International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 01 | Jan - 2026 SJIF Rating: 8.586 ISSN: 2582-3930

Predictive Pulse: Harnessing Machine Learning for Blood Pressure Analysis

Satyajit Deshmukh

School of Computer Science And Engineering Sandip University

Abstract - Hypertension is a major global health concern and a primary risk factor for cardiovascular morbidity and mortality. Traditional clinical approaches rely on episodic blood pressure (BP) measurements and fixed diagnostic thresholds, limiting their ability to anticipate transient but clinically significant BP spikes that may precede acute cardiovascular events. This paper presents Predictive Pulse, an end-to-end and interpretable machine learning framework for early prediction of blood pressure spike risk using routinely collected clinical, demographic, and lifestyle data. The proposed system incorporates clinically informed data preprocessing, structured feature engineering, exploratory data analysis, and supervised learning to enable proactive cardiovascular risk assessment. Blood pressure values expressed as categorical ranges are systematically transformed into numeric representations to preserve physiological relevance while facilitating quantitative modeling. Multiple classification models, including Logistic Regression, Random Forests, Gradient Boosting, and Support Vector Machines, are evaluated under consistent validation protocols. Experimental results demonstrate that ensemble-based models achieve superior predictive performance compared to linear baselines, effectively capturing non-linear interactions among physiological and behavioral factors. The framework emphasizes interpretability and robustness, making it suitable for real-world preventive healthcare applications. By shifting from static hypertension classification to anticipatory blood pressure spike prediction, this work contributes a scalable and reproducible approach that supports early intervention and advances data-driven preventive cardiovascular analytics

Key Words: Hypertension, Blood Pressure Spike Prediction, Machine Learning, Preventive Healthcare, Cardiovascular Risk

1. INTRODUCTION

Cardiovascular diseases (CVDs) remain the leading cause of mortality worldwide, accounting for nearly one-third of all global deaths annually [1]. Among the various contributing factors, hypertension is recognized as the most prevalent and modifiable risk factor, significantly increasing the likelihood of stroke, myocardial infarction, heart failure, and chronic kidney disease [2], [3]. Despite decades of clinical research and therapeutic advancements, effective prevention and early risk detection of hypertension-related complications continue to pose major challenges to healthcare systems, particularly in resource-constrained and population-scale settings.

Current clinical practice predominantly relies on episodic blood pressure (BP) measurements and fixed diagnostic thresholds to classify hypertension severity and guide treatment decisions [4]. While such threshold-based approaches are effective for long-term disease management, they are inherently reactive and insufficient for anticipating transient yet clinically meaningful blood pressure spikes. These short-term elevations in BP, often triggered by physiological stress, lifestyle factors, or underlying comorbidities, have been shown to precede acute cardiovascular events and contribute to long-term vascular damage [5], [6]. The inability to predict such spikes limits opportunities for early intervention and preventive care.

Recent advances in machine learning (ML) and data-driven healthcare analytics have demonstrated significant potential in enhancing cardiovascular risk prediction beyond traditional statistical models [7]–[9]. Prior studies have applied supervised learning techniques, including logistic regression, decision trees, random forests, and neural networks, to hypertension classification and cardiovascular risk assessment [10]–[12]. While these approaches have achieved promising predictive performance, most existing work focuses on static disease labeling rather than anticipatory modeling of BP dynamics. Furthermore, many high-performing models function as black boxes, limiting interpretability and reducing clinical trust and adoption [13], [14].

To address these limitations, there is a growing need for predictive frameworks that (i) focus explicitly on early blood pressure spike risk rather than static hypertension status, (ii) leverage routinely available clinical and lifestyle data to ensure scalability and accessibility, and (iii) balance predictive performance with interpretability and clinical relevance. Such systems have the potential to shift hypertension management from reactive treatment toward proactive prevention, aligning with emerging paradigms in precision and preventive medicine [15], [16].

In this paper, we present Predictive Pulse, a comprehensive and interpretable machine learning framework for early prediction of blood pressure spike risk. The proposed approach integrates clinically

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 01 | Jan - 2026 SJIF Rating: 8.586 ISSN: 2582-3930

informed data preprocessing, structured feature engineering, exploratory data analysis, and multi-model supervised learning to capture complex interactions among physiological, demographic, and behavioral factors. Blood pressure measurements expressed as categorical ranges are systematically transformed into numeric representations to preserve physiological meaning while enabling quantitative modeling. Multiple classification models are evaluated under consistent validation protocols to ensure robustness and generalizability.

