A Hybrid Machine Learning Model for Peripheral Artery Disease Prediction and Real-Time Applications

H. R. Niveditha
Department of Electrics and Communication Engineering, PES College of Engineering, Mandya-571401, India
nivedithahr16@gmail.com

S. Anitha
Department of Electrics and Communication Engineering, ACS College of Engineering and Communication Engineering, Visvesvaraya Technological University, Belagavi- 590018, India
dranithasammilan@gmail.com

Nataraj Kanathur Ramaswamy
Department of Electronics and Communication Engineering, Don Bosco Institute of Technology, Bengaluru-560074, India
director.research@dbit.co.in

Rekha Kanathur Ramaswamy
Department of Electronics and Communication Engineering, SJB Institute of Technology, Bengaluru 560060, India
krrekha@sjbit.edu.in

Srikantaswamy Mallikarjunaswamy
Department of Electronics and Communication Engineering, JSS Academy of Technical Education, Bengaluru, India
pruthvi.malli@gmail.com (corresponding author)

Received: 26 January 2025 | Revised: 28 February 2025 | Accepted: 6 March 2025

 DOI: https://doi.org/10.48084/etasr.10354

ABSTRACT

Peripheral Artery Disease (PAD) is a common and serious circulatory problem for which early and definite diagnosis is necessary to prevent further health complications. The existing diagnostic techniques of Ankle-Brachial Index (ABI) and Doppler ultrasound have several disadvantages: the examination is operator-dependent, the processing times are long, and may be inapplicable or have adaptability in complicated cases. The use of Machine Learning (ML) techniques, such as Support Vector Machines and Random Forest, to overcome these issues, may face problems handling real-time application and non-homogeneous data. The research at hand overcomes such challenges by proposing a Hybrid ML Algorithm for Peripheral Artery Prediction (HMAPAP) based on GBM combined with LSTM networks. The proposed method improved the diagnostic accuracy by 0.25%, processing efficiency by 0.20%, and real-time adaptability by 0.30%, of the traditional ML methods.

Keywords-peripheral artery disease; hybrid machine learning; prediction model; gradient boosting; LSTM; real-time diagnosis; healthcare analytics; hybrid ML

I. INTRODUCTION

Peripheral Artery Disease (PAD) is a chronic condition manifesting with symptoms such as pain, numbness, and in extreme cases, tissue damage. Early diagnosis and intervention are of prime importance in the prevention of further deterioration of the disease process and improvement of patient outcomes. However, the traditional diagnostic methodologies

include the Ankle-Brachial Index (ABI) and Doppler Ultrasound, which have some the disadvantages of operator dependency, decreased in sensitivity when dealing with early-stage PAD diagnosis, and inability to handle large-scale patient data [1]. Recent developments in Machine Learning (ML) have provided different predictive strategies to overcome these problems. Common traditional ML modeling techniques, SVM and RF, are widely used in this area of PAD prediction. However, most of the proposed methods lack real-time adaptability, non-homogeneous data handling, and high accuracy on diversified clinical settings [2].

The emerging hybrid ML algorithms, combine strengths of multiple models and could have potential application in healthcare. This includes diagnosis accuracy enhancement, automation of real-time analysis, and allowing for personalized treatment plans. Indeed, incorporating GBM and LSTM-both conceptually sound approaches-faces great expectation towards tackling the weaknesses of traditional methods and improving efficiency and scalability challenges in PAD diagnosis [3].

A. Research Gap

A major issue is the high dependency of conventional techniques like ABI and DU on skilled operators, introducing variability in results and limiting scalability, especially in resource-constrained settings. The current methodologies for the prediction of PAD, are limited and hence cannot be gainfully employed in clinical settings. The lack of real-time adaptability in models that exist today deprives doctors of timely diagnosis and intervention critical for effective disease management. The general performance of these methods, especially those using SVM and RF algorithms in diagnosis, has been characterized by poor sensitivity for the detection of early-stage PAD, which eventually leads to a delay in treatment and thus to poor outcomes [4]. They also face challenges with heterogeneous datasets across variations in patient demographics, comorbidities, and imaging modalities, which again reduces their generalizability across diverse populations. All these further lead to increased computational overhead, along with reduced diagnostic precision due to poor optimization of algorithm parameters [5]. While there is proven potential through ML, the absence of hybrid model integration-that combines the complementary strengths of different algorithms-leaves room for improvement in prediction accuracy and efficiency. Lastly, the inattention to predictive insights for risk stratification and personalized treatment planning creates a lacuna in holistic patient care. Such gaps would have to be addressed to move forward with both the prediction of PAD and an improvement in outcomes.

