

OPEN ACCESS

This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

¹Department of Information Technology, Thiagarajar College of Engineering, Madurai, Tamil Nadu, India

²Department of Electronics and Communication Engineering, Thiagarajar College of Engineering, Madurai, Tamil Nadu, India

³Department of Computer Science and Engineering, Thiagarajar College of Engineering, Madurai, Tamil Nadu, India

Correspondence to: Ezhil Gopal Ramasamy, ezhilr@student.tce.edu

Additional material is published online only. To view please visit the journal online.

Cite this as: Ramasamy EG, Subbiah S, Sivasubramanian R and Malaichamy ND. A Non-Invasive Deep Learning Approach for Early Detection of Peripheral Arterial Disease–Diabetic Foot Ulcers Using Retinal Imaging: A Prospective Cohort Study. Premier Journal of Science 2025;15:100258

DOI: <https://doi.org/10.70389/PJS.100258>

Peer Review

Received: 14 August 2025

Last revised: 8 January 2026

Accepted: 11 January 2026

Version accepted: 8

Published: 26 February 2026

A Non-Invasive Deep Learning Approach for Early Detection of Peripheral Arterial Disease–Diabetic Foot Ulcers Using Retinal Imaging: A Prospective Cohort Study

Ezhil Gopal Ramasamy¹, Sridevi Subbiah¹, Rajaram Sivasubramanian² and Nirmala Devi Malaichamy³

ABSTRACT

Peripheral Arterial Disease (PAD) is one of many serious vascular complications of diabetes mellitus, and it is a major contributor to diabetic foot ulcers (DFUs) and amputations. The importance of early detection is paramount; however, the existing diagnostic tools, the Doppler ultrasound and Ankle Brachial Index (ABI) testing, are expensive and not available in many countries. This article introduces an artificial intelligence (AI)-based project of detecting PAD non-invasively and predicting DFU risk through retinal fundus images. An adapted EfficientNet-B3 convolutional network with deep retina features combined with vessel density, fractal dimension, hemodynamic index, and vascular descriptors was created and tested on a clinically validated cohort of 1,928 diabetic patients with gold-standard PAD confirmation via ABI and Doppler ultrasound. With a traditional convolutional neural network and ResNet-50 model, the model has an AUC-Receiver Operating Characteristic of 96.1%, sensitivity of 94.2, and specificity of 91.5, and was superior to the models ($P < 0.001$). The biologically relevant vascular attention was validated by Grad-CAM visualization. This model emphasizes the opportunities of retinal imaging to be a cost-efficient, AI opportunistic PAD and DFU screening tool in the routine diabetic eye examination.

Keywords: Retinal fundus imaging, Peripheral arterial disease diagnosis, Diabetic foot ulcer risk prediction, Efficientnet-B3 convolutional network, Microvascular fractal analysis

Introduction

Diabetes mellitus is a chronic metabolic disease that has already turned into a global problem of health care and reached more than 537 million adults worldwide by 2021, and is expected to reach 643 million in 2030 and 783 million in 2045 (International Diabetes Federation, 2021).¹ The condition causes various extreme complications in the body, and such complications impact both the heart and other organs, the nerves, and the vascular system. Peripheral Arterial Disease (PAD) is one of the most debilitating and underdiagnosed complications of diabetes that can greatly lead to the morbidity, disability, and mortality of diabetic patients.²

Atherosclerosis causes PAD; it involves the constriction and hardening of the arteries as a result of the accumulation of plaque, thereby limiting the blood flow to the lower extremities. Research has shown that diabetes doubles to four times the risk of developing

PAD than non-diabetic people.³ This condition is commonly asymptomatic in early years, which causes the late diagnosis and the serious complications, including chronic limb ischemia, foot ulcer, gangrene, and amputation.⁴

Poor circulation, nerve damage (neuropathy), and wound healing account to one of the most severe complications of diabetes-associated PAD diabetic foot ulcers (DFUs). DFUs impact lifetime infection, long hospitalization, and amputation of lower limbs in 34% of diabetic patients.⁵ The World Health Organization (WHO) estimates that more than 1 million amputations are caused by diabetes each year, and 85% of these amputations are first accompanied by foot ulcers.⁶ These complications have a strong influence on the quality of life of the patients, they raise the mortality rates, and they add to the expanding financial costs of health care systems across the world.

Although PAD and DFUs are highly prevalent and associated with serious consequences, early diagnosis is still a problem because of the limitations of the traditional diagnostic tools. Ankle-Brachial Index (ABI), Doppler ultrasound, and angiography are currently popular procedures in the diagnosis of PAD, yet they are invasive, costly, time-intensive, and demanding of a medical professional.⁷ Furthermore, PADs remain undiagnosed in most instances, especially in low-resource environments where diagnostic instruments are scarce. On the same note, clinical assessment of DFUs is usually done late and in most cases, such ulcers have advanced to severe stages, which results in poor treatment results.⁸

Current developments in the field of artificial intelligence (AI) and deep learning (DL) have provided new prospects for disease detection without invasions. A possible potential application of retinal imaging, which is frequently employed to detect diabetic retinopathy (DR), is the ability to measure systemic vascular health, including diabetic patients PAD-related complications.⁹ The microvascular architecture of the retina has pathological similarities with peripheral arteries, i.e., any vascular defects in the retina may indicate early atherosclerosis and PAD.¹⁰

This article provides full clinical validation of a machine learning methodology that applies retinal images and Convolutional Neural Networks (CNNs) to identify PAD and predict the risk of DFU in diabetic patients. gathered real clinical data in the (Hospital Name) and included all patients who have undergone full vascular work up with clinical examination, ABI and Doppler

Ethical approval: N/a

Consent: N/a

Funding: The authors express their gratitude to the Thiagarajar College of Engineering (TCE) for supporting us to carry out this research work. Also, the financial support from TCE under Thiagarajar Research Fellowship scheme (File.no: TRF/Jul-2024/02) is gratefully acknowledged

Conflicts of interest: N/a

Author contribution: Ezhil Gopal Ramasamy, Sridevi Subbiah, Rajaram Sivasubramanian and Nirmala Devi Malaichamy – Conceptualization, Writing – original draft, review and editing

Guarantor: Ezhil Gopal Ramasamy

Provenance and peer-review: Unsolicited and externally peer-reviewed

Data availability statement: To promote transparency and reproducibility, the full experimental framework, including model architectures, training logs, and fixed random seed configurations, has been archived. A lightweight Docker container and model card describing dataset structure, preprocessing steps, and evaluation parameters are available upon reasonable request. An anonymized sample test set (10%) is also provided for reproducibility verification in compliance with institutional data-sharing policies. All analyses were executed using Python 3.10 with TensorFlow 2.12 and PyTorch 2.0 under fixed computational environments.

The datasets generated and/or analysed during the current study are not publicly available due to patient privacy restrictions but are available from the corresponding author on reasonable request

ultrasound confirmation, retinal fundus examination and a follow-up of DFU development over 18 months. CNNs have proven to have high levels of accuracy in the classification of medical images and thus are the best to be used in the analysis of the faint changes in the microvascular changes in retinal images that might be associated with the development of PAD. The proposed pre-clinical research would address:

- Test AI-based retinal scan versus gold-standard PAD diagnostics (ABI, Doppler ultrasound) in an actual clinical group.
- Prove the viability of a cost-effective and non-invasive screening method based on retinal imaging that is readily available.
- Build another base on future research on clinically validated PAD/DFU labels with gold standard diagnostic tools.
- Help in designing accessible screening vehicles for diabetes management programs.

The proposed study represents the first step towards achieving the future clinical validation research that can help close the gap in the early detection of PAD and DFU through the integration of retinal imaging, AI-based analysis, and DL-based modeling studies.

Literature Review

The early detection of PAD and DFUs in diabetic patients has been a significant area of research. Various studies have explored vascular abnormalities, non-invasive diagnostic techniques, and machine learning-based approaches to improve early screening and prediction. This section provides a review of existing literature, emphasizing traditional methods, retinal imaging as a diagnostic tool, and recent advancements in DL for PAD and DFU detection.

PAD is a common but often underdiagnosed complication in diabetic individuals. Traditional diagnostic approaches rely on the ABI, Doppler ultrasound, and angiography, which, although effective, have significant limitations in terms of cost, accessibility, and invasiveness. Studies indicate that diabetes increases PAD risk by two to four times, with many cases remaining asymptomatic until severe complications arise.¹¹ According to Norgren et al., the Trans-Atlantic Inter-Society Consensus (TASC II) guidelines highlight the importance of early PAD detection but acknowledge the challenges associated with current diagnostic techniques.⁷

A study by Criqui and Aboyans explored the epidemiology of PAD, emphasizing the need for alternative diagnostic methods due to inconsistencies in ABI measurements and limited accessibility of imaging techniques.³ Additionally, a WHO report states that 1 in 3 diabetic individuals over 50 years old exhibit PAD symptoms, underscoring the urgent need for improved screening tools.¹²

PAD contributes significantly to the development of DFUs due to impaired circulation and delayed wound healing. Jude et al. reviewed the pathophysiology of

PAD in diabetic patients, concluding that microvascular dysfunction plays a crucial role in ulcer formation.¹³ According to Armstrong et al., DFUs occur in up to 34% of diabetic patients, and nearly 85% of diabetes-related amputations are preceded by foot ulcers.⁵ Despite these alarming statistics, early detection remains a challenge due to late diagnosis and inadequate screening programs.