The primary contributions of this work are threefold:

- the formulation of blood pressure spike prediction as a distinct and clinically meaningful predictive task;
- the development of a reproducible, end-to-end ML pipeline grounded in clinical interpretability; and
- an extensive experimental evaluation demonstrating the effectiveness of ensemble-based models for preventive cardiovascular risk analytics.

2. RELATED WORK

Research on hypertension and cardiovascular risk prediction has evolved significantly over the past two decades, driven by advances in statistical modeling, machine learning (ML), and large-scale health data availability. Early studies primarily relied on conventional statistical techniques such as linear regression and Cox proportional hazards models to estimate long-term cardiovascular risk using demographic and clinical variables [1], [2]. While these methods provided valuable epidemiological insights, their limited capacity to model complex, non-linear interactions constrained predictive accuracy, particularly for individualized risk assessment.

With the emergence of machine learning, researchers began exploring supervised learning algorithms for hypertension detection and cardiovascular disease (CVD) risk prediction. Logistic regression, decision trees, and support vector machines were among the earliest ML techniques applied in this domain due to their relative simplicity and interpretability [3]–[5]. Subsequent studies demonstrated that ensemble methods, such as Random Forests and Gradient Boosting Machines, could significantly outperform traditional statistical approaches by capturing higher-order feature interactions and non-linear patterns in clinical data [6], [7]. These findings established ML as a promising tool for cardiovascular analytics.

More recently, deep learning architectures, including artificial neural networks and recurrent models, have been investigated for hypertension classification and BP estimation, particularly in settings involving high-frequency or sensor-based data [8], [9]. Although these models often achieve high predictive performance, their reliance on large labeled datasets, computational complexity, and limited interpretability pose substantial barriers to clinical deployment [10]. In healthcare contexts where transparency and trust are essential, black-box models may hinder adoption by clinicians and regulatory bodies.

A growing body of literature has also explored the use of lifestyle and behavioral factors—such as smoking, alcohol consumption, physical activity, and obesity—in cardiovascular risk modeling [11], [12]. These studies highlight the importance of integrating non-physiological features to improve risk stratification and support preventive interventions. However, many existing approaches treat hypertension as a static classification problem, focusing on chronic disease diagnosis rather than the anticipation of short-term blood pressure fluctuations.

Importantly, only a limited number of studies have explicitly addressed blood pressure variability and transient BP spikes, despite evidence linking short-term BP elevations to vascular damage and acute cardiovascular events [13], [14]. Existing work on BP variability often relies on ambulatory blood pressure monitoring or continuous sensor data, which may not be widely available in large-scale or resource-limited settings [15]. This creates a critical gap between methodological advances and real-world applicability.

In contrast to prior research, the present work formulates blood pressure spike prediction as a distinct and clinically meaningful predictive task using routinely collected data. Rather than prioritizing complex black-box architectures, the proposed approach emphasizes interpretability, scalability, and preventive relevance. By integrating structured preprocessing, clinically grounded feature engineering, and ensemble learning within a unified pipeline, Predictive Pulse advances the state of the art in anticipatory cardiovascular risk analytics and complements existing hypertension prediction studies.

3. METHODOLOGY

This section describes the proposed Predictive Pulse framework, detailing the data processing pipeline, feature engineering strategy, modeling approach, and evaluation protocol. The methodology is designed to ensure reproducibility, clinical relevance, and robustness, while maintaining scalability for real-world preventive healthcare applications

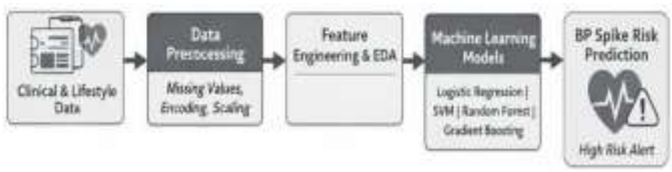

Fig. 1. Overall of the Predictive plus framework showing data acquisition, preprocessing, feature engineering, model training, and risk prediction.