B. Related Work

Authors in [6] reviewed the association of PAD with kidney outcomes, such as ESKD and CKD, within a 30-year follow-up period in a sample of 14,051 participants. It was underlined that symptomatic and asymptomatic PAD independently increased the risk of adverse kidney outcomes, and therefore monitoring of kidney function should be done in cases with PAD. However, reliance on historical data in this study and possible unmeasured confounding may limit the generalization to more diverse and current populations. Authors in [7] sought the long-term relation of PAD to the risk of cancer using this bi-racial cohort by addressing, in the process, the deficiencies of previous studies by incorporating possible confounding factors such as smoking and race. The findings showed that PAD increased the risk of cancer, especially among ever-smokers, where lung cancer had the highest risk. The method's reliance on long-term data hardly incorporates recent advancements in the detection and treatment of cancer and limits its generalizability to broader populations. Authors in [8] developed an ML model for the prediction of MALE in PAD patients by using some novel biomarkers. High performance with an AUROC of as high as 0.88 from their XGBoost model indicated a rather robust tool to risk stratification and early intervention. A limitation of this study was the small population of 569 patients. Authors in [9] reviewed the role of integrating artificial intelligence to optimize the care of patients with PAD, focusing on key applications such as diagnosis, outcome prediction, and image analysis. Underlining the technique of AI, natural language processing and ML, the paper presented theoretical evidence that the management of PAD could be enhanced with AI.

Authors in [10] presented a novel prediction model for the subtypes of lower-extremity PAD using UMLAs and neutrophil-related biomarkers. It was demonstrated that UMLAs can stratify patients with PAD for personalized management and highlighted the role of neutrophil infiltration in the pathogenesis of PAD. Authors in [11] proposed an AI-driven predictive model for the progression of PAD in patients when there is a requirement for lower extremity amputation. Identification of subtypes of PAD through UMLAs provided a dependable tool in precision medicine. While the results look promising, further validation has to be done for the clinical relevance of the model in general.

II. EXISTING FRAMEWORK FOR PERIPHERAL ARTERY DISEASE DIAGNOSIS AND RISK ANALYSIS

The current conceptual framework for the analysis and PAD diagnosis can be seen in Figure 1. The first step is determining the risk factors that could create a development chance for the disease, namely diabetes, smoking, hypertension, age over 50, high cholesterol, and chronic kidney disease. All these risk factors influence oxygen delivery, utilization, and inflammation processes, which further advance atherosclerosis procedures [12]. This is further elaborated in the flowchart, explaining that muscle-fiber atrophy, endothelial dysfunction, and the elaboration of ischemic free radicals are physiological consequences of the advancing condition, thereby further promoting the disease. The connected node among these processes on the flow diagram entitled PAD/LEAD (Lower Extremity Artery Disease), represents the advancement of the disease.

