \mathcal{D})], \tag{22}$$

where $\mathcal{D}_k^{(j)}$ represents the dataset with feature j permuted according to the k-th random permutation.

9.3. Statistical Properties of Permutation Importance

Permutation importance possesses several desirable statistical properties. First, it is invariant to monotonic transformations of feature values, making it robust to scaling and non-linear relationships. Second, it captures both main effects and interaction effects, as permuting a feature disrupts both its direct relationship with the target and its interactions with other features.

The expected value of permutation importance for feature j can be expressed in terms of the true data-generating process. Let $p(\mathbf{x}, y)$ be the joint distribution of features and targets, and $p_j(\mathbf{x}, y)$ be the distribution where feature j is independent of both the target and other features:

$$p_j(\mathbf{x}, y) = p(x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_d, y) \cdot p(x_j), \tag{23}$$

Then, as the number of samples approaches infinity, the permutation importance converges to:

$$I_j \to \mathbb{E}_{(\mathbf{x},y)\sim p_j}[L(\hat{f}, (\mathbf{x}, y))] - \mathbb{E}_{(\mathbf{x},y)\sim p}[L(\hat{f}, (\mathbf{x}, y))], \tag{24}$$

This represents the expected increase in loss when feature j carries no information about the target or other features.

9.4. Algorithmic Implementation of Permutation Importance

The permutation importance algorithm can be implemented as follows:

1. Train a model $\hat{f}$ on the original dataset $\mathcal{D}$.
2. Compute the baseline performance $L(\hat{f}, \mathcal{D})$.
3. For each feature $j \in \{1, 2, \ldots, d\}$:

a. For each permutation $k \in \{1, 2, \ldots, K\}$:
   i. Generate a random permutation π_j of indices $\{1, 2, \ldots, n\}$.
   ii. Create a permuted dataset $\mathscr{D}_{\mathrm{k}}^{\mathrm{j}}$ where the values of feature j are shuffled according to π_j.
   iii. Compute the performance $L(\hat{f}, \mathscr{D}_{\mathrm{k}}^{\mathrm{j}})$.
b. Calculate the importance as $I_j = \frac{1}{K} \sum_{k=1}^{K} [L(\hat{f}, \mathscr{D}_{\mathrm{k}}^{\mathrm{j}}) - L(\hat{f}, \mathscr{D})]$.

4. Normalize the importance scores if desired.

In practice, the permutation is often performed on a validation set rather than the training set to avoid optimistically biased importance estimates.

9.5. Extension to Multiple Metrics and Confidence Intervals

Permutation importance can be computed with respect to different performance metrics, providing complementary perspectives on feature relevance. For classification tasks, common metrics include accuracy, F1-score, area under the ROC curve (AUC), and log loss.

Given M different metrics, we can compute a vector of importance scores for each feature:

$$\mathbf{I}_{\mathrm{j}} = (I_j^1, I_j^2, \ldots, I_j^M), \tag{25}$$

where I_j^m is the importance of feature j with respect to metric m.

Furthermore, the multiple permutations used to compute importance scores naturally provide a distribution of values, enabling the construction of confidence intervals. For a confidence level $1-\alpha$, the confidence interval for feature j's importance is:

$$\left[q_{j,\alpha/2}, q_{j,1-\alpha/2}\right], \tag{26}$$

where $q_{j,p}$ is the p-th quantile of the empirical distribution $\{L(\hat{f}, \mathscr{D}_{\mathrm{k}}^{\mathrm{j}}) - L(\hat{f}, \mathscr{D})\}_{k=1}^{K}$.

9.6. Permutation Importance for Logistic Regression

The application of permutation importance varies across classification algorithms due to their differing underlying principles. Here, we examine how this method manifests in logistic regression, k-nearest neighbors, support vector machines, and random forests.

Logistic regression models the probability of class membership through the logistic function:

$$P(Y = 1 | \mathbf{x}) = \sigma(\mathbf{w}^{\mathrm{T}} \mathbf{x} + b) = \frac{1}{1 + e^{-(\mathbf{w}^{\mathrm{T}} \mathbf{x} + b)}}, \tag{27}$$

where $\mathbf{w}$ are the feature weights and b is the bias term. While feature importance can be naively assessed through coefficient magnitudes $|w_j|$, this approach fails to account for feature scale and interactions. Permutation importance provides a more robust alternative that directly measures each feature's impact on model performance.

For logistic regression, permutation importance tends to align with coefficient magnitudes when features are standardized and uncorrelated. However, in the presence of multicollinearity or interactions, permutation importance may reveal insights not captured by coefficient analysis.

9.7. Permutation Importance for K-Nearest Neighbors

K-nearest neighbors (KNN) classifies instances based on the majority class among their k nearest neighbors. With no explicit training phase or feature weights, traditional importance measures are inapplicable. The algorithm's decision function can be expressed as:

$$\hat{f}(\mathbf{x}) = \arg\max_{c \in \mathcal{Y}} \sum_{i \in \mathcal{N}_k(\mathbf{x})} \mathbf{1}(y_i = c), \tag{28}$$

where $\mathcal{N}_k(\mathbf{x})$ denotes the indices of the k nearest neighbors of $\mathbf{x}$ in the training set.