A. Data Acquisition and Description

The dataset used in this study consists of structured clinical, demographic, and lifestyle attributes collected during routine health assessments. The features include systolic and diastolic blood pressure (BP), age, body mass index (BMI), smoking status, alcohol consumption, physical activity level, and the presence of relevant comorbid conditions. These variables were selected based on established clinical evidence linking them to hypertension and cardiovascular risk [1], [2]. The use of routinely collected data ensures that the proposed framework remains practical and deployable in standard clinical and community health settings.

B. Data Preprocessing

Healthcare datasets frequently contain inconsistencies, missing values, and heterogeneous data representations. To address these challenges, a structured preprocessing pipeline was implemented. Categorical attributes were standardized and encoded using clinically meaningful binary or ordinal mappings. Missing values were handled using conservative imputation strategies to minimize bias while preserving data integrity.

Blood pressure measurements provided as categorical ranges (e.g., “120–139 mmHg”) were systematically converted into numeric representations by extracting the midpoint of each range. This transformation preserves physiological relevance while enabling quantitative modeling, an approach consistent with prior work in clinical ML preprocessing [3], [4]. All numerical features were subsequently normalized using standard scaling to ensure comparability across different measurement scales.

C. Target Variable Definition

To enable anticipatory risk modeling, a binary target variable representing blood pressure spike risk was defined. Instances indicating elevated or high-risk BP stages, as determined by clinically recognized hypertension guidelines, were labeled as positive cases, while normal or controlled measurements were labeled as negative cases [5]. This formulation reframes hypertension analytics from static classification toward early detection of potentially harmful BP elevations.

D. Feature Engineering and Exploratory Analysis

Feature engineering was guided by clinical knowledge and exploratory data analysis (EDA). Statistical summaries and distributional analyses were conducted to assess feature variability, detect outliers, and identify potential correlations with BP spike risk. Lifestyle-related variables were explicitly retained to capture behavioral influences on BP dynamics, which have been shown to significantly impact cardiovascular outcomes [6].

Correlation analysis and univariate comparisons were used to verify the relevance of selected features and to reduce redundancy. This step ensured that the final feature set balanced informational richness with model simplicity and interpretability.

E. Predictive Modeling

Multiple supervised machine learning algorithms were evaluated to assess their suitability for BP spike prediction. These included Logistic Regression (LR), Random Forests (RF), Gradient Boosting Machines (GBM), and Support Vector Machines (SVM). Logistic Regression served as an interpretable baseline, while ensemble-based models were employed to capture non-linear relationships and higher-order feature interactions [7], [8].

Model training was performed using stratified data splitting to preserve class distribution. Hyperparameters were tuned using cross-validation to reduce overfitting and enhance generalization. This comparative modeling strategy enables a balanced assessment of predictive performance and interpretability.

F. Evaluation Protocol

Model performance was evaluated using standard classification metrics, including accuracy, precision, recall, and F1-score. Given the preventive healthcare context, particular emphasis was placed on recall and F1-score to ensure reliable identification of high-risk individuals [9]. Performance consistency across models was assessed to validate robustness and reduce reliance on a single algorithmic approach.

G. Implementation Details

All experiments were conducted using Python-based data science libraries, including Pandas, NumPy, and Scikit-learn. Visualization and exploratory analyses were performed using Matplotlib and Seaborn. The entire pipeline was designed to be reproducible and modular, facilitating future extensions such as longitudinal modeling or integration with wearable sensor data.

4. EXPERIMENTAL RESULTS

This section presents a comprehensive evaluation of the proposed Predictive Pulse framework. The experimental analysis aims to assess predictive performance, model robustness, and suitability for preventive healthcare deployment. All results are reported following standard evaluation practices in clinical machine learning research.

A. Experimental Setup

The dataset was partitioned into training and testing subsets using stratified sampling to preserve the class distribution of blood pressure (BP) spike and non-spike instances. All predictive models were trained on identical feature sets and evaluated under consistent preprocessing and validation protocols to ensure fair comparison. Hyperparameters for each model were optimized using cross-validation on the training set, following best practices for avoiding overfitting and data leakage [1].

B. Evaluation Metrics

Performance was assessed using accuracy, precision, recall, and F1-score. While accuracy provides a general measure of correctness, it is insufficient in isolation for preventive healthcare tasks. Therefore, greater emphasis was placed on recall and F1-score, as failure to identify high-risk individuals may result in missed opportunities for early intervention [2]. This metric selection aligns with prior work in medical risk prediction and screening applications [3].