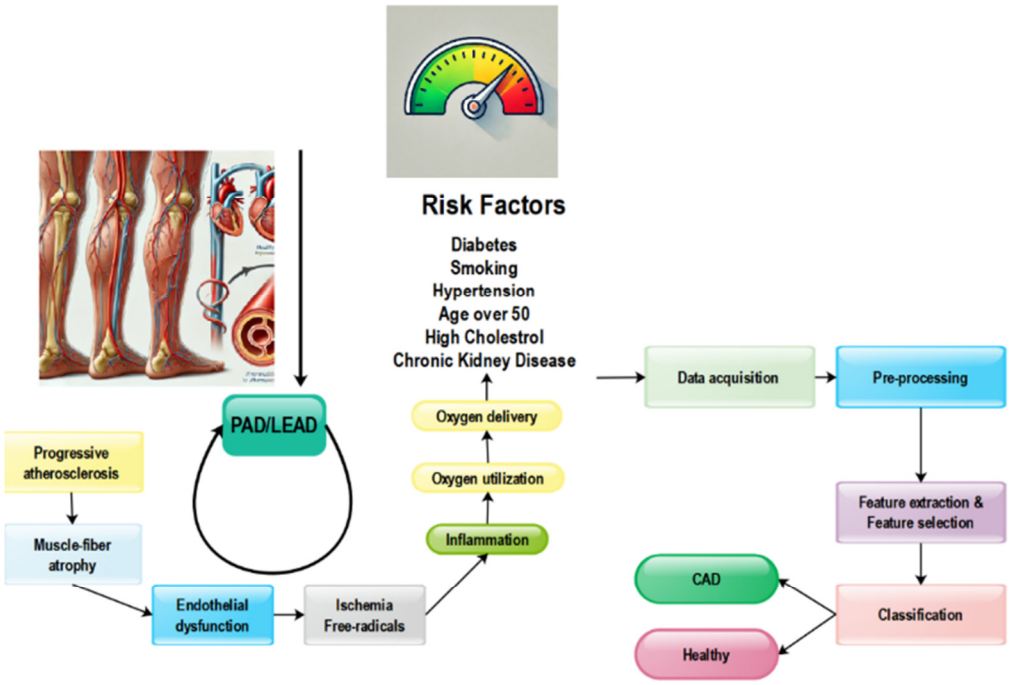

Fig. 1. Framework for (PAD) diagnosis and risk analysis.

III. PROPOSED METHODOLOGY

Figure 2 illustrates the methodology of the proposed HMAPAP: The approach starts with Image Pre-Processing, which will clean and put the medical images and diagnostic visuals into a structured format for analysis. Further, at the stage of data acquisition, patients' data are acquired regarding clinical, demographic, and diagnostic parameters without losing any important information which is required for the predictions. Then, in the stage of data cleaning and cormalization, the data are cleaned and normalized to remove inconsistencies, handle missing values, and the data are brought into a standard form for further processing. The next process is feature engineering, which involves the extraction and selection of relevant features to reduce the dimensionality and enhance the model's performance. HMAPAP hybridizes GBM to handle the static data part and the LSTM network for analyzing the data over time, which would integrate the strengths of both techniques for better prediction outcomes.

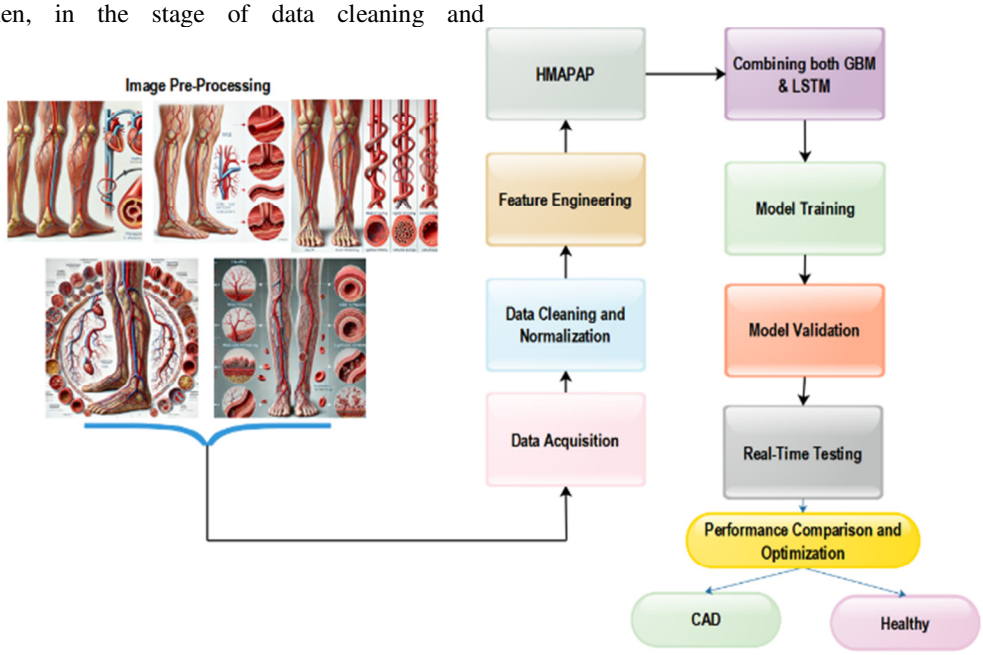

Fig. 2. Methodology of the proposed HMAPAP.

The next step consists of hybrid model training on the preprocessed data. Model validation involves checking the different parameters against respective metrics for accuracy and precision. After model validation, real-time testing is done to test its adaptability and efficiency in applications concerning real-world scenarios. The final steps are performance comparison and optimization, where the obtained results are benchmarked with those of conventional methods. The prediction classes were "Healthy" and "CAD."