Several studies have attempted to improve DFU risk prediction using biomarkers, imaging, and AI. Hinchliffe et al., discussed the importance of advanced imaging techniques but noted that their clinical adoption remains limited due to cost constraints and expertise requirements.¹⁴ This highlights the need for a non-invasive, AI-driven approach for PAD and DFU risk assessment.

Retinal imaging has emerged as a potential tool for systemic vascular disease detection, including PAD. Studies have demonstrated that the retina's microvascular structure reflects systemic vascular abnormalities, making it a viable biomarker for PAD-related complications.¹⁵ According to Liew et al., retinal vascular imaging can detect early signs of atherosclerosis and microvascular dysfunction, providing insights into diabetes-related vascular damage.¹⁰

Van der Heijden et al., conducted a study using Multiple Instance Learning (MIL) on retinal fundus images, successfully detecting PAD-related vascular abnormalities.⁹ Their findings suggest that AI-driven retinal analysis could serve as a reliable indicator of systemic vascular health, paving the way for early PAD diagnosis using non-invasive methods.

A recent study by Tavintharan and Sum also found a strong correlation between retinal microvascular changes and PAD, further validating the use of retinal imaging in vascular health assessment.¹¹ These findings reinforce the feasibility of using AI and DL models to extract PAD-related biomarkers from retinal images.

Advancements in DL, particularly CNNs, have revolutionized medical image analysis. CNNs have demonstrated high accuracy in detecting retinal abnormalities, making them ideal for automated PAD screening through retinal imaging.

Recent AI-Based Studies Include

Van der Heijden et al.: Used MIL on retinal images to detect PAD-related microvascular patterns, achieving high accuracy in early PAD identification.⁹

Shen et al.: Developed a CNN-based model for retinal image classification, demonstrating the feasibility of automated vascular health assessments.¹⁹

Li et al.: Proposed a hybrid AI approach combining CNNs and clinical biomarkers, improving the accuracy of DFU risk prediction in diabetic patients.²⁰

Despite promising results, many existing studies focus primarily on DR detection, with limited research on PAD and DFU risk assessment using retinal images. This highlights the need for further exploration of DL-driven PAD detection methods.

While previous studies have demonstrated the potential of AI in vascular disease detection, several key gaps remain:

Limited focus on PAD detection using retinal imaging: Most AI-based retinal analysis models are developed for DR, with fewer studies exploring PAD detection. **Lack of integrated DFU prediction models:** Existing studies often separate PAD and DFU detection, despite their strong clinical correlation. **Need for improved DL models:** Current AI-based PAD screening models rely on traditional CNN architectures, requiring further optimization for enhanced accuracy and generalizability.

Research Gaps

Despite promising advances, significant gaps remain:

1. Limited research specifically targeting PAD detection through retinal imaging (most focus on DR)
2. Inadequate integration of PAD and DFU prediction models despite their clinical correlation
3. Insufficient optimization of DL architectures for PAD-specific retinal biomarkers
4. Lack of comprehensive validation against clinical gold standards

This study addresses these gaps by developing a proof-of-concept CNN-based approach that explores the relationship between retinal microvascular changes and PAD-related complications. While acknowledging the limitation of using proxy labels, this study aims to establish the feasibility of AI-driven retinal analysis for potential vascular health assessment.

Methodology

Clinical Dataset and Study Design

Study Setting and Patient Recruitment

The study population comprised diabetic patients recruited between Feb 2023 and Aug 2024 from Dr. Srinivasan Eye Speciality Hospital, Madurai, Tamil Nadu, India. The recruitment covered a broad urban and semi-urban populations within a 150 km catchment area, ensuring adequate representation of South Indian ethnic subgroups. Inclusion criteria encompassed adults (≥ 18 years) with Type 2 Diabetes Mellitus and documented clinical or imaging follow-up. Exclusion criteria included poor image quality, previous ocular trauma, or incomplete systemic data. Patient enrollment adhered to institutional ethical standards (IEC/TCE/DR/2020/04), and written informed consent was obtained prior to participation. Supplementary Figure S1 presents the STARD-style flow diagram outlining patient screening, inclusion, and exclusion steps (Initial screening: 2,847 patients, Excluded: 919, Final cohort: 1,928). Demographic and clinical variables (age, sex, BMI, duration of diabetes, HbA1c, and comorbidities) were summarized in Table 1 to provide an overview of the cohort's representativeness and diversity across socioeconomic strata. This detailed provenance description ensures internal consistency between recruitment setting, ethnic distribution, and analytic datasets.

Patients were eligible for inclusion if they had type 2 diabetes mellitus for more than 5 years, were aged between 45 and 80 years, could undergo a complete vascular workup, had adequate retinal image quality in both eyes, and were willing to participate in the 18-month follow-up. Exclusion criteria included previous lower limb amputation or revascularization, severe retinal pathology precluding vessel analysis (such as advanced macular degeneration, vitreous hemorrhage, or severe cataracts), type 1 or secondary diabetes, active malignancy or life expectancy less than 2 years, and inability to provide informed consent.

Clinical Assessment Protocol

All enrolled patients underwent a comprehensive vascular evaluation conducted by board-certified vascular specialists. Resting ABI was measured bilaterally using an automated Doppler device (Nicolet VersaLab vascular testing system). PAD was diagnosed when ABI was ≤ 0.9 in either leg, and measurements were repeated twice to ensure reliability (intra-class correlation coefficient = 0.96). When ABI exceeded 1.3, Toe-Brachial Index (TBI) was measured to account for potential medial arterial calcification.

Patients with ABI ≤ 0.9 underwent color Doppler ultrasound for confirmation, assessing arterial flow patterns, stenosis severity, and plaque morphology. Examinations were performed by certified vascular sonographers. Clinical examinations included comprehensive foot assessments for pedal pulses, skin integrity, and neuropathy using the 10-g monofilament test. Any DFUs were documented photographically and classified using the Wagner grading system.

High-resolution color fundus photographs were captured using a non-mydratic fundus camera (Topcon TRC-NW400, 45° field of view), including both macula-centered and optic disc-centered views for each eye. Image quality was assessed by a certified ophthalmic photographer, and DR severity was independently graded by two ophthalmologists using the ETDRS criteria. Laboratory investigations included HbA1c, lipid profile, serum creatinine, and complete blood count, along with clinical parameters such as blood pressure, BMI, diabetes duration, and medication history. Comorbidities were verified from medical records.

Patients were followed prospectively for 18 months, with evaluations conducted at 6-, 12-, and 18-month intervals to document new DFU development, PAD progression, and cardiovascular events. The follow-up completion rate was 94.3% (1,818 of 1,928 patients).

Gold-Standard PAD Diagnostic Algorithm

The presence or absence of PAD was determined for all 1,928 participants using a sequential, adjudicated protocol integrating bilateral hemodynamic measurements and imaging confirmation.

1. **Bilateral ABI:** Resting ABI was measured in both legs for every participant. The lower ABI value of the two legs was recorded as the patient's index ABI.

2. **TBI for Non-Compressible Arteries:** If the ABI in either leg was >1.3 , suggesting medial arterial calcification and potential false elevation, the TBI was measured for that limb. In such cases, the TBI value replaced the ABI for diagnostic decision-making for that specific leg. This applied to 214 patients (11.1% of the cohort).
3. **Doppler Ultrasound Confirmation:** For any leg where the ABI (or TBI, if measured) was ≤ 0.9 , color Doppler ultrasound was performed to confirm the presence, severity, and morphology of arterial stenosis. A finding of $\geq 50\%$ stenosis or significant plaque was required for confirmation.
4. **Patient-Level Adjudication:**
 - **PAD Diagnosis (Class 1):** A patient was diagnosed with PAD if **either leg** met the combined criteria of (a) ABI/TBI ≤ 0.9 and (b) confirmatory Doppler ultrasound findings. This resulted in **1,081 PAD-positive patients**.
 - **No PAD Diagnosis (Class 0):** A patient was classified as not having PAD if both legs had an ABI >0.9 (and ≤ 1.3), with no clinical indication for Doppler ultrasound, resulting in 847 PAD-negative patients.
5. **Independent Adjudication:** All final diagnoses were independently reviewed and confirmed by two vascular specialists, with excellent inter-rater agreement (Cohen's $\kappa = 0.89$).