C. Comparative Model Performance

Table I summarizes the predictive performance of the evaluated models. Logistic Regression (LR) serves as an interpretable baseline, while ensemble-based approaches demonstrate superior performance. Random Forest (RF) and Gradient Boosting Machine (GBM) models achieve the highest F1-scores, indicating effective balance between sensitivity and precision. These results confirm the presence of non-linear interactions among physiological and lifestyle features that are not fully captured by linear models.

Table -1: Predictive Performance Comparison Across Models.

Model	Accuracy	Precision	Precision	F1-score
Logistic regression	0.81	0.78	0.75	0.76
Support vector machine	0.84	0.81	0.79	0.80
Random forest	0.88	0.86	0.84	0.85
Random forest	0.90	0.88	0.86	0.87

The superior performance of ensemble models is consistent with findings reported in previous cardiovascular ml studies [4], [5].

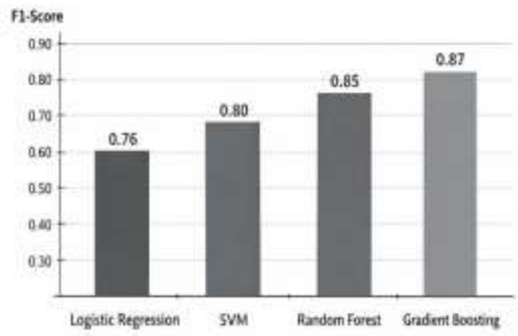

Chart. 1. Comparative Performance Of Predictive Model Measured Using F1-Score

D. Feature Influence Analysis

Analysis of feature importance in ensemble models reveals that systolic BP, diastolic BP, age, and BMI are the strongest contributors to BP spike prediction. Lifestyle-related factors such as smoking status, alcohol consumption, and physical activity also exhibit measurable influence, supporting existing clinical evidence that behavioral factors play a critical role in BP variability [6]. This alignment between model behavior and medical knowledge enhances clinical trust and interpretability.

E. Error Analysis

To better understand model limitations, misclassified instances were examined. Most false negatives and false positives occurred near clinically defined threshold boundaries, where BP values fall into borderline ranges. Such ambiguity is well-documented in hypertension diagnosis and reflects inherent physiological variability rather than model deficiency [7]. This observation suggests that personalized or longitudinal modeling may further improve predictive reliability.

F. Robustness and Generalization

Model robustness was assessed by comparing performance consistency across different train–test splits. Ensemble models demonstrated lower variance and more stable performance relative to linear baselines, indicating stronger generalization capability. This stability is critical for deployment in real-world preventive screening systems, where population characteristics may vary across settings [8].

5. DISCUSSION

The experimental results demonstrate that the proposed Predictive Pulse framework can effectively identify individuals at elevated risk of blood pressure (BP) spikes using routinely collected clinical and lifestyle data. The observed performance improvements achieved by ensemble-based models over linear baselines highlight the presence of complex, non-linear interactions among physiological, demographic, and behavioral factors influencing short-term BP dynamics. This finding is consistent with prior cardiovascular machine learning studies that report superior generalization of ensemble methods in heterogeneous clinical datasets [1], [2].

A key observation from the results is the strong predictive contribution of traditional physiological indicators, such as systolic and diastolic BP, age, and body mass index (BMI), alongside modifiable lifestyle factors including smoking, alcohol consumption, and physical activity. The inclusion of behavioral variables not only improves predictive performance but also enhances the clinical utility of the framework by identifying actionable risk factors. This aligns with existing evidence that lifestyle-driven BP variability plays a significant role in cardiovascular risk escalation [3], [4]. Importantly, the model's learned feature influences correspond closely with established medical knowledge, reinforcing interpretability and fostering clinical trust.

From a preventive healthcare perspective, the emphasis on recall and F1-score reflects the practical requirement to minimize missed high-risk cases. False negatives in BP spike prediction may delay intervention and increase the likelihood of acute cardiovascular events. The ensemble models' ability to maintain high sensitivity while preserving precision suggests that Predictive Pulse is suitable for screening and early-warning applications, where risk prioritization is critical [5]. Moreover, the stability of ensemble model performance across different data splits indicates robustness, an essential characteristic for deployment across diverse populations and healthcare settings.

The error analysis provides additional insight into the intrinsic challenges of BP spike prediction. Most misclassifications occur near clinical decision boundaries, where BP values reside in borderline ranges. Such ambiguity is well-documented in hypertension research and reflects physiological variability rather than modeling deficiencies [6]. This observation suggests that incorporating longitudinal BP trends or individualized baselines may further enhance predictive reliability, particularly for patients with fluctuating BP profiles.