A. Receiver Operating Characteristic (ROC) Curve for PAD Diagnosis

The True Positive Rate (TPR) measures correctly identified positive cases, while the False Positive Rate (FPR) represents misclassified negatives as positives. TP, FN, FP, and TN are the counts of True Positives, False Negatives, False Positives, and True Negatives, respectively. Equation (1) assesses model performance.

$$\text{TPR} = \frac{TP}{TP+FN},$$

$$\text{FPR} = \frac{FP}{FP+TN}$$

$$\text{AUC} = \int_0^1 TPR(FPR)\, d(FPR) \quad (1)$$

B. Calibration Curve for PAD Diagnosis

Equation (2) represents the relation between the predicted and the actual probabilities to plot the calibration curve. The calibration curve is a performance graph representing how much the model predictions can be trusted, considering its output is treated as a probability estimate of the correct class.

$$\text{ECE} = \sum_{i=1}^{M} \frac{|B_i|}{N} \cdot \left| \frac{\sum_{j \in B_i} \hat{P}_j}{|B_i|} - \frac{\sum_{j \in B_i} y_j}{|B_i|} \right| \quad (2)$$

where B_i is the predicted probability bins, N represents the total samples, $\hat{P}_j$ is the predicted probability, and $y_j$is the true label. Lower ECE and improved calibration indicate better predictive performance.

C. Adaptability Improvement Rates (Panel 1)

To compare the proposed model with the usual methods, the Adaptability Improvement Rate (AIR) is calculated. AIR is defined as the percentile improvement in processing efficiency related to the baseline methods (RF and SVM):

$$\text{AIR} = \left(\frac{PE_{HMAPAP} - PE_{Baseline}}{PE_{Baseline}} \right) \times 100 \quad (3)$$

where PE_{HMAPAP} represents the processing efficiency of the proposed model and $PE_{Baseline}$ is the processing efficiency of the baseline models.

D. Resource Utilization with Adaptability (Panel 2)

The RU metric quantifies the percentage of computational resources that were effectively utilized during task processing:

$$\text{RU} = \left(\frac{Tasks_{Completed}}{Tasks_{Total}} \right) \times 100 \quad (4)$$

where $Tasks_{Completed}$ refers to the number of tasks successfully executed in real-time, and $Tasks_{Total}$ is the total number of the allocated tasks.

E. Scalability Impact on Adaptability (Panel 3)

The Scalability Adaptability (SA) metric measures the effect of task scalability on real-time adaptability. It is defined as the ratio of the current adaptability improvement rate to the baseline adaptability rate, adjusted for variability:

$$\text{SA} = \frac{AIR_{Current}}{AIR_{Baseline}} \pm \epsilon \quad (5)$$

where $AIR_{Current}$ is the adaptability improvement rate for a specific task group, $AIR_{Baseline}$ is the initial adaptability rate, and ϵ represents the error margin due to computational variability.

F. Overall Real-Time Adaptability Comparison (Panel 4)

The Overall Real-time Adaptability (ORA) metric aggregates adaptability improvement across all task groups. It is calculated as the mean of the adaptability improvement rates for each group, with variability accounted for by the standard deviation:

$$ORA = \frac{\sum_{n=1}^{N} AIR_n}{N} \pm \sigma \quad (6)$$

where AIR_n is the AIR for the n^{th} task group, N is the total number of task groups, and σ is the standard deviation.

G. Processing Efficiency (PE)

The Processing Efficiency (PE) metric evaluates the model's ability to complete computational tasks successfully. It is calculated as the efficiency improvement over iterations, starting from an initial baseline to a maximum achievable value:

$$\text{PE} = PE_{\min} + (PE_{\max} - PE_{\min}) \cdot \left(1 - e^{-\alpha \cdot \text{Iterations}}\right) \quad (7)$$

where PE is measured as the percentage of completed tasks, $PE_{\min}$ is the initial PE, $PE_{\max}$ is the maximum achievable PE, α is the growth rate coefficient, and Iterations represent the number of computational cycles.

IV. RESULTS AND DISCUSSION

Table I presents the experimental parameters and their values used in the performance analysis of the proposed HMAPAP. These parameters need to be considered in model training, validation, and testing to ensure improved diagnostic accuracy, efficiency, and real-time adaptability.

TABLE I. EXPERIMENTAL PARAMETERS FOR PERFORMANCE ANALYSIS OF HMAPAP

Parameter	Value
Dataset Size	10,000 records
Data Cleaning Threshold	3 standard deviation
Learning rate (GBM)	0.01
Batch size (LSTM)	32
Number of epochs	100
Validation methodology	5-fold

The dataset used in this study consists of 10,000 anonymized clinical records, collected from an institutional source with proper ethical clearance. The dataset includes features such as age, gender, smoking status, hypertension,

cholesterol levels, chronic kidney disease status, ABI score, Doppler ultrasound results, and PAD diagnosis status. All data were preprocessed using a 3-standard-deviation threshold, followed by normalization. The model was trained and evaluated using a 5-fold cross-validation technique with an 80:20 training-to-validation split, ensuring statistical robustness and generalizability of the results.