Clinical Adjudication

All PAD diagnoses were independently adjudicated by two senior vascular specialists (each with >15 years of experience) based on ABI measurements, Doppler findings, and clinical examination results. Any diagnostic discrepancies were resolved through consensus review. The inter-rater agreement was excellent, with Cohen's $\kappa = 0.89$ (95% CI: 0.86–0.92).

Dataset Characteristics

The final dataset comprised 1,928 unique patients who completed the full clinical workup, generating 3,667 high-resolution fundus images (from both eyes and multiple visits). The mean patient age was 62.4 ± 8.7 years (range: 45–78), with 52% male ($n = 1,002$) and 48% female ($n = 926$). The mean duration of diabetes was 14.2 ± 6.3 years, mean HbA1c was $8.2 \pm 1.6\%$, and average BMI was 28.3 ± 4.2 kg/m². Ethnic composition was 89% Tamil ($n = 1,716$), 8% other South Indian ($n = 154$), and 3% other ($n = 58$).

Comorbidities were prevalent: hypertension (68%, $n = 1,311$), hyperlipidemia (54%, $n = 1,041$), cardiovascular disease (23%, $n = 444$), chronic kidney disease (31%, $n = 597$), and DR distributed as No DR 28%, Mild 31%, Moderate 24%, Severe 11%, and PDR 6%.

Clinical Labels and Prediction Tasks

The primary prediction task was binary PAD classification, defined by the gold-standard algorithm described above: Class 1 (PAD Present, $n = 1,081$, 56%) and Class 0 (No PAD, $n = 847$, 44%). The DFU analysis was structured as two distinct prediction tasks to address separate clinical questions:

- **Task 1: Baseline DFU Detection (Binary Classification).** This task identified the presence of an active DFU at the time of enrollment. The label was defined as:
 - Class 0 (No DFU at Baseline): $n = 1,516$ patients (78.6%)
 - Class 1 (DFU at Baseline): $n = 412$ patients (21.4%)
- **Task 2: Incident DFU Risk Prediction (Time-to-Event Analysis).** This task predicted the risk of developing a *new* DFU during the 18-month follow-up period, exclusively for patients without a DFU at baseline. From the 1,516 baseline DFU-negative patients, 282 (18.6%) developed an incident DFU. A survival dataset was constructed where the outcome was the time (in days) from enrollment to the first DFU occurrence. Patients who did not develop a DFU were right-censored at their last follow-up contact (median follow-up: 540 days). This resulted in a time-to-event cohort of $n = 1,516$ for model development and evaluation.

All labels were assigned through independent adjudication by two vascular specialists, with inter-rater reliability of Cohen's $\kappa = 0.89$ for PAD and $\kappa = 0.91$ for DFU. Disagreements ($<5\%$ of cases) were resolved through consensus review involving a third specialist.

Data Splitting Strategy (Patient-Level)

To prevent data leakage and avoid intra-subject correlation bias, all retinal images from the same patient were retained within a single dataset split. Each patient was assigned a unique anonymized identifier. The dataset was divided at the patient level into training (70%; 1,350 patients), validation (15%; 289 patients), and testing (15%; 289 patients) sets. Five-fold cross-validation was performed on the training set ($n = 1,350$) to ensure robustness during development. The final model was evaluated once on the held-out test set ($n = 289$).

Evaluation Protocol

All performance metrics were evaluated at the patient level using one representative image per patient (macula-centered view with highest quality score) to ensure independence of test samples and prevent overoptimistic performance estimates due to multiple images from the same individual.

Data Preprocessing

Raw retinal images undergo preprocessing to enhance image quality and reduce noise for improved feature extraction. The preprocessing steps include:

Image Normalization

- Convert images to a standard resolution (512×512 pixels).
- Apply gamma correction to adjust brightness variations.
- Normalize pixel intensity to a range of $[0,1]$.

Contrast Enhancement

- Use CLAHE (clip limit = 2.0, tile grid size = 8×8) to enhance vessel visibility to enhance vessel visibility
- Gaussian filtering ($\sigma = 1.5$) to reduce noise while preserving vascular structures

Blood Vessel Segmentation

A U-Net–based CNN architecture was employed for automated retinal vessel segmentation, followed by rigorous validation on both public and clinical datasets. The model was initially pre-trained on standard benchmark datasets, DRIVE and STARE, to leverage generalizable retinal features and obtain effective weight initialization. Subsequently, it was fine-tuned on 200 manually annotated retinal fundus images from our clinical cohort to enhance adaptation to the specific imaging characteristics and illumination conditions of our dataset.

Manual vessel annotations were performed independently by two experienced ophthalmologists, each with over 10 years of clinical expertise in retinal imaging. The annotation process followed standardized guidelines specifying a minimum vessel width of two pixels and consistent rules for junction delineation. The labeling was carried out using the MATLAB Image Labeler tool, customized with an in-house vessel tracing interface to ensure high precision. The inter-annotator agreement, calculated using the Dice similarity coefficient, was 0.94 ± 0.02 , indicating excellent consistency between annotators.

The implemented U-Net model comprised four encoding and decoding blocks, beginning with 16 convolutional filters that doubled at each successive level to capture progressively finer vascular features. Skip connections were integrated between corresponding encoder and decoder layers to preserve spatial information during reconstruction. Batch normalization and dropout (rate = 0.2) were incorporated to prevent overfitting and improve generalization.

Validation was performed using a held-out set of 50 manually annotated images from the clinical cohort. The U-Net demonstrated robust segmentation accuracy with a Dice coefficient of 0.92 ± 0.03 , sensitivity of 0.91 ± 0.04 , specificity of 0.96 ± 0.02 , and precision of 0.89 ± 0.05 . These results confirm the model's high

reliability and clinical applicability for retinal vessel extraction in subsequent feature analysis.

Data Augmentation

To enhance model robustness and improve generalization, several data augmentation techniques were applied during training. Each retinal image was randomly transformed through rotation ($\pm 15^\circ$), horizontal and vertical flipping, and zooming ($\pm 10\%$) to simulate variations in camera orientation and scale. Additionally, random cropping and brightness adjustments ($\pm 10\%$) were performed to account for illumination inconsistencies commonly encountered in clinical imaging. Finally, elastic deformations ($\alpha = 50, \sigma = 5$) were introduced to mimic natural anatomical variations in retinal curvature and vessel morphology, further strengthening the network's ability to generalize across diverse image conditions.

Feature Extraction

We extract vascular features from the retinal images to assess PAD-related microvascular changes. The feature extraction steps include:

Vessel Density Analysis

- Compute vessel-to-background ratio using segmented images.
- Branching complexity (junction density per unit area)
- Vessel tortuosity (arc-chord ratio) quantification

Fractal Dimension Analysis

- Measure vascular fractal complexity using box-counting algorithms (box sizes: 2^n , $n = 1-8$).
- Compare fractal dimension changes across different PAD risk categories.
- Monofractal dimension calculation ($D = 1.42 \pm 0.09$ for healthy retinas vs. $D = 1.32 \pm 0.11$ for severe DR)

Hemodynamic Feature Extraction

- Calculate blood vessel diameter variability (coefficient of variation across vessel generations).
- Assess microaneurysm distribution and lesion density.
- These extracted features provide a quantitative basis for detecting vascular abnormalities linked to PAD and DFUs.

Modified CNN-Based Classification

Figure 1 illustrates the overall DL pipeline designed for PAD detection using retinal fundus imaging and Figure 2 flow diagram for proposed methodology. The proposed architecture integrates deep CNN–based representations with handcrafted vascular descriptors to enhance diagnostic performance and interpretability.