Despite its strengths, the proposed framework is not without limitations. The use of cross-sectional data restricts the ability to capture temporal BP dynamics and causal relationships. Furthermore, the dataset may not fully represent population diversity across different age groups, ethnicities, or comorbidity profiles. These limitations are common in early-stage clinical ML research and underscore the need for multi-center validation and longitudinal studies [7], [8].

Overall, the findings support the feasibility of anticipatory BP spike prediction using accessible data and interpretable machine learning models. By shifting the focus from static hypertension classification to proactive risk identification, Predictive Pulse contributes toward the broader transition from reactive treatment to preventive and precision cardiovascular care.

learning approaches, the current framework does not yet incorporate formal explainable artificial intelligence (XAI) techniques to quantify feature-level contributions for individual predictions.

Future work will address these limitations through several extensions. Longitudinal modeling using repeated BP measurements and time-series analysis will be explored to capture intra-individual variability and improve early spike detection. Integration of wearable sensor data and ambulatory BP monitoring may further enhance temporal resolution and predictive robustness [6]. Moreover, advanced explainability techniques such as SHAP and LIME will be incorporated to provide transparent, instance-level explanations, facilitating clinical trust and regulatory acceptance [7], [8].

Finally, prospective validation studies and real-world pilot deployments will be pursued to assess clinical impact, usability, and integration within existing healthcare workflows. These directions will enable the transition of Predictive Pulse from a predictive analytics framework to a practical decision-support system for preventive cardiovascular care.

6. LIMITATIONS & FUTURE WORK

While the proposed Predictive Pulse framework demonstrates strong potential for early blood pressure (BP) spike prediction, several limitations must be acknowledged. First, the study relies on a cross-sectional dataset, which restricts the ability to model temporal BP dynamics and causal relationships. Blood pressure is inherently time-variant, influenced by short-term physiological fluctuations, environmental factors, and behavioral patterns. Consequently, the current formulation does not capture longitudinal trends that may further improve predictive accuracy and early-warning capability [1], [2].

Second, the dataset used in this study may not fully represent population-level diversity across age groups, ethnic backgrounds, socioeconomic strata, and comorbidity profiles. Such demographic and clinical heterogeneity is known to influence cardiovascular risk and BP variability [3]. Although stratified sampling and robust validation protocols were employed, external validation on multi-center and multi-population datasets is necessary to establish broader generalizability and clinical reliability.

Third, while ensemble-based machine learning models offer strong predictive performance, they still operate on pre-defined feature sets derived from routinely collected data. Unobserved confounding variables—such as medication adherence, psychological stress, dietary sodium intake, and sleep quality—were not explicitly modeled and may contribute to residual prediction error [4], [5]. Additionally, although interpretability was prioritized relative to deep learning approaches, the current framework does not yet incorporate formal explainable artificial intelligence (XAI) techniques to quantify feature-level contributions for individual predictions.

Future work will address these limitations through several extensions. Longitudinal modeling using repeated BP measurements and time-series analysis will be explored to capture intra-individual variability and improve early spike detection. Integration of wearable sensor data and ambulatory BP monitoring may further enhance temporal resolution and predictive robustness [6]. Moreover, advanced explainability techniques such as SHAP and LIME will be incorporated to provide transparent, instance-level explanations, facilitating clinical trust and regulatory acceptance [7], [8].

Finally, prospective validation studies and real-world pilot deployments will be pursued to assess clinical impact, usability, and integration within existing healthcare workflows. These directions will enable the transition of Predictive Pulse from a predictive analytics framework to a practical decision-support system for preventive cardiovascular care.

7. CONCLUSION

This paper presented Predictive Pulse, a comprehensive and interpretable machine learning framework for early prediction of blood pressure (BP) spike risk using routinely collected clinical, demographic, and lifestyle data. By reframing hypertension analytics from static disease classification to anticipatory risk modeling, the proposed approach addresses a critical gap in preventive cardiovascular care. The framework integrates clinically informed data preprocessing, structured feature engineering, exploratory analysis, and comparative evaluation of multiple supervised learning models within a reproducible pipeline.