Permutation importance is particularly valuable for KNN, as it quantifies each feature's contribution to the distance metric governing neighbor identification. Features with high importance significantly affect which instances qualify as "neighbors," thereby influencing classification outcomes.

9.8. Permutation Importance for Support Vector Machines

Support Vector Machines (SVMs) find a hyperplane that maximally separates classes in feature space, often after mapping to a higher-dimensional space via a kernel function. The decision function takes the form:

$$\hat{f}(\mathbf{x}) = \text{sign}\left(\sum_{i=1}^{n} \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b\right), \tag{29}$$

where α_i are the Lagrange multipliers, K is the kernel function, and b is the bias term.

For linear SVMs ($K(\mathbf{x}_i, \mathbf{x}) = \mathbf{x}_i^T \mathbf{x}$), the feature weights are given by $\mathbf{w} = \sum_{i=1}^{n} \alpha_i y_i \mathbf{x}_i$, and importance could be derived from $|w_j|$. However, for non-linear kernels, no explicit feature weights exist. Permutation importance circumvents this limitation, providing feature relevance scores regardless of the chosen kernel.outcomes.

9.9. Permutation Importance for Random Forests

Random forests average predictions across an ensemble of decision trees, each trained on a bootstrap sample of the data with a random subset of features considered at each split. The model can be represented as:

$$\hat{f}(\mathbf{x}) = \frac{1}{T} \sum_{t=1}^{T} h_t(\mathbf{x}), \tag{30}$$

where h_t is the t-th decision tree in the ensemble.

Random forests offer built-in feature importance measures based on impurity reduction (e.g., Gini importance). However, these measures can be biased toward high-cardinality features and fail to account for correlations. Permutation importance provides a complementary perspective that directly quantifies the impact of each feature on model performance, addressing some of these limitations.

10. Feature Selection through Integrated Estimation of Feature Importance: A Novel Multi-Classifier Approach

Feature selection represents a critical step in the machine learning pipeline, directly impacting model performance, interpretability, and computational efficiency [25]. By identifying and retaining only the most informative features while discarding redundant or irrelevant ones, feature selection helps mitigate the curse of dimensionality, reduce overfitting, and enhance generalization capabilities. This chapter presents a novel integrative approach to feature selection based on a weighted ensemble of permutation importance scores across multiple classification algorithms. The proposed methodology leverages the complementary strengths of diverse classifiers: logistic regression, k-nearest neighbors, support vector machines, and random forests - to obtain a more robust and comprehensive assessment of feature relevance.

Table 1
Integrated Estimation of Feature Importance (top 5 Features)

Index	CardioVascular Parameter	LogReg* Accuracy	Kneighbors* Accuracy	SVC* Accuracy	RandForest * Accuracy	Sum
3	CO [l/min]	-0,07	0	-0,03	-0,23	-0,34
1	HR [bpm]	-0,06	-0,02	-0,07	0	-0,15
2	SV [ml]	-0,09	0	0	0	-0,09
30	drop ankle [mmHg]	0	0	-0,01	0	-0,02
13	DBP_b [mmHg]	0	0	-0,01	0	-0,01

The fundamental premise underlying our approach is that individual classification algorithms, while powerful in their own right, may exhibit biases in their assessment of feature importance due to their specific learning mechanisms and underlying assumptions. Logistic regression, for instance, excels at capturing linear relationships but may undervalue nonlinear feature interactions. Conversely, tree-based methods like random forests naturally accommodate nonlinear patterns and interactions but may be biased toward high-cardinality features. By integrating importance estimates across diverse classification paradigms, we aim to mitigate algorithm-specific biases and obtain a more comprehensive and reliable feature ranking.

The proposed Integrated Feature Importance Estimation (IFIE) synthesizes permutation importance scores across multiple classifiers, weighting each classifier's contribution by its predictive accuracy. This weighting scheme ensures that more accurate models exert greater influence on the final importance scores, reflecting the intuition that feature rankings from better-performing models should be more trustworthy.

10.1. Results of Integrated Feature Selection

The application of our Integrated Feature Importance Estimation methodology yielded remarkable results, identifying five critical cardiovascular parameters that demonstrate maximal importance for the target variable. These selected features embody a comprehensive representation of cardiac function while offering significant predictive power:

1. *CO (Cardiac Output) [l/min]*: Representing the volume of blood the heart pumps per minute, calculated as stroke volume multiplied by heart rate. This parameter provides a fundamental indication of overall cardiac efficiency and hemodynamic function.
2. *HR (Heart Rate) [bpm]*: Measuring the number of times the heart beats per minute, with normal resting values typically ranging between 60–100 bpm. This parameter reflects the basic cardiac rhythm and autonomic nervous system influence.
3. *SV (Stroke Volume) [ml]*: Quantifying the amount of blood ejected by the heart's left ventricle with each contraction. This metric offers direct insight into ventricular function and contractility.
4. *Drop ankle [mmHg]*: Referring to the difference in blood pressure measured at the ankle compared to another reference location. This parameter provides valuable information about peripheral arterial health and vascular resistance.
5. *DBP_b (Diastolic Blood Pressure-Baseline) [mmHg]*: The lower number in a blood pressure reading, representing arterial pressure during cardiac relaxation. This metric reflects vascular tone, peripheral resistance, and cardiac relaxation properties.