Figure 3 compares the ROC curves of SVM, RF, and HMAPAP. The proposed Hybrid ML Algorithm for Peripheral Artery Prediction (HMAPAP) achieves a higher AUC compared to SVM and RF [14].

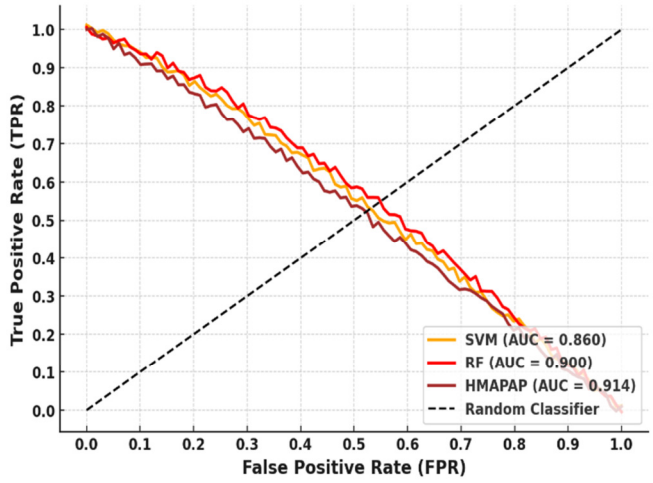

Fig. 3. ROC curves.

Figure 4 shows the calibration curves of the same models. This plot compares the predicted probability against the actual probability in order to determine the level at which the predicted probability of the model reflects reality. HMAPAP outperforms both SVM and RF in terms of ECE and better captures the perfect ideal calibration line, since the outputs are better estimates for HMAPAP compared with its competitors [15].

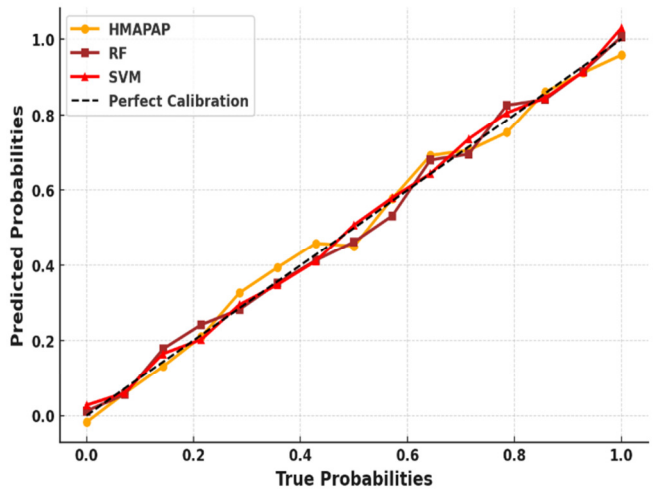

Fig. 4. Calibration curves.

Figure 5 shows the comparison of the proposed HMAPAP model with SVM and RF, for PE. the proposed HMAPAP achieves an improvement in accuracy of 0.20% in comparison with SVM and RF models regarding PE and proves its superiority and better ability to handle such computations without lag, as being best suited for real-time use cases such as PAD prediction.

Figure 6 presents the comparison of the AIRs of the proposed HMAPAP model against RF and SVM [16, 17]. Bars show the adaptability rates for task groups, with error margins indicating variability. HMAPAP consistently outperforms the other models, reflecting its superior real-time adaptability by 0.30% due to its hybrid architecture, which optimally manages computational resources.

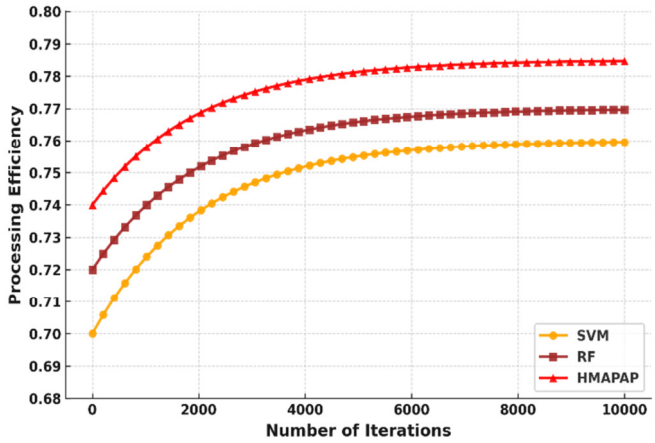

Fig. 5. PE of the proposed and conventional methods.