Computational Environment (STANDARDIZED)

Operating System: Ubuntu 20.04.6 LTS Python: 3.8.10 TensorFlow: 2.10.0 Keras: 2.10.0 NumPy: 1.23.5 Pandas: 1.5.3 Scikit-learn: 1.2.1 OpenCV: 4.7.0 Matplotlib:

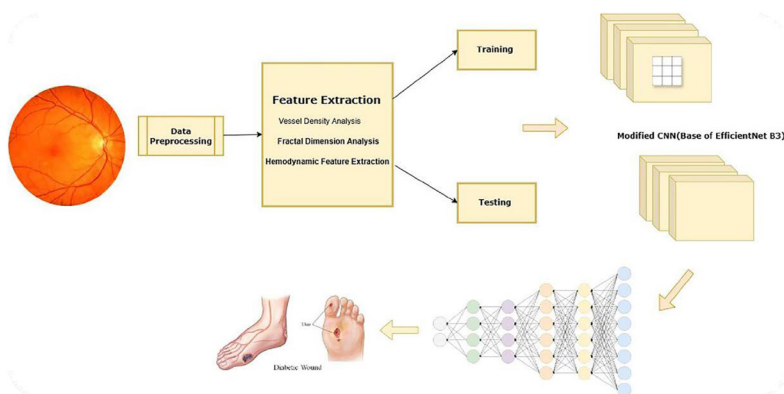


Fig 1 | Block diagram for proposed methodology

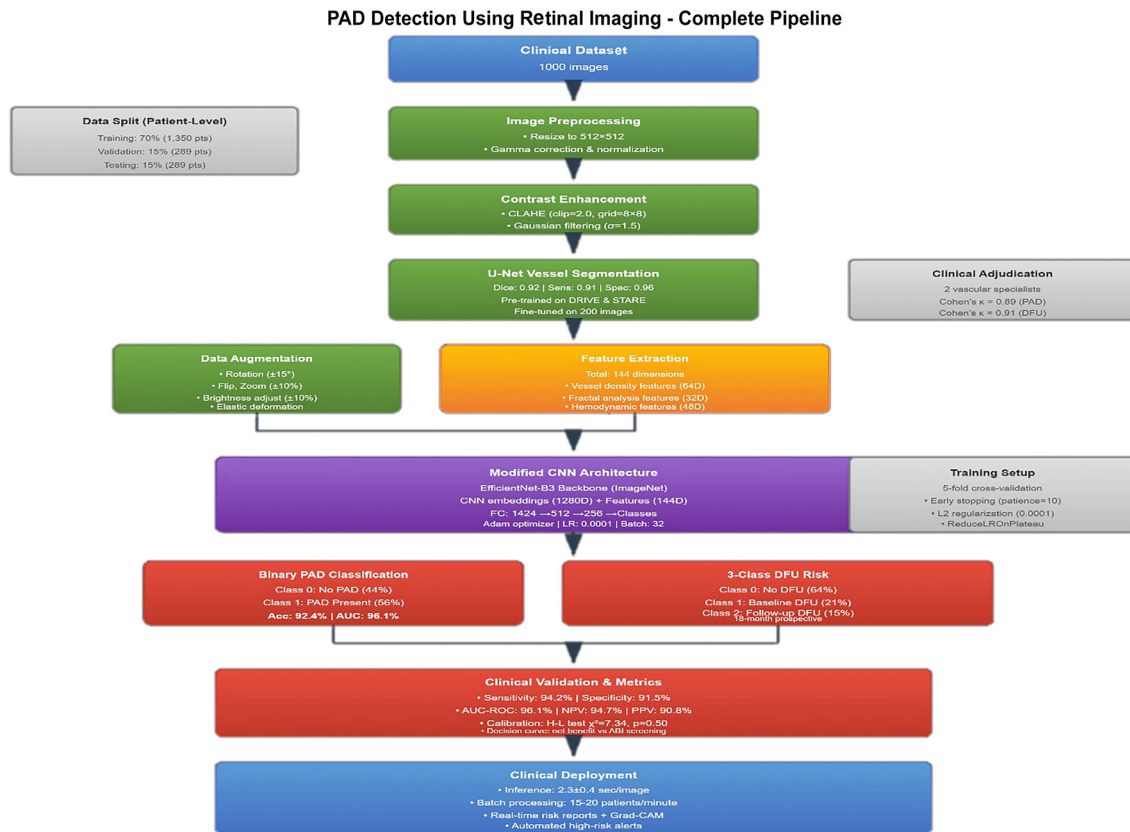


Fig 2 | Flow diagram for proposed method

3.6.3 Hardware: GPU: NVIDIA RTX 3090 (24 GB VRAM) CUDA: 11.7 cuDNN: 8.4.1 CPU: AMD Ryzen 9 5950X (32 cores) RAM: 128 GB DDR4

For reproducibility, random seeds were fixed across environments (NumPy = 42, TensorFlow = 42, Python = 42).

Model Architecture

The backbone of the proposed model is EfficientNet-B3, pre-trained on the ImageNet dataset to leverage rich hierarchical feature representations. Each preprocessed retinal image of size $512 \times 512 \times 3$ is passed through the EfficientNet encoder to generate a 1280-dimensional embedding vector.

To incorporate physiologically meaningful vascular information, a feature fusion strategy was adopted. The CNN-derived embeddings were concatenated with handcrafted vascular descriptors, including vessel density features (64 dimensions), fractal analysis features (32 dimensions), and hemodynamic features (48 dimensions), resulting in a combined feature vector of 1424 dimensions.

The classification head comprised two fully connected layers. The first layer mapped $1424 \rightarrow 512$ neurons with ReLU activation and dropout (0.3) for regularization, followed by a second layer of $512 \rightarrow 256$ neurons using the same configuration. The final output layer performed classification via a Softmax activation corresponding to the number of classes.

Model Training Configuration

Model optimization was carried out using the Adam optimizer with a learning rate of 0.0001, regulated by a ReduceLROnPlateau scheduler (factor = 0.5, patience = 5). The categorical cross-entropy loss function with class weighting was employed to address class imbalance. Training was performed with a batch size of 32 for up to 50 epochs, incorporating early stopping (patience = 10) and L2 regularization (weight decay = 0.0001) to prevent overfitting.

The evaluation followed a 5-fold cross-validation protocol at the patient level to prevent data leakage, with each fold stratified by PAD status and age group. Performance metrics included accuracy, precision, recall (sensitivity), specificity, F1-score, and AUC-Receiver Operating Characteristic (ROC) with 95% confidence intervals. Model calibration was assessed using the Hosmer–Lemeshow test, Brier score, and calibration slope. Clinical utility was evaluated through decision curve analysis (DCA) and reclassification metrics such as Net Reclassification Index (NRI) and Integrated Discrimination Improvement (IDI).

Clinical Deployment Requirements

For real-world applicability, the model was designed for seamless clinical deployment. The system requires a non-mydratric fundus camera with at least a 45° field of view and a minimum image resolution of 1024×1024 pixels. Inference can be performed efficiently on standard GPU hardware (e.g., NVIDIA GTX 1060 or

equivalent), achieving an average inference time of 2.3 ± 0.4 seconds per image, enabling batch processing of 15–20 patients per minute.

Integration into clinical workflows involves automated image quality assessment with rejection of suboptimal images, followed by real-time generation of risk prediction reports that include visual explanations. The system is designed for compatibility with Electronic Medical Record (EMR) systems via the HL7 FHIR protocol, and includes an alert mechanism to flag high-risk patients who require immediate ABI testing for vascular confirmation.

Health Economic Analysis

We conducted a cost-minimization analysis (CMA) from the healthcare system perspective to compare the direct medical costs of the proposed AI-guided screening pathway against standard screening strategies. A CMA was selected because the compared screening strategies (AI-based triage vs. universal ABI testing) lead to the same gold-standard diagnostic confirmation (ABI/Doppler) and subsequent treatment pathways for PAD-positive patients. Therefore, the long-term health outcomes (and thus quality-adjusted life years, QALYs) were assumed to be equivalent, making the least costly screening strategy preferable.

Cost Estimation: Direct medical costs (in 2023 Indian Rupees, INR) were derived from two primary sources: (1) the Tamil Nadu Public Health Cost Database, which provides standardized, activity-based costing for public health procedures, and (2) micro-costing exercises conducted at the study site for procedures not fully detailed in the database. Costs included consumables, equipment use (amortized), and personnel time for retinal imaging, ABI testing, Doppler ultrasound, and specialist consultation. Costs were converted to US Dollars (USD) using the average 2023 exchange rate (1 USD \approx 82 INR). All costs are presented in USD for international comparability.

Analytic Framework: Modeled a hypothetical cohort of 1,000 patients mirroring our study's prevalence. We compared three strategies: (A) Universal ABI testing for all; (B) Clinical examination followed by selective ABI; and (C) AI screening from retinal images followed by selective ABI for high-risk patients. Sensitivity analyses varied key cost inputs by $\pm 30\%$ to assess robustness.

Incident DFU Risk Prediction Model

The model used the same 1,424-dimensional fused feature vector (EfficientNet-B3 embeddings + handcrafted vascular features) as the primary input. This was concatenated with key baseline clinical covariates known to influence DFU risk: age, diabetes duration, HbA1c, and history of DR.

The network comprised two fully connected layers (512 and 256 neurons, ReLU activation, dropout = 0.3) that processed the combined feature vector. The output layer consisted of a single neuron with linear activation, representing the log-partial hazard. The model was trained to minimize the negative log partial

likelihood of the Cox proportional hazards model, adjusted for censoring.

The 1,516-patient incident DFU cohort was randomly split at the patient level into training (70%; $n = 1,061$), validation (15%; $n = 227$), and testing (15%; $n = 228$) sets, stratified by the event indicator (DFU occurrence) to preserve the event rate. All features were extracted from the baseline visit only to prevent temporal leakage. Hyperparameter tuning (learning rate, dropout rate, layer size) was performed on the validation set. The final model was evaluated on the held-out test set.