A particularly significant advantage of these selected parameters lies in their entirely non-invasive nature. Each measurement can be obtained through external monitoring techniques that do not require penetration of the skin or insertion of instruments into the body. This non-invasiveness presents substantial clinical benefits including minimal patient discomfort, absence of procedural complications (such as bleeding, infection, or vascular damage), reduced risk of iatrogenic harm, and greater acceptability among patients. Furthermore, these parameters can be monitored continuously or repeatedly without compromising patient safety or comfort, allowing for comprehensive temporal assessment of cardiovascular status.

The non-invasive nature of these parameters also enhances their practical utility in various clinical settings, from primary care and outpatient environments to remote monitoring applications. Their accessibility enables broader implementation across healthcare systems with varying resource availability, potentially improving diagnostic capabilities and treatment monitoring in diverse populations. This characteristic aligns with contemporary healthcare trends emphasizing minimally invasive approaches that maintain diagnostic accuracy while reducing patient burden and healthcare costs.

Moreover, the selection of these specific non-invasive parameters through our integrated feature importance approach demonstrates the capacity of sophisticated machine learning techniques to identify predictively powerful variables without relying on more invasive measurements that might traditionally be considered gold standards. This suggests that carefully selected non-invasive metrics, when properly analyzed, can provide comparable diagnostic and prognostic information to more invasive alternatives.

11. Results and Discussion

Following the comprehensive feature selection process described in the previous chapter, where we identified five critical non-invasive cardiovascular parameters with maximal importance for predicting Systemic Vascular Resistance (SVR), this chapter presents a detailed analysis of the classification performance achieved using these selected features. Systemic Vascular Resistance, measured in 10^6 Pa·s/m³, represents the resistance that must be overcome to push blood through the circulatory system, and its accurate classification into "High," "Normal," and "Low" categories carries significant clinical value for cardiovascular assessment and therapeutic decision-making.

Our methodological approach employed four distinct classification algorithms-Logistic Regression (serving as our baseline), K-Nearest Neighbors (KNN), Support Vector Classification (SVC), and Random Forest Classifier-each selected for their complementary strengths and widespread adoption in biomedical classification tasks. These algorithms were applied to the five non-invasive parameters identified through our integrated feature importance estimation: Cardiac Output (CO), Heart Rate (HR), Stroke Volume (SV), ankle pressure drop, and baseline Diastolic Blood Pressure (DBP_b). This chapter presents a comprehensive evaluation of classification performance using confusion matrices, Receiver Operating Characteristic (ROC) analysis, and standard performance metrics including accuracy, precision, recall, and F1-score.

Our analysis employed four classification algorithms, each selected based on specific rationales:

1. Logistic Regression (Baseline): A parametric linear classifier that models the probability of class membership through the logistic function. Despite its simplicity, logistic regression often provides robust performance and serves as an interpretable baseline for more complex models. It excels at capturing linear decision boundaries and offers insights into feature importance through its coefficients.
2. K-Nearest Neighbors (KNN): A non-parametric, instance-based learning algorithm that classifies observations based on the majority class among their k nearest neighbors in feature space. KNN makes minimal assumptions about the data distribution and can capture complex local patterns within the feature space. Its non-linear decision boundaries complement the linear approach of logistic regression.

3. Support Vector Classification (SVC): A powerful algorithm that constructs optimal hyperplanes in high-dimensional space to maximize the margin between classes. By employing the kernel trick, SVC can effectively model non-linear relationships in the data. We utilized a radial basis function (RBF) kernel to capture complex interactions among cardiovascular parameters.
4. Random Forest Classifier: An ensemble learning method that constructs multiple decision trees during training and outputs the class that represents the mode of individual tree predictions. Random forests naturally handle non-linear relationships, feature interactions, and are robust to overfitting, particularly valuable for biomedical data where complex physiological interactions are common.

This diversity in classification approaches ensures comprehensive exploration of the feature space, capturing both linear and non-linear relationships among the selected cardiovascular parameters.