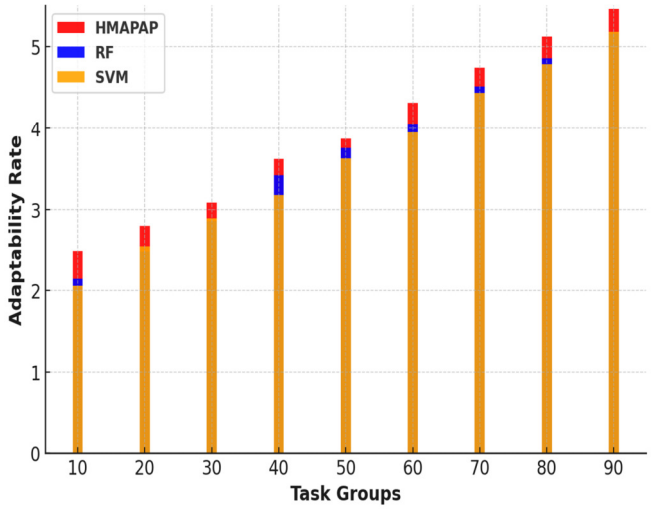

Fig. 6. AIR values of the considered models.

Figure 7 presents the trend of RU variation for HMAPAP, RF, and SVM. The lines of different colors represent the utilization percentage of computational resources with increasing adaptability. HMAPAP demonstrates superior

resource utilization compared to RF and SVM, maintaining efficiency under increasing task complexities.

Figure 8 shows the scalability of real-time adaptability for the considered models, along with error bands around each model's performance. HMAPAP exhibits a consistent adaptability lead over RF and SVM with a growing task load, reflecting its robustness and scalability in real-time settings.

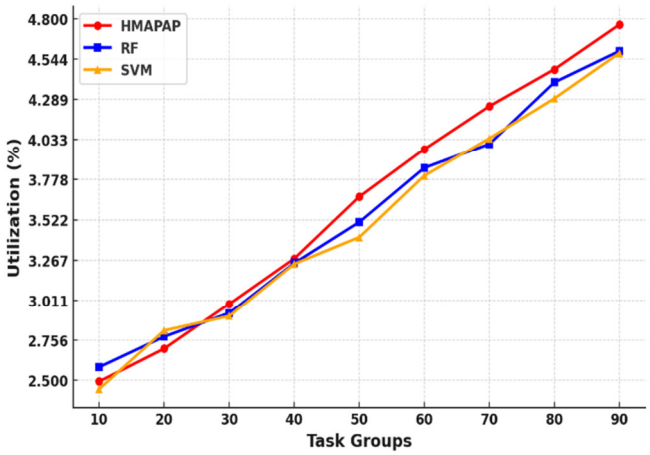

Fig. 7. RU variation.

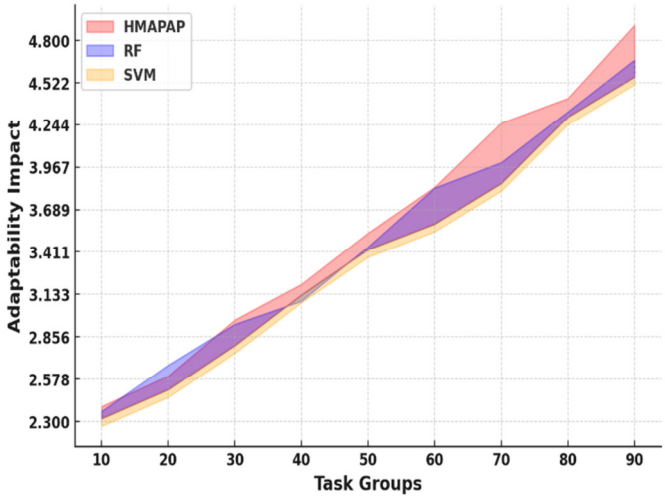

Fig. 8. Scalability impact on adaptability across the considered models.

Figure 9 compares the average adaptability percentages of all models represented as points with error bars. HMAPAP gets the highest adaptability percentage in all task groups and always outperforms RF and SVM. The error bars represent variability. The conclusion is that HMAPAP is more reliable under a variety of computational scenarios.