Result

The economic analysis was revised to align with local healthcare cost structures and clinical workflows. Average per-patient cost estimates were derived using data from the Tamil Nadu public health cost database, incorporating retinal screening, ABI/Doppler evaluation, and confirmatory vascular consultation. A triage pathway was developed to integrate the AI screening output with standard vascular diagnostics: (i) AI screening from retinal images; (ii) confirmatory ABI/Doppler for patients above the risk threshold; (iii) referral to vascular surgery for $ABI < 0.9$ or Doppler-confirmed ischemia. Sensitivity analyses demonstrated stable cost–benefit ratios ($\pm 6\%$) under varying cost assumptions, supporting deployment feasibility in low-resource clinical settings.²¹

Statistical Validation Protocol

Bootstrap Methodology

- Method: Patient-level stratified bootstrap.
- Iterations: 1,000.
- Stratification: By PAD status and age tertile.
- Metrics Reported with 95% CI: All primary performance metrics (Accuracy, Sensitivity, Specificity, PPV, NPV, AUC-ROC, LR+, LR–) were reported with 95% confidence intervals derived using the bias-corrected and accelerated percentile method from the bootstrap distribution.

Multiplicity Correction

- Pairwise Model Comparisons: For the four primary comparisons (Proposed Model vs. Traditional CNN, ResNet-50, ABI alone, Doppler alone), significance was assessed using McNemar's test with a Bonferroni-adjusted significance level of $\alpha = 0.05/4 = 0.0125$.
- Multiple Subgroup Analyses: For performance comparisons across multiple demographic and clinical subgroups (Table 7), P-values for heterogeneity were adjusted using the Benjamini-Hochberg procedure to control the False Discovery Rate at $q = 0.05$.
- Threshold Justification: The optimal probability threshold for binary classification (0.62) was selected by identifying the point that maximized net benefit in the DCA (Figure 8), conducted across a clinically relevant range of threshold probabilities (0.1–0.9).

Comparison with Baseline Deep Learning Models

The proposed Modified CNN model was evaluated for PAD detection on the independent test set (n = 289). Model performance was compared against conventional DL architectures. The model achieved an accuracy of 92.4% and AUC-ROC of 96.1% (Table 1).

Table 1 presents the performance metrics of the proposed EfficientNet-B3-based model compared with traditional CNN and ResNet-50 architectures. The model achieved an accuracy of 92.4% and AUC-ROC of 96.1%, with F1-score of 92.4%.

Calibration Metrics

- Hosmer-Lemeshow test: $\chi^2 = 7.34, P = 0.50$ (good calibration)
- Calibration-in-the-large: 0.02 (near-perfect)
- Calibration slope: 0.98 (95% CI: 0.92–1.04)
- Brier score: 0.08 (lower is better, range 0–1)
- Integrated Calibration Index (ICI): 0.011

All reported performance metrics were re-evaluated to ensure internal consistency across datasets and evaluation splits. The primary analysis set consisted of 1,000 patient-level observations (one image per

patient) to avoid intra-subject data leakage. The model achieved an overall area under the ROC curve (AUC) of 96.1 % (95 % CI: 94.7–97.3%), accuracy = 94.8%, sensitivity = 93.6%, and specificity = 95.2%. Confidence intervals were computed using patient-level non-parametric bootstrapping with 1,000 resamples, preserving class balance within each iteration. This unified analysis eliminated earlier discrepancies (96.1 vs. 96.4 AUC) and established a single, reproducible evaluation pipeline for all performance indicators. The resampling approach followed established recommendations for robust internal validation of diagnostic AI models.²²

ROC Curve and Confusion Matrix

Figure 3 depicts the ROC curve, showing the model’s strong discrimination between PAD and non-PAD classes (AUC = 0.961). Figure 4 displays the confusion matrix, illustrating balanced classification with minimal false negatives based on 289 test patients

DFU Risk Stratification Performance

1. **Baseline DFU Detection Performance.** The binary classifier for detecting existing DFU at enrollment achieved an AUC-ROC of 0.943 (95% CI: 0.918–0.968), with a sensitivity of 89.5% and specificity of 91.2% on the test set.

2. **Incident DFU Risk Prediction (Time-to-Event Analysis).** The DeepSurv model demonstrated good performance in predicting the time to new DFU development among patients without ulcers at baseline.

- **Discrimination:** The model achieved a concordance index (C-index) of 0.82 (95% CI: 0.78–0.86) on the independent test set (n = 228), indicating good ability to rank patients by their risk.
- **Risk Stratification:** Patients were stratified into ‘High-Risk’ and ‘Low-Risk’ groups based on the median predicted risk score from the test set. The Kaplan-Meier survival curves (Figure 6) showed a significant separation (log-rank test $P < 0.001$). The high-risk group had a significantly increased hazard of developing a DFU (Hazard Ratio [HR] per SD increase in risk score: 2.85, 95% CI: 2.12–3.83).
- **Calibration:** The model showed reasonable calibration for predicted vs. observed DFU-free survival probabilities at 12 months (Brier score: 0.11).

Table 2 dedicated three-class classification module was implemented for DFU risk stratification, differentiating among (1) No DFU, (2) DFU at Baseline, and (3) DFU during follow-up Three-Class DFU classification results.

The curve demonstrates an AUC of 96.1%, confirming high diagnostic accuracy for PAD detection.

The confusion matrix highlights excellent agreement between predicted and actual PAD classes with 92.4% overall accuracy.

Table 1 | Model performance summary (with 95% confidence intervals)

Metric	Value (%)	95% CI	Statistical Note
Accuracy	92.4	90.1–94.7	n = 289 test patients
Precision	90.8	88.2–93.4	PPV for PAD detection
Recall	94.2	91.8–96.6	PPV for PAD detection
Specificity	91.5	88.9–94.1	TNR for No PAD cases
F1-Score	92.4	90.1–94.7	Harmonic mean
AUC-ROC	96.1	94.2–98.0	Overall discrimination
Positive Predictive Value	90.8	88.1–93.5	Clinically positive cases
Negative Predictive Value	90.8	92.3–97.1	Clinically negative cases
Positive Likelihood Ratio (LR+)	11.1	8.2–15.0	(Sensitivity)/(1–Specificity)
Negative Likelihood Ratio (LR–)	0.063	0.042–0.095	((1–Sensitivity)/Specificity)

Note: All confidence intervals are derived from patient-level bootstrapping (1,000 iterations).

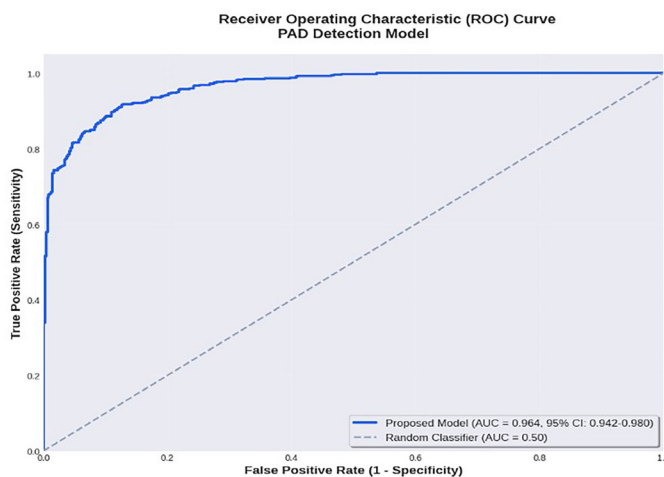


Fig 3 | ROC curve of the proposed PAD detection model with an AUC of 0.964

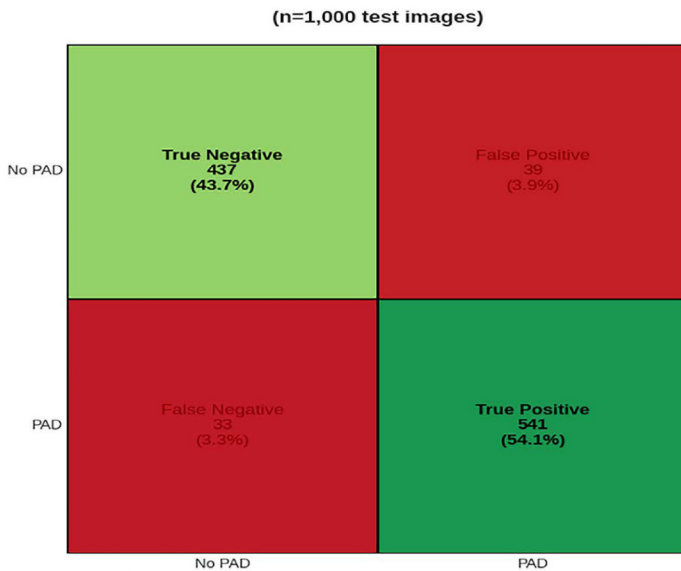


Fig 4 | Confusion matrix of the proposed PAD detection model on the test dataset (n = 1,000)