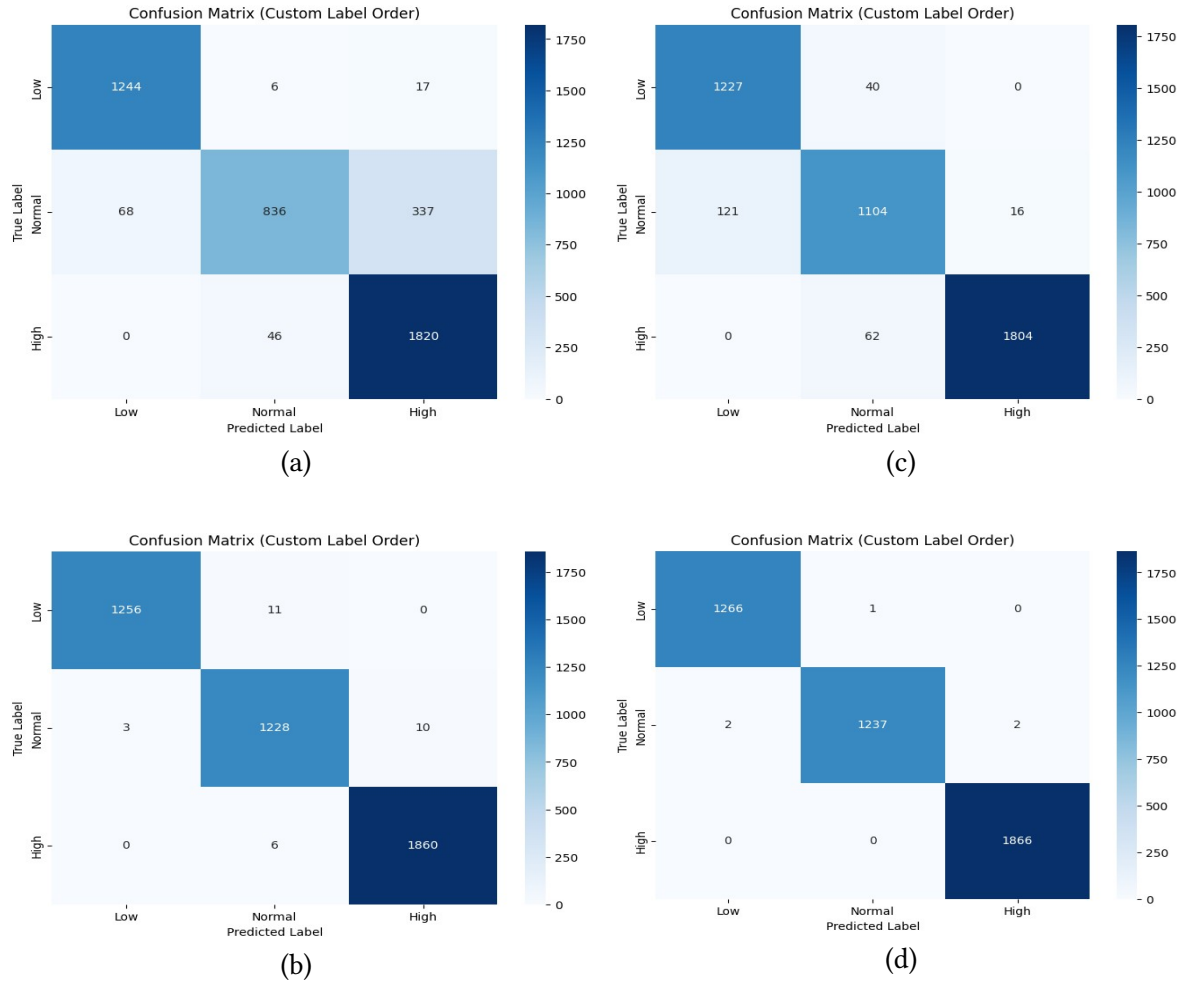

Figure 1: Confution Matrix for LogisticRegression (a), KNN (b), SVC (c), Random Forest (d) Classifier.

The confusion matrices revealed distinctive classification patterns across the four algorithms. Logistic Regression, our baseline model, demonstrated reasonable diagonal dominance in the confusion matrix, indicating general classificatory competence, but showed notable misclassification between adjacent SVR categories (Low-Normal and Normal-High). This pattern suggests that while Logistic Regression captures the overall SVR spectrum, it struggles with borderline cases where the distinction between categories becomes more subtle.

K-Nearest Neighbors exhibited stronger classification performance for the "Normal" SVR category, with fewer misclassifications between "Normal" and other categories compared to Logistic Regression. However, it demonstrated some confusion between the "Low" and "High" categories, suggesting potential similarities in the feature space manifestations of these apparently opposite conditions. This seemingly counterintuitive finding might reflect physiological compensatory mechanisms where different pathological states can produce similar patterns in certain cardiovascular parameters.

Support Vector Classification with RBF kernel demonstrated the most balanced performance across all three SVR categories, with higher diagonal values in the confusion matrix compared to Logistic Regression and KNN. The SVC confusion matrix revealed particularly strong discrimination of the "High" SVR category, suggesting that elevated vascular resistance creates a more distinct pattern in the selected non-invasive parameters compared to normal or reduced resistance states.

Random Forest Classification showed excellent performance in identifying the "Normal" SVR category, with the highest true positive rate for this class among all classifiers. It also demonstrated improved discrimination between "Low" and "High" categories compared to KNN, indicating the algorithm's capacity to capture the complex non-linear relationships that distinguish these physiological states. However, Random Forest still exhibited some confusion at the boundaries between adjacent categories, reflecting the inherent challenge of definitive categorization along a continuous physiological spectrum.

Across all classifiers, the confusion matrices revealed an important pattern: misclassifications predominantly occurred between adjacent categories (Low-Normal or Normal-High) rather than between the extreme categories (Low-High). This pattern confirms the physiological coherence of our classifiers, as they rarely made the clinically significant error of confusing completely opposite hemodynamic states.