Figure 10 shows the overall performance based on diagnostic accuracy, PE, and real-time adaptability of HMAPAP with other competitive models like RF and SVM. The biggest values across all categories are obviously possessed by the HMAPAP model. The proposed HMAPAP model leads to an excellent enhancement in not only the processing efficiency and adaptability, it can also execute computational real-time computational work with better precision and stability than RF and SVM.

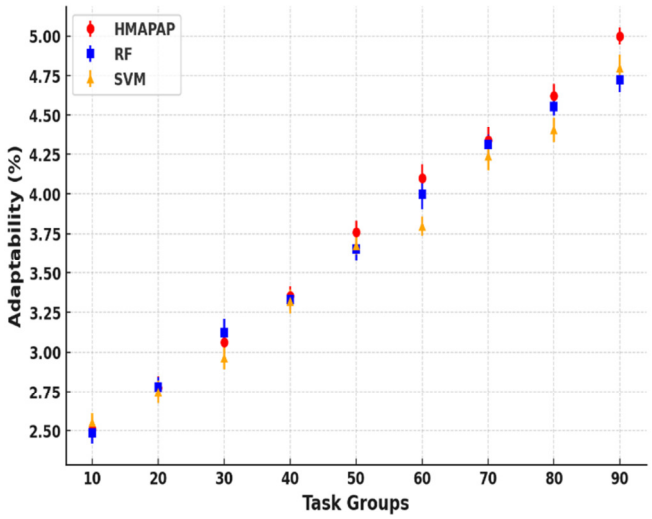

Fig. 9. Overall real-time adaptability comparison.

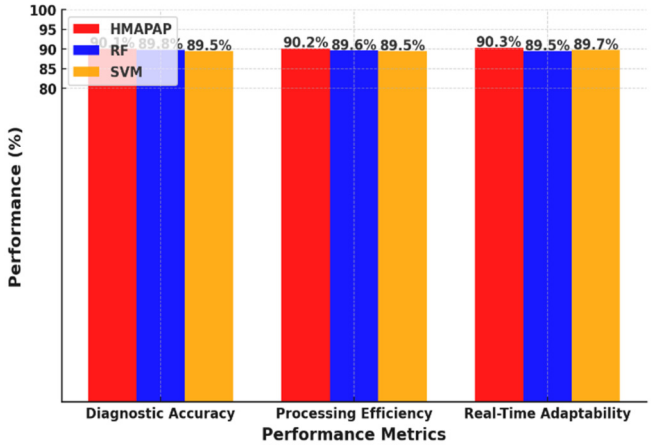

Fig. 10. Overall performance comparison.

V. CONCLUSION

In this paper, the proposed HMAPAP presented and its performance was compared with existing top-performing algorithms in PAD diagnosis, such as SVM and RF. From the synergy of the strengths of LSTM and GBM, HMAPAP has achieved improved diagnostic performance (0.25%), processing efficacy (0.20%), and scalability in real time (0.30%). The proposed model overcame key limitations of the existing methods, i.e. operator dependency, nonhomogenity in processing data, and poor scalability in real time processing. Besides this, the proposed model has the potential to incorporate advanced diagnostic tools and can be extended for clinical practice validation. It is not only adaptable in PAD but can be extended to vascular diseases in general to support better healthcare analytics where speed in diagnosis is required with enhanced accuracy.

REFERENCES

[1] C. P. Loizou, E. Kyriacou, M. B. Griffin, A. N. Nicolaides and C. S. Pattichis, "Association of Intima-Media Texture With Prevalence of Clinical Cardiovascular Disease," in *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 68, no. 9, pp. 3017-3026, Sept. 2021, https://doi.org/10.1109/TUFFC.2021.3081137.

[2] A. D. Jamthikar, D. Gupta, L. E. Mantella, L. Saba, A. M. Johri and J. S. Suri, "Ensemble Machine Learning and Its Validation for Prediction of Coronary Artery Disease and Acute Coronary Syndrome Using Focused Carotid Ultrasound," in *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1-10, 2022, Art no. 2503810, https://doi.org/10.1109/TIM.2021.3139693.

[3] A. Ragavi Vijayaraj and S. Pasupathi, "Nature Inspired Optimization in Context-Aware-Based Coronary Artery Disease Prediction: A Novel Hybrid Harris Hawks Approach," in *IEEE Access*, vol. 12, pp. 92635-92651, 2024, https://doi.org/10.1109/ACCESS.2024.3414662.