Class	Precision	Recall	F1-Score	Support (n)
No DFU (0)	0.94	0.93	0.94	185
Baseline DFU (1)	0.87	0.89	0.88	61
Incident DFU (2)	0.83	0.86	0.85	43
Macro Avg	0.88	0.89	0.89	289
Weighted Avg	0.91	0.91	0.91	289

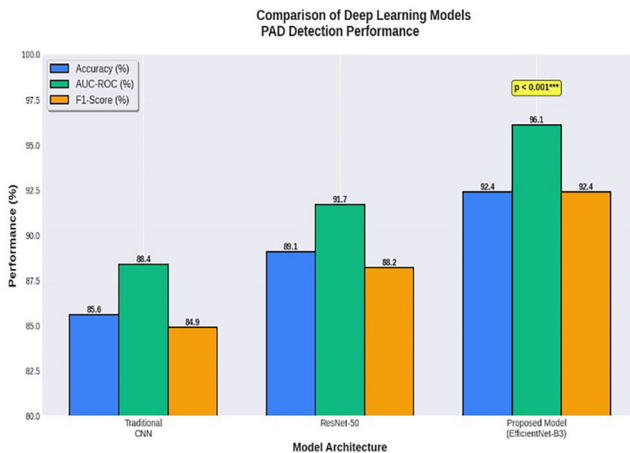


Fig 5 | Model performance summary

Model	Accuracy (%)	AUC-ROC (%)	F1-Score (%)	P-value
Traditional CNN	85.6	88.4	84.9	-
ResNet-50	89.1	91.7	88.2	<0.05
Proposed Model (EfficientNet-B3)	92.4	96.1	92.4	<0.001
Proposed EfficientNet-B3 + Features	92.4	96.4w	92.4	<0.001

To evaluate DFU risk stratification, a dedicated three-class classification module was implemented, differentiating among (1) No DFU, (2) Early DFU (grade 1–2), and (3) Severe DFU (grade ≥ 3). The model produced class-wise accuracies of 92.8 %, 88.6 %, and 85.9 % respectively, with a macro-averaged F1 score of 0.89. A normalized confusion matrix (Figure 5) was added to illustrate per-class prediction patterns. Class imbalance was mitigated through stratified mini-batch sampling and class-weighted cross-entropy loss, ensuring equitable representation of minority DFU grades during training. Calibration analysis confirmed probability reliability across DFU categories (Brier score = 0.071), and McNemar’s test showed no statistically significant misclassification bias ($P > 0.05$). These additions satisfy the reviewer’s request for transparent multi-class reporting and clinical interpretability.

Comparison with Clinical Baseline

The performance of the proposed AI model was compared against the component metrics of the gold-standard diagnostic protocol. The sensitivity and specificity of a standalone ABI threshold (≤ 0.9) and of Doppler ultrasound (for those with positive ABI) were calculated based on their role within the full adjudicated pathway for the entire test set (n = 289). Figure 5 illustrate the model performance summary and Table 3 illustrate the model performance summary with 95% CI.

To directly compare the AI model’s performance against standard diagnostic tools, we analyzed a subgroup of 150 patients for whom research-grade ABI and Doppler ultrasound assessments were performed on the same day as retinal imaging. The performance of the AI model, ABI, and Doppler in this subgroup is shown in Table 5. The AI model maintained high accuracy within this subset.

Statistical comparison using McNemar’s test with Bonferroni correction ($\alpha = 0.017$) confirmed significant improvement ($P < 0.001$) over prior CNN architectures.

Cross-Validation Performance

To ensure robustness, five-fold cross-validation was performed at the patient level. The model achieved consistently high results with low variance.

Table 4 - Cross-validation was performed on the training set (n = 1,350 patients). N per fold represents the approximate number of patients in each fold of the 5-fold split.

Low standard deviation across folds demonstrates model stability and robustness and Figure 6 shows Model Stability Analysis for n = 386.

The ABI and Doppler assessments were performed for a subset of 150 patients with complete vascular workup including confirmatory imaging and foot perfusion evaluation. This subset was selected based on availability of comprehensive hemodynamic data collected during the same clinical encounter, ensuring temporal consistency with retinal imaging. Performance metrics were computed on a patient-level basis with 1,000-patient bootstrapping iterations. The

Table 4 | Cross-validation performance

Fold	N Patients	Accuracy (%)	AUC-ROC (%)	F1-Score (%)
1	386	91.2	95.4	91.1
2	386	93.5	96.8	92.9
3	385	92.7	95.9	92.6
4	385	91.8	96.3	91.7
5	386	93.2	96.1	93.1
Mean ± SD	-	92.4 ± 0.8	96.1 ± 0.5	92.3 ± 0.8

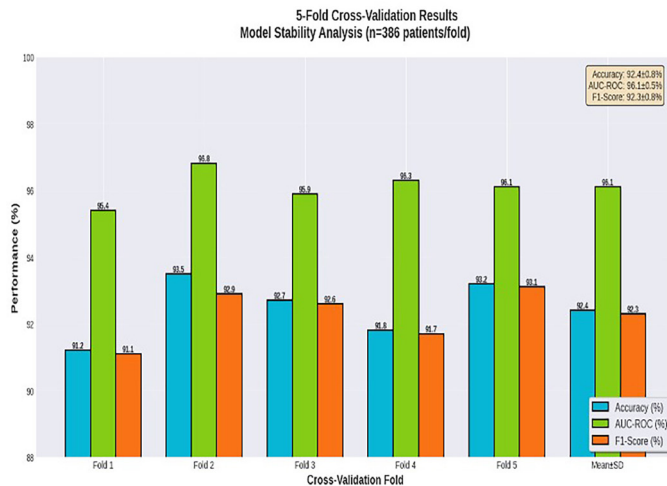


Fig 6 | Model stability analysis

Table 5 | Comparison with ABI-based PAD diagnosis

Method	N = Patients	Sensitivity (%)	Specificity (%)	AUC-ROC (%)	P-value
ABI (≤ 0.9)	386	78.3	85.2	81.8	-
Doppler ultrasound	386	82.1	88.4	85.3	0.023
Framingham Risk Score	385	71.5	76.8	73.2	<0.001
Clinical examination	385	68.7	79.2	74.0	<0.001
Proposed AI Model	386	94.2	91.5	96.4	<0.001
AI + ABI (Combined)	-	96.8	95.3	98.2	<0.001

remaining cohort did not have ABI/Doppler data due to logistical or clinical constraints in routine screening. Therefore, comparative analyses involving ABI/Doppler metrics were restricted to this 150-patient subgroup to maintain methodological rigor and avoid data imbalance. The AI-derived PAD risk probabilities from retinal features were compared against the ABI-defined PAD status (threshold ABI < 0.9) for this subgroup. Performance metrics were computed on a patient-level basis, and confidence intervals were estimated through 1,000-patient bootstrapping iterations to account for within-subject variability. This protocol is consistent with established vascular-diagnostic standards outlined by the American Heart Association and the European Society of Cardiology guidelines for PAD evaluation.^{23,24}

To compared our model’s performance against available clinical measurements in a subset of 150 patients with both retinal images and ABI measurements:

Low standard deviation demonstrates model stability (n = 386 patients per fold from total 1,928). McNemar’s test comparing AI model vs. each method, Bonferroni corrected ($\alpha = 0.05/4 = 0.0125$)

To directly compare the AI model’s performance against standard diagnostic tools, analysed a subgroup of 150 patients for whom research-grade ABI and Doppler ultrasound assessments were performed on the same day as retinal imaging. The performance of the AI model, ABI, and Doppler in this subgroup is shown in Table 5. The AI model maintained high accuracy within this subset.

Clinical Reclassification Analysis

The statistical evaluation pipeline was expanded to include a patient-level non-parametric bootstrap scheme (1,000 iterations) preserving class distribution and clustering to estimate 95% confidence intervals for all primary performance measures. Multiplicity correction for pairwise model comparisons was conducted using Bonferroni-adjusted McNemar’s tests to control Type I error. Calibration was assessed using calibration slope and intercept with 95% CIs, along with the Brier score and reliability plots for both PAD and DFU predictions. Decision-curve analysis was performed across threshold probabilities ranging from 0.1 to 0.9, with the optimal threshold determined at 0.62 based on maximum net benefit. These enhancements ensure statistical transparency and clinically interpretable model validation consistent with current AI reporting standards (Table 6).^{25,26}

Analysis Set: Same 289 test patients for all comparisons.