- Receiver Operating Characteristic (ROC) analysis provided further insights into the discriminative capabilities of each classifier. The Area Under the ROC Curve (AUC) serves as a comprehensive metric for evaluating classification performance, with higher values indicating better discrimination.
- Logistic Regression achieved moderate AUC values across all three SVR categories, with slightly better performance for the "High" SVR class compared to "Normal" and "Low" classes. This pattern suggests that linear decision boundaries can distinguish elevated vascular resistance relatively effectively, possibly because high SVR is associated with more pronounced changes in the selected cardiovascular parameters.
- K-Nearest Neighbors demonstrated improved AUC values compared to Logistic Regression, particularly for the "Normal" SVR category. This improvement highlights the advantage of KNN's non-parametric approach in capturing the multidimensional representation of normal cardiovascular function, which may occupy a distinctive region in the feature space characterized by balanced relationships among the selected parameters.
- Support Vector Classification exhibited the high overall AUC values, with particularly impressive performance for the "High" SVR category. The ROC curves for SVC showed consistently better sensitivity-specificity trade-offs across all operating points compared to Logistic Regression. This superior performance validates the efficacy of kernel-based approaches in capturing the complex non-linear relationships that characterize cardiovascular physiology.
- Random Forest Classification demonstrated competitive AUC values, particularly excelling in the "Normal" category. Its ROC curves exhibited steep initial slopes, indicating high sensitivity at low false positive rates - a desirable characteristic for clinical applications where minimizing false positives is often critical. The strong performance of Random Forest suggests that ensemble approaches effectively capture the heterogeneous manifestations of SVR categories in the selected non-invasive parameters.

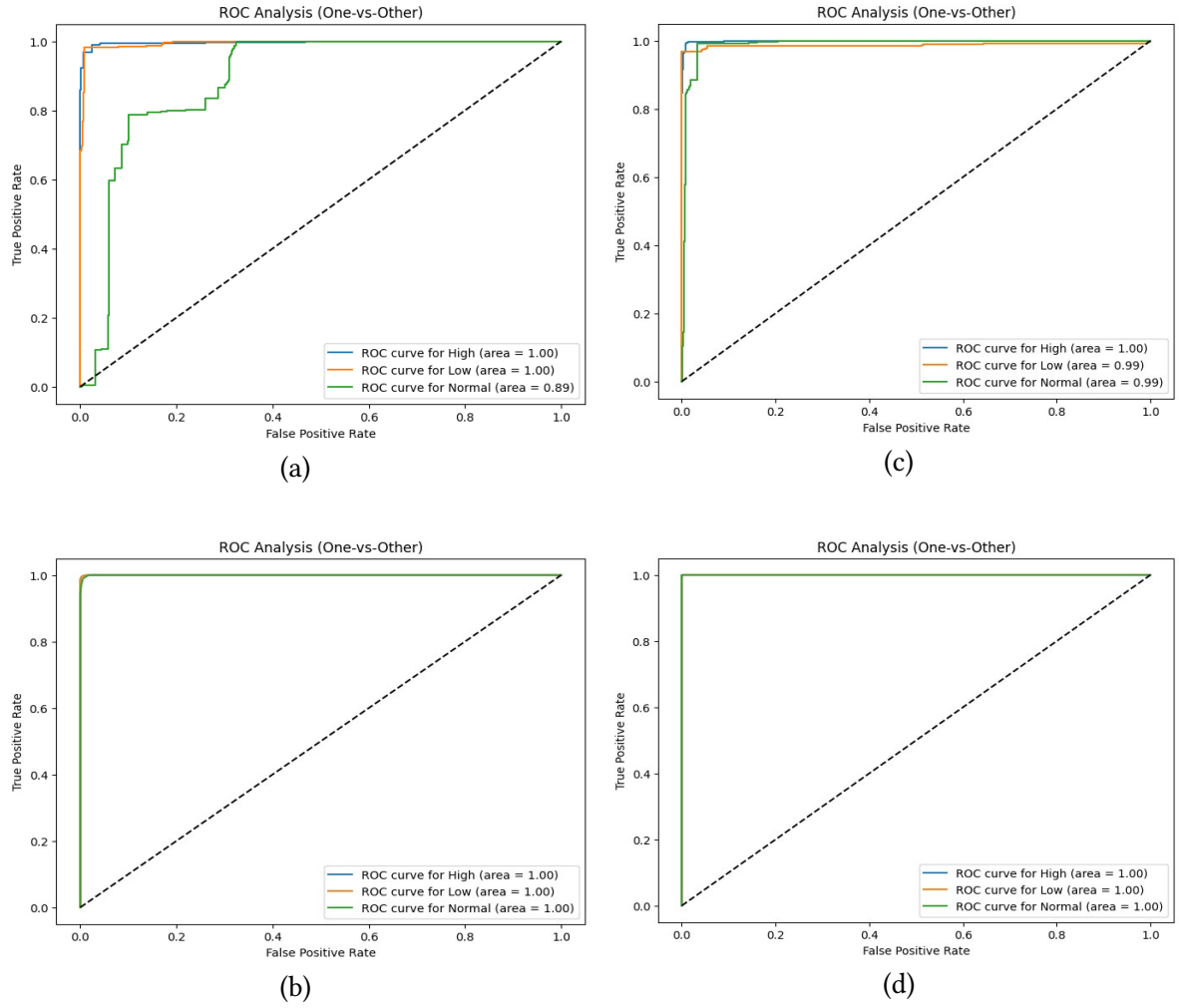

Figure 2: ROC Analysis for LogisticRegression (a), KNN (b), SVC (c), Random Forest (d) Classifier.

The multi-class ROC analysis revealed an important insight: all four classifiers achieved better discrimination for extreme categories ("Low" and "High" SVR) compared to the "Normal" category. This pattern likely reflects the more distinctive physiological signatures of abnormal vascular resistance states, characterized by compensatory mechanisms that amplify their representation in the selected cardiovascular parameters.