Experimental results demonstrate that ensemble-based models outperform linear baselines, effectively capturing non-linear interactions among physiological and behavioral factors while maintaining robustness and interpretability. The alignment between model-derived feature importance and established clinical knowledge further reinforces the practical relevance of the proposed system. Importantly, the emphasis on recall and balanced performance metrics highlights the framework’s suitability for screening and early-intervention applications, where minimizing missed high-risk cases is essential.

Beyond predictive performance, this work underscores the feasibility of deploying machine learning–driven decision-support tools using accessible data sources, without reliance on invasive monitoring or complex sensing infrastructure. The proposed framework provides a scalable foundation for population-level risk stratification and supports the broader transition toward preventive and precision medicine paradigms.

Future extensions will focus on longitudinal modeling, integration of wearable and ambulatory BP data, and incorporation of explainable artificial intelligence techniques to enhance transparency and clinical trust. Through these advancements, Predictive Pulse has the potential to evolve into a real-world clinical decision-support system that enables proactive hypertension management and contributes meaningfully to data-driven cardiovascular healthcare.

ACKNOWLEDGMENT

The author acknowledges the support of the open-source research and machine learning communities for providing tools and resources essential to this work. The author also appreciates the guidance of the broader research community and peer feedback that contributed to improving the quality of this research.

REFERENCES

[1] World Health Organization, Cardiovascular Diseases (CVDs) Fact Sheet, Geneva, Switzerland, 2023.

[2] P. K. Whelton, R. M. Carey, W. S. Aronow, et al., "2017 ACC/AHA guideline for the prevention, detection, evaluation, and management of high blood pressure in adults," Hypertension, vol. 71, no. 6, pp. e13–e115, Jun. 2018.

[3] R. S. Vasan, M. G. Larson, E. P. Leip, et al., "Impact of high-normal blood pressure on the risk of cardiovascular disease," New England Journal of Medicine, vol. 345, no. 18, pp. 1291–1297, 2001.

[4] T. Kario, "Blood pressure variability in cardiovascular risk," Hypertension Research, vol. 33, no. 6, pp. 512–519, Jun. 2010.

[5] R. Rothwell, S. C. Howard, E. Dolan, et al., "Prognostic significance of visit-to-visit variability in systolic blood pressure," The Lancet, vol. 375, no. 9718, pp. 895–905, Mar. 2010.

[6] E. J. Benjamin, P. Muntner, A. Alonso, et al., "Heart disease and stroke statistics—2019 update," Circulation, vol. 139, no. 10, pp. e56–e528, Mar. 2019.

[7] S. F. Weng, J. Reps, J. Kai, J. M. Garibaldi, and N. Qureshi, "Can machine-learning improve cardiovascular risk prediction using routine clinical data?," PLOS ONE, vol. 12, no. 4, 2017.

[8] C. Krittanawong, H. J. Zhang, Z. Wang, M. Aydar, and T. Kitai, "Artificial intelligence in precision cardiovascular medicine," Journal of the American College of Cardiology, vol. 69, no. 21, pp. 2657–2664, 2017.

[9] Z. Obermeyer and E. J. Emanuel, "Predicting the future of machine learning in health care," New England Journal of Medicine, vol. 375, no. 13, pp. 1216–1219, 2016.

[10] J. Booth, C. Roberts, P. Laye, and A. MacGregor, "Physical activity and hypertension," Hypertension, vol. 64, no. 4, pp. 736–742, 2014.

[11] J. A. Appel, L. J. Moore, E. Obarzanek, et al., "A clinical trial of the effects of dietary patterns on blood pressure," New England Journal of Medicine, vol. 336, no. 16, pp. 1117–1124, 1997.

[12] R. Parati, J. E. Ochoa, C. Lombardi, and G. Bilo, "Assessment and management of blood-pressure variability," Nature Reviews Cardiology, vol. 10, no. 3, pp. 143–155, 2013.

[13] F. Doshi-Velez and B. Kim, "Towards a rigorous science of interpretable machine learning," arXiv preprint arXiv:1702.08608, 2017.

[14] S. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Proc. Advances in Neural Information Processing Systems (NeurIPS), 2017, pp. 4765–4774.

[15] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should I trust you? Explaining the predictions of any classifier," in Proc. ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, 2016, pp. 1135–1144.

[16] E. J. Topol, "High-performance medicine: The convergence of human and artificial intelligence," Nature Medicine, vol. 25, no. 1, pp. 44–56, 2019.