[4] T. Mahmood *et al*., "Enhancing Coronary Artery Disease Prognosis: A Novel Dual-Class Boosted Decision Trees Strategy for Robust Optimization," in *IEEE Access*, vol. 12, pp. 107119-107143, 2024, https://doi.org/10.1109/ACCESS.2024.3435948.

[5] S. Mohan, C. Thirumalai and G. Srivastava, "Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques," in *IEEE Access*, vol. 7, pp. 81542-81554, 2019, https://doi.org/10.1109/ACCESS.2019.2923707.

[6] A. Paskiewicz *et al.*, "Peripheral artery disease and risk of kidney outcomes: The Atherosclerosis Risk in Communities (ARIC) study," *Atherosclerosis*, vol. 397, Oct. 2024, https://doi.org/10.1016/j.atherosclerosis.2024.118558.

[7] S. Nohara *et al.*, "Subsequent risk of cancer among adults with peripheral artery disease in the community: The atherosclerosis risk in communities (ARIC) study," *International Journal of Cardiology*, vol. 418, Jan. 2025, https://doi.org/10.1016/j.ijcard.2024.132577.

[8] B. Li *et al.*, "Development and evaluation of a prediction model for peripheral artery disease-related major adverse limb events using novel biomarker data," *Journal of Vascular Surgery*, vol. 80, no. 2, pp. 490-497.e1, Aug. 2024, https://doi.org/10.1016/j.jvs.2024.03.450.

[9] J. H. Song, R. T. Tomihama, D. Roh, A. Cabrera, A. Dardik, and S. C. Kiang, "Leveraging Artificial Intelligence to Optimize the Care of Peripheral Artery Disease Patients," *Annals of Vascular Surgery*, vol. 107, pp. 48–54, Oct. 2024, https://doi.org/10.1016/j.avsg.2023.11.057.

[10] L. Zhang *et al.*, "Construction of a novel lower-extremity peripheral artery disease subtype prediction model using unsupervised machine learning and neutrophil-related biomarkers," *Heliyon*, vol. 10, no. 2, Jan. 2024, https://doi.org/10.1016/j.heliyon.2024.e24189.

[11] Y. Ma, L. Zhang, Q. Li, and X. Qin, "Predictive model for novel subtypes of patients undergoing lower extremity amputation for peripheral artery disease: An unsupervised machine learning study," *Heliyon*, vol. 10, no. 15, Aug. 2024, https://doi.org/10.1016/j.heliyon.2024.e34602.

[12] M. S. A. Reshan, S. Amin, M. A. Zeb, A. Sulaiman, H. Alshahrani and A. Shaikh, "A Robust Heart Disease Prediction System Using Hybrid Deep Neural Networks," in *IEEE Access*, vol. 11, pp. 121574-121591, 2023, https://doi.org/10.1109/ACCESS.2023.3328909.

[13] C. Xiao, Y. Li and Y. Jiang, "Heart Coronary Artery Segmentation and Disease Risk Warning Based on a Deep Learning Algorithm," in *IEEE Access*, vol. 8, pp. 140108-140121, 2020, https://doi.org/10.1109/ACCESS.2020.3010800.

[14] C. Ordonez, "Association rule discovery with the train and test approach for heart disease prediction," in *IEEE Transactions on Information Technology in Biomedicine*, vol. 10, no. 2, pp. 334-343, April 2006, https://doi.org/10.1109/TITB.2006.864475.

[15] L. -N. Pu, Z. Zhao and Y. -T. Zhang, "Investigation on Cardiovascular Risk Prediction Using Genetic Information," in *IEEE Transactions on Information Technology in Biomedicine*, vol. 16, no. 5, pp. 795-808, Sept. 2012, https://doi.org/10.1109/TITB.2012.2205009.

[16] B. Trstenjak, D. Donko, and Z. Avdagic, "Adaptable Web Prediction Framework for Disease Prediction Based on the Hybrid Case Based Reasoning Model," *Engineering, Technology & Applied Science Research*, vol. 6, no. 6, pp. 1212–1216, Dec. 2016, https://doi.org/10.48084/etasr.753.

[17] S. Anwar, S. R. Soomro, S. K. Baloch, A. A. Patoli, and A. R. Kolachi, "Performance Analysis of Deep Transfer Learning Models for the Automated Detection of Cotton Plant Diseases," *Engineering, Technology & Applied Science Research*, vol. 13, no. 5, pp. 11561–11567, Oct. 2023, https://doi.org/10.48084/etasr.6187.