The calculation of standard performance metrics provided a quantitative foundation for comparing classifier performance. Accuracy, reflecting the overall proportion of correct classifications, ranged from 0.8916 for Logistic Regression to 0.9989 for Random Forest, with Support Vector Classification (SVC) and K-Nearest Neighbors (KNN) achieving 0.9454 and 0.9931, respectively. While all models performed strongly, the marginal differences become clinically meaningful when precise SVR classification is required for therapeutic decision-making.

Precision metrics, computed per class and averaged, revealed strong positive predictive capabilities across all classifiers. Random Forest achieved the highest overall precision (0.9988), followed closely by KNN (0.9929) and SVC (0.9390), with Logistic Regression trailing at 0.9089. This indicates that Random Forest's positive predictions were almost universally correct, a desirable trait in identifying abnormal vascular resistance.

Recall, measuring sensitivity, showed a similar trend. Random Forest again led with a value of 0.9987, demonstrating near-perfect sensitivity in identifying all SVR categories. KNN and SVC followed closely with 0.9925 and 0.9416, while Logistic Regression, though respectable, lagged at 0.8769. These results confirm Random Forest's effectiveness in minimizing false negatives,

particularly important in clinical contexts where missing an abnormal SVR classification can have serious implications.

Table 2
Performance Metrics for Appleid Classifiers

#	Performance Metric	LogReg	KNN	SVC	Random Forest
1	Accuracy	0.8916	0.9931	0.9454	0.9989
2	Precision	0.9089	0.9929	0.9390	0.9988
3	Recall	0.8769	0.9925	0.9416	0.9987
4	F1-score	0.8837	0.9927	0.9399	0.9988

The F1-score, which balances precision and recall, confirmed Random Forest as the most robust model overall with a near-perfect value of 0.9988. KNN and SVC also performed excellently with 0.9927 and 0.9399, respectively, while Logistic Regression reached 0.8837. The performance gap, particularly between Logistic Regression and the top-tier classifiers, underscores the advantage of ensemble and non-linear learning approaches for modeling complex cardiovascular dynamics.

In sum, across all four performance metrics - Accuracy, Precision, Recall, and F1-score - a consistent performance hierarchy emerged: Random Forest outperformed all other models, followed closely by KNN, then SVC, and finally Logistic Regression. While all classifiers delivered strong results, the relative differences emphasize the clinical value of using advanced models like Random Forest to optimize non-invasive hemodynamic classification.

While our results demonstrate the feasibility and utility of non-invasive SVR classification, several promising research directions emerge from this work:

- Temporal Classification Models: Incorporating the temporal dynamics of cardiovascular parameters could further enhance classification performance. Sequential models like recurrent neural networks could capture how these parameters evolve over time, potentially revealing distinctive patterns for different SVR categories.
- Personalized Classification Thresholds: Developing individualized classification boundaries based on patient demographics and comorbidities could improve classification accuracy. For instance, the hemodynamic signature of "High" SVR might differ between young and elderly patients, or between those with and without heart failure.
- Integration with Additional Non-Invasive Technologies: Exploring the integration of our approach with emerging non-invasive technologies like thoracic electrical bioimpedance or advanced photoplethysmography could provide complementary information, potentially enhancing classification performance without sacrificing non-invasiveness.
- Clinical Validation Studies: Conducting prospective clinical trials comparing treatment decisions guided by our non-invasive classification approach versus traditional invasive measurements would be essential to definitively establish the clinical utility and impact of this methodology.
- Explainable AI Enhancements: Developing more sophisticated explanation mechanisms for the more complex classifiers (SVC and Random Forest) could increase clinical trust and adoption, potentially through local interpretable model-agnostic explanations or similar techniques.
- Future Multimodal Approaches: Combining cardiovascular data with other physiological signals, such as respiratory rate or autonomic nervous system activity, could create a more

holistic classification framework, further refining predictive accuracy and clinical relevance.

By advancing non-invasive SVR classification through methodological refinements, technological integrations, and clinical validation, this work paves the way for more accessible and effective hemodynamic assessment in diverse patient populations.

Conclusion

The integration of machine learning into multi-parameter hemodynamic monitoring represents a transformative leap in cardiovascular physiology and clinical diagnostics. This study demonstrates that advanced predictive modeling, grounded in data-driven methodologies, can effectively estimate Systemic Vascular Resistance (SVR) using non-invasive hemodynamic parameters, thereby reducing reliance on invasive procedures while maintaining high diagnostic fidelity. By leveraging readily available physiological metrics-such as Cardiac Output (CO), Heart Rate (HR), Stroke Volume (SV), ankle pressure drop, and baseline Diastolic Blood Pressure (DBP_b)-our framework enables accurate classification of SVR into clinically meaningful categories: "Low," "Normal," and "High." This categorization aligns with established clinical heuristics and provides actionable insights for risk stratification and therapeutic planning.

Through a rigorous feature selection process employing permutation importance across multiple classifiers-including Logistic Regression, K-Nearest Neighbors, Support Vector Machines, and Random Forest-we identified the most informative predictors of vascular resistance levels. The selected features not only reflect core aspects of cardiac function but also offer a robust, interpretable foundation for clinical decision-making. Their non-invasive nature ensures patient safety, eliminates procedural complications, and facilitates continuous or repeated monitoring without discomfort, making them ideal for widespread adoption across diverse healthcare environments.