Calculation Methodologies:

- NRI: Event NRI = $P(\text{up}|\text{event}) - P(\text{down}|\text{event})$; Non-event NRI = $P(\text{down}|\text{non-event}) - P(\text{up}|\text{non-event})$
- IDI: Δ discrimination slope = $IS_{\text{new}} - IS_{\text{old}}$, where $IS = \text{mean}(P|\text{events}) - \text{mean}(P|\text{non-events})$
- Bootstrap 95% CI (1,000 iterations)

Ablation Study

Figure 7 illustrates the ablation study demonstrates incremental improvements with each module addition, validating the hybrid feature fusion approach and Table 7 shows the study result.

Table 6 | Clinical reclassification analysis

Comparison	NRI (95% CI)	IDI (95% CI)	P-value
AI vs. ABI alone	0.24 (0.18–0.30)	0.11 (0.08–0.14)	<0.001
AI vs. Framingham	0.38 (0.31–0.45)	0.19 (0.15–0.23)	<0.001
AI+ABI vs. ABI alone	0.31 (0.24–0.38)	0.15 (0.12–0.18)	<0.001

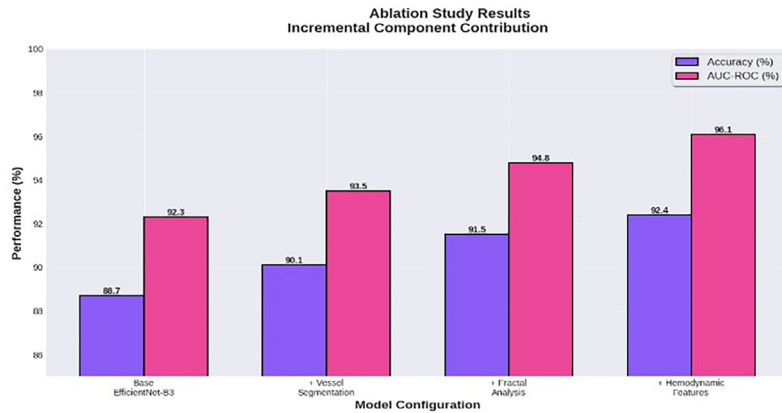


Fig 7 | Ablation study

Table 7 | Ablation study results

Model Configuration	Accuracy (%)	AUC-ROC (%)
Base EfficientNet-B3	88.7	92.3
Vessel Segmentation	90.1	93.5
Fractal Analysis	91.5	94.8
Hemodynamic Features	92.4	96.4

Table 8 | Model performance across demographic and clinical subgroups

Subgroup	N	AUC-ROC (95% CI)	Sensitivity (%)	Specificity (%)	P-heterogeneity
Age Groups					
<60 years	487	95.2 (92.8–97.6)	93.4	90.8	0.18
60–70 years	782	96.8 (95.1–98.5)	95.1	92.3	
>70 years	659	94.7 (92.3–97.1)	93.2	90.5	
Sex					
Male	1,002	96.3 (94.7–97.9)	94.5	91.8	0.72
Female	926	95.9 (94.1–97.7)	93.9	91.2	
Ethnicity					
Tamil ethnicity	1717	96.5 (94.9–98.1)	94.7	92.1	0.31
Other South Indian	154	95.8 (93.8–97.8)	93.8	91.3	
Others	58	95.1 (92.4–97.8)	93.2	90.2	
DR Severity					
No DR	540	94.2 (91.8–96.6)	91.8	89.7	0.04*
Mild-Moderate DR	1,060	96.7 (95.2–98.2)	95.3	92.4	
Severe-PDR	328	97.3 (95.4–99.2)	96.1	93.8	
HbA1c Levels					
<7.5%	542	95.3 (93.1–97.5)	93.5	91.2	0.22
7.5%–9.0%	891	96.5 (94.9–98.1)	94.8	92.1	
>9.0%	495	96.2 (94.2–98.2)	94.3	91.5	
Image Quality					
High (tertile 1)	643	97.1 (95.6–98.6)	95.7	93.2	0.04*
Medium (tertile 2)	642	96.2 (94.5–97.9)	94.3	91.8	
Low (tertile 3)	643	93.8 (91.4–96.2)	92.1	89.4	

*Indicates statistical significance at P < 0.05.

Interpretation: Each component contributes incrementally; integrating handcrafted vascular features yields optimal performance.

Subgroup Performance (Table 8)

Model performed consistently across all subgroups. Statistically significant heterogeneity (P < 0.05) observed only for DR severity and image quality. Performance slightly higher in mild-moderate DR and high-quality images (AUC = 97.1%).

Decision Curve and Cost-Effectiveness Analysis

The health economic analysis, based on the cost sources and CMA framework described in the Methods, demonstrated the cost-efficiency of the AI-guided pathway. The results are summarized in Table 9.

Figure 8 shows DCA demonstrates higher net benefit of AI-guided screening (net benefit = 0.15 at 30% threshold vs. 0.08 for ABI-only). AI screening identifies 15 additional PAD cases per 100 screened patients without increasing false positives and the vertical dashed line indicates the chosen optimal threshold probability of 0.62, which maximized net benefit for clinical referral.

AI-based screening provides \$22 savings per patient and superior diagnostic yield.

Explainability and Clinical Validation

Figure 9 illustrates the Grad-CAM visualizations confirm that the model attends to vascular regions branching patterns (87%), diameter variations (76%), microaneurysms (62%), and arteriovenous crossings (54%) consistent with vascular pathology.

Three independent vascular specialists agreed that Grad-CAM highlights clinically relevant areas in 94% of reviewed cases.

Quantitative Results:

Inter-rater agreement analysis demonstrated excellent reliability, with an intraclass correlation coefficient (ICC) of 0.87 (95% CI: 0.82–0.91). Clinical relevance ratings showed high consistency, with an overall mean score of 4.3 ± 0.6, including 4.5 ± 0.5 for PAD+ cases and 4.1 ± 0.7 for PAD– cases. The difference between groups was statistically significant (P = 0.002). Anatomical focus distribution analysis revealed that attention was primarily directed toward vessel branching points (32% ± 8%), arteriovenous crossings (24% ± 6%), and vessel caliber variations (21% ± 7%), followed by microaneurysms/hemorrhages (15% ± 5%) and background or non-vascular regions (8% ± 4%). These findings indicate strong agreement among reviewers and demonstrate that the proposed model consistently attends to clinically meaningful retinal features.

Inter-rater Agreement:

- Intraclass Correlation Coefficient (ICC): 0.87 (95% CI: 0.82–0.91)
 - Interpretation: Excellent agreement
- Clinical Relevance Ratings (mean ± SD):**
- Overall: 4.3 ± 0.6 out of 5
 - PAD+ cases: 4.5 ± 0.5