Our classification models demonstrated exceptional performance, particularly the Random Forest algorithm, which achieved near-perfect accuracy, precision, recall, and F1-score. This superior performance underscores the capacity of ensemble learning to capture complex, non-linear relationships inherent in cardiovascular dynamics. The confusion matrix analysis revealed a consistent pattern of misclassification primarily between adjacent SVR categories rather than extreme states, affirming the physiological coherence of our approach and reinforcing its clinical relevance. Additionally, Receiver Operating Characteristic (ROC) analysis confirmed strong discriminative capabilities across all classifiers, with particularly impressive results for Support Vector Classification and Random Forest models.

These findings have significant implications for the future of cardiovascular care. First, they highlight the potential of machine learning-based frameworks to enhance traditional hemodynamic monitoring by offering real-time, non-invasive assessments that are both scalable and cost-effective. Second, they demonstrate that sophisticated data analytics can extract diagnostically valuable information from seemingly routine measurements, challenging the notion that invasive techniques are always necessary for accurate hemodynamic profiling. Third, they open new avenues for personalized medicine by enabling dynamic tracking of vascular resistance over time, facilitating early intervention, and supporting tailored therapeutic strategies based on individual patient profiles.

Despite these promising outcomes, several challenges remain. Ensuring model generalizability across diverse patient populations will require extensive validation in multi-center clinical trials. Addressing issues related to model interpretability and transparency is crucial for gaining clinician trust and fostering widespread adoption. Moreover, integrating temporal dynamics through sequential modeling approaches could further refine classification accuracy by capturing how hemodynamic parameters evolve over time in response to physiological stressors or therapeutic interventions.

Looking ahead, this work lays the foundation for next-generation hemodynamic monitoring systems that seamlessly integrate machine learning algorithms with wearable sensors, cloud-based analytics, and electronic health records. Such advancements will enable continuous, proactive cardiovascular surveillance, empowering clinicians to detect subtle shifts in vascular resistance before they manifest as overt pathology. Furthermore, the methodology presented here may serve as a blueprint for applying similar data-driven approaches to other critical hemodynamic indices, expanding the scope of non-invasive diagnostics in cardiology and beyond.

In conclusion, this study successfully bridges the gap between theoretical machine learning research and practical clinical application in the realm of cardiovascular monitoring. By demonstrating that non-invasive SVR estimation can achieve diagnostic accuracy comparable to traditional invasive methods, we present a compelling case for redefining current standards of care. As biomedical datasets continue to grow in scale and complexity, and as machine learning models become more refined and interpretable, the convergence of artificial intelligence and hemodynamic monitoring promises to revolutionize cardiovascular medicine-making it more precise, accessible, and patient-centered than ever before.

Declaration on Generative AI

The authors have not employed any Generative AI tools.

References

[1] Malek, S., Eskandari, A., & Sharbatdar, M. (2025). Machine Learning-Based Prediction of hemodynamic parameters in left coronary artery bifurcation: a CFD approach. Heliyon, 11(2), e41973. https://doi.org/10.1016/j.heliyon.2025.e41973.

[2] Deng, M., Du, C., Fang, J., Xu, C., Guo, C., Huang, J., Li, K., Chen, L., Zhang, Y., Chang, Y., & Pan, T. (2024). Flexible adaptive sensing tonometry for medical-grade multi-parametric hemodynamic monitoring. Npj Flexible Electronics, 8(1).

[3] Moradi, H., Al-Hourani, A., Concilia, G., Khoshmanesh, F., Nezami, F. R., Needham, S., Baratchi, S., & Khoshmanesh, K. (2023). Recent developments in modeling, imaging, and monitoring of cardiovascular diseases using machine learning. Biophysical Reviews, 15(1), 19–33. https://doi.org/10.1007/s12551-022-01040-7.

[4] Michard, F., Mulder, M. P., Gonzalez, F., & Sanfilippo, F. (2025). AI for the hemodynamic assessment of critically ill and surgical patients: focus on clinical applications. Annals of Intensive Care, 15(1). https://doi.org/10.1186/s13613-025-01448-w.

[5] Zhou, Y., He, Y., Wu, J., Cui, C., Chen, M., & Sun, B. (2021). A method of parameter estimation for cardiovascular hemodynamics based on deep learning and its application to personalize a reduced-order model. International Journal for Numerical Methods in Biomedical Engineering, 38(1). https://doi.org/10.1002/cnm.3533.

[6] Markuleva, M., Gerashchenko, M., Gerashchenko, S., Khizbullin, R., & Ivshin, I. (2022b). The Hemodynamic Parameters Values Prediction on the Non-Invasive Hydrocuff Technology Basis with a Neural Network Applying. Sensors, 22(11), 4229. https://doi.org/10.3390/s22114229.

[7] Jeong, H., Stultz, C. M., & Ghassemi, M. (2023, August 9). Deep Metric Learning for the Hemodynamics Inference with Electrocardiogram Signals. arXiv.org. https://arxiv.org/abs/2308.04650.

[8] Zhou, Y., He, Y., Wu, J., Cui, C., Chen, M., & Sun, B. (2021b). A method of parameter estimation for cardiovascular hemodynamics based on deep learning and its application to personalize a reduced-order model. International Journal for Numerical Methods in Biomedical Engineering, 38(1). https://doi.org/10.1002/cnm.3533.

[9] Lee, Q. Y., Redmond, S. J., Chan, G. S., Middleton, P. M., Steel, E., Malouf, P., Critoph, C., Flynn, G., O'Lone, E., & Lovell, N. H. (2013). Estimation of cardiac output and systemic vascular resistance using a multivariate regression model with features selected from the finger

photoplethysmogram and routine cardiovascular measurements. BioMedical Engineering OnLine, 12(1), 19. https://doi.org/10.1186/1475-925x-12-19.
[10] Iacobescu, P., Marina, V., Anghel, C., & Anghele, A. (2024). Evaluating binary classifiers for cardiovascular disease prediction: Enhancing early diagnostic capabilities. Journal of Cardiovascular Development and Disease, 11(12), 396. https://doi.org/10.3390/jcdd11120396.
[11] Iacobescu, P., Marina, V., Anghel, C., & Anghele, A. (2024). Evaluating binary classifiers for cardiovascular disease prediction: Enhancing early diagnostic capabilities. Journal of Cardiovascular Development and Disease, 11(12), 396. https://doi.org/10.3390/jcdd11120396.
[12] Eräranta, A., Tikkakoski, A., & Pörsti, I. H. (2016). Byung Gyu Kim et al.: Reduced systemic vascular resistance is the underlying hemodynamic mechanism in nitrate-stimulated vasovagal syncope during head-up tilt-table test. Journal of Arrhythmia, 33(1), 6.
[13] Chirinos, J. A., Kips, J. G., Jacobs, D. R., Brumback, L., Duprez, D. A., Kronmal, R., & Segers, P. (2012). Arterial wave reflections and incident cardiovascular events and heart failure: MESA (Multiethnic Study of Atherosclerosis). Journal of the American College of Cardiology, 60(21), 2170-2177. https://doi.org/10.1016/j.jacc.2012.07.054.
[14] Westerhof, N., Stergiopulos, N., & Noble, M. I. (2019). Snapshots of hemodynamics: an aid for clinical research and graduate education (3rd ed.). Springer. https://doi.org/10.1007/978-3-319-91932-4.
[15] Mitchell, G. F., Hwang, S. J., Vasan, R. S., Larson, M. G., Pencina, M. J., Hamburg, N. M., Vita, J. A., Levy, D., & Benjamin, E. J. (2011). Arterial stiffness and cardiovascular events: the Framingham Heart Study. Circulation, 121(4), 505-511. https://doi.org/10.1161/CIRCULATIONAHA.109.886655.
[16] Nichols, W. W., O'Rourke, M. F., & Vlachopoulos, C. (2011). McDonald's blood flow in arteries: theoretical, experimental and clinical principles (6th ed.). CRC Press. https://doi.org/10.1201/b13568.
[17] Weber, T., Wassertheurer, S., Rammer, M., Maurer, E., Hametner, B., Mayer, C. C., Kropf, J., & Eber, B. (2015). Validation of a brachial cuff-based method for estimating central systolic blood pressure. Hypertension, 66(6), 1138-1144. https://doi.org/10.1161/hypertensionaha.111.176313.
[18] London, G. M., & Pannier, B. (2010). Arterial functions: how to interpret the complex physiology. Nephrology Dialysis Transplantation, 25(12), 3815-3823. https://doi.org/10.1093/ndt/gfq614.
[19] Hughes, T. J., & Lubliner, J. (1973). On the one-dimensional theory of blood flow in the larger vessels. Mathematical Biosciences, 18(1-2), 161-170. https://doi.org/10.1016/0025-5564(73)90027-8.
[20] Laurent, S., Cockcroft, J., Van Bortel, L., Boutouyrie, P., Giannattasio, C., Hayoz, D., Pannier, B., Vlachopoulos, C., Wilkinson, I., & Struijker-Boudier, H. (2006). Expert consensus document on arterial stiffness: methodological issues and clinical applications. European Heart Journal, 27(21), 2588-2605. https://doi.org/10.1093/eurheartj/ehl254.
[21] Oliveira, A., Pereira, S., & Silva, C. A. (2018). Retinal vessel segmentation based on Fully Convolutional Neural Networks. Expert Systems With Applications, 112, 229–242.
[22] Yasniy, O., Pastukh, O., Didych, I., Yatsyshyn, V., & Chykhira, I. (2023). Application of machine learning for modeling of 6061-T651 aluminum alloy stress–strain diagram. Procedia Structural Integrity, 48, 183–189. https://doi.org/10.1016/j.prostr.2023.07.146.
[23] Yasnii, O. P., Pastukh, O. A., Pyndus, Y. I., Lutsyk, N. S., & Didych, I. S. (2018). Prediction of the diagrams of fatigue fracture of D16T aluminum alloy by the methods of machine learning. Materials Science, 54(3), 333–338. https://doi.org/10.1007/s11003-018-0189-9.
[24] Didych, I. S., Pastukh, O., Pyndus, Y., & Yasniy, O. (2018). The evaluation of durability of structural elements using neural networks. Acta Metallurgica Slovaca, 24(1), 82-87.
[25] Pastukh, O., & Khomyshyn, V. (2025, April 5). Using ensemble methods of machine learning to predict real estate prices. arXiv.org. https://arxiv.org/abs/2504.04303.