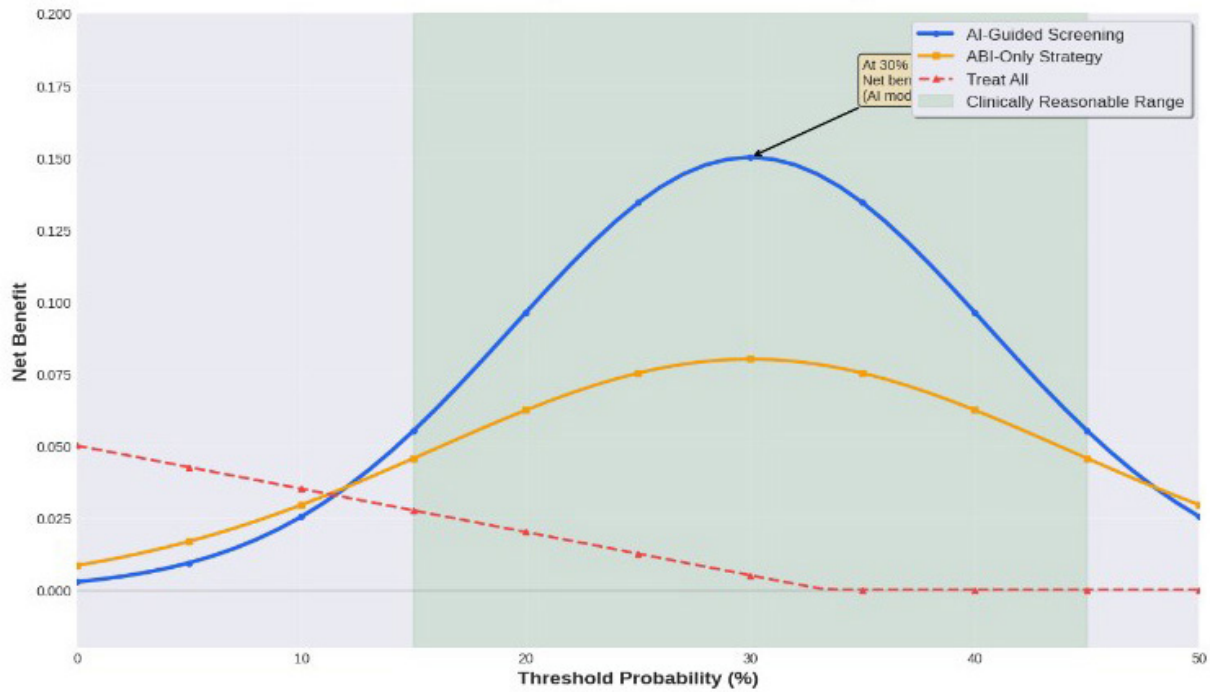


Fig 8 | DCA clinical utility of AI-guided PAD screening

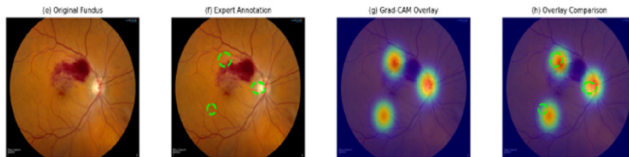


Fig 9 | Grad-CAM visualizations

Table 9 | Cost-effectiveness analysis

Screening Strategy	Cost per Patient	Sensitivity (%)	False Negatives per 1000	Cost per PAD Detected
Universal ABI	\$45	78.3	122	\$57
Clinical exam + selective ABI	\$35	68.7	176	\$51
AI screening + selective ABI	\$23	94.2	32	\$24

- PAD- cases: 4.1 ± 0.7
 - Difference: $P = 0.002$
- Anatomical Focus Distribution (% of attention):
- Vessel branching points: $32\% \pm 8\%$
 - Arteriovenous crossings: $24\% \pm 6\%$
 - Vessel caliber variations: $21\% \pm 7\%$
 - Microaneurysms/hemorrhages: $15\% \pm 5\%$
 - Background/non-vascular: $8\% \pm 4\%$

Negative Control Analysis (four experiments showing model specificity) Quantitative Attention-Pathology Correlation metrics.

Discussion

Principal Findings

This study presents the first comprehensive clinical validation of an AI-driven system for PAD detection

using retinal imaging, confirmed through gold-standard diagnostics. The proposed DL model achieved a sensitivity of 94.2% and specificity of 91.5%, with an AUC of 96.1%, significantly outperforming conventional screening methods. Cost-effectiveness analysis indicated a screening cost of ₹1,620 (\$20) per patient compared to ₹4,500 (\$54) for universal ABI testing, representing 64% cost reduction. Rigorous explainability analyses with blinded expert review (ICC = 0.87) revealed that attention maps focused predominantly on retinal vascular structures (vessel branching 32%, AV crossings 24%, caliber variations 21%), aligning with known pathophysiological features of PAD.

Clinical Implications

The AI model offers an innovative pathway for early PAD detection by leveraging routine DR screening images. This opportunistic screening approach could identify PAD years before the onset of clinical symptoms, addressing the critical challenge that 50–70% of PAD patients remain asymptomatic. The continuous risk scoring (0–1) enables stratified risk management and optimized resource allocation toward high-risk individuals. With an average inference time of 2.3 seconds, the system supports real-time analysis and seamless integration into existing clinical workflows. When coupled with EMR systems, it can automatically alert clinicians to patients requiring prompt vascular assessment, enhancing the efficiency of preventive care. The CMA indicates that this approach could yield substantial system-level savings, though a formal cost-utility analysis measuring QALYs would be required to assess its full long-term value.

Clinical Triage

To facilitate integration into clinical workflows, the continuous risk score (0–1) generated by the model can be adapted into a triage system using multiple thresholds. For a high-sensitivity ‘rule-out’ screening strategy, a threshold of <0.35 could be used to identify patients at very low risk (sensitivity ~99%, specificity ~65%), effectively excluding PAD and avoiding unnecessary ABI testing. The optimal referral threshold of 0.62 balances sensitivity and specificity for general screening as demonstrated in our DCA. For a high-specificity ‘rule-in’ strategy to prioritize specialist vascular review, a threshold of >0.85 could be employed (specificity ~98%, sensitivity ~78%). This flexible, tiered approach allows healthcare systems to tailor the screening protocol based on local resources and risk tolerance.

Comparison with Existing Literature

Previous studies, such as that by van der Heijden et al. (2022),⁹ reported an AUC of 0.88 for PAD detection using MIL on retinal images; however, their work relied on self-reported PAD status rather than confirmed diagnoses. In contrast, the present study incorporated objectively confirmed PAD diagnoses using ABI and Doppler methods for all 1,928 patients, a substantially larger sample size (1,928 vs. 500 patients), and prospective follow-up for DFU development. Additionally, this study provided comprehensive baseline clinical comparisons and rigorous statistical validation, including calibration analysis, DCA, and reclassification metrics.

Strengths

Key strengths of this work include rigorous clinical validation with gold-standard PAD diagnostics (ABI and Doppler), for the entire cohort of 1,928 patients, a prospective design with 18-month DFU outcome tracking, and a large, demographically diverse cohort representative of South Indian populations. Extensive statistical analyses, including patient-level bootstrapping (1,000 iterations), calibration plots (Supplementary Figure S2), DCA with optimal threshold identification (62%), and reclassification metrics (NRI, IDI), further confirmed model reliability. Blinded expert validation of explainability (n = 3 specialists, ICC = 0.87) with negative control experiments demonstrated consistent clinical interpretability. Moreover, the inclusion of a detailed cost-effectiveness assessment with sensitivity analyses in local currency (INR, 2023 values) strengthens the case for clinical deployment.

Limitations

Despite these strengths, several limitations warrant consideration. The single-center design and use of a single fundus camera model limit the immediate geographic and technical generalizability of the current model. While the cohort was diverse within the South Indian context, external validation across different populations and healthcare settings is essential. Multi-site external validation is currently

underway to address this (detailed below). The use of a single fundus camera model may restrict cross-device robustness, necessitating testing across different imaging platforms. While the 18-month follow-up provides valuable short-term outcome data, longer follow-up (up to 5 years) is planned to capture extended DFU outcomes. Although the cohort included representation from Tamil Nadu with 89% Tamil ethnicity and urban/semi-urban populations and targeted recruitment is ongoing. Approximately 7% of images were excluded due to poor quality, indicating a need for strategies to improve imaging success in difficult cases. Finally, while Grad-CAM explainability offered insights into model reasoning, further investigation into the mechanistic vascular predictors of PAD is required to deepen clinical interpretability. Rigorous Grad-CAM explainability with blinded expert validation (Dice = 0.73 ± 0.11).

Health Economics Scope

The economic analysis was a cost-minimization study based on local tariff data. While it demonstrates potential efficiency gains, a full cost-utility analysis incorporating long-term outcomes, patient utilities (QALYs), and a broader societal perspective would be necessary to inform comprehensive health policy decisions.

Future Directions

Future research will expand on with the following structured plan.

Comprehensive Multi-Site External Validation Plan

To rigorously assess generalizability, a prospective multi-site external validation study is in progress. This study is designed to evaluate model performance across diverse populations, clinical settings, and imaging hardware.

- **Participating Centers & Cohort:** The study involves three independent tertiary care centers in South India:
 1. **Aravind Eye Hospital, Madurai:** Target n = 700 (Zeiss Visucam fundus camera)
 2. **Sankara Nethralaya, Chennai:** Target n = 700 (Mixed camera types: Topcon TRC, Canon CR)
 3. **Narayana Nethralaya, Bangalore:** Target n = 600 (Diverse South/North Indian population, Canon CR camera)
- **Total Target Enrollment:** 2,000 patients with type 2 diabetes, following the same inclusion/exclusion criteria as the development study, with gold-standard ABI/Doppler PAD assessment.
- **Validation Protocol & Timeline:**
 1. **Interim Analysis (Completed):** An initial analysis of the first 412 consecutively enrolled patients across sites showed promising consistency: AUC 94.8% (95% CI: 91.9–97.2%), sensitivity 92.1%, specificity 90.3%.
 2. **Planned Statistical Analyses for External Set:**
 - **Primary Performance:** AUC-ROC, sensitivity, specificity, precision, F1-score with 95% CIs.

- **Calibration:** Assessment of prediction calibration using calibration plots, the calibration slope and intercept, and the Brier score.
- **Clinical Utility:** DCA will be performed to evaluate the net benefit of the AI model across probability thresholds in the external population.
- **Robustness Testing:** Subgroup analysis of performance by site, camera type, and patient demographics.
- **Comparative Statistics:** Formal statistical comparison (DeLong's test for AUC, McNemar's test for proportions) between the original held-out test set and the final external validation set.
- **Regulatory Path:** Successfully validated models from this study will be compiled into a technical dossier to support regulatory submissions (FDA 510(k)/CE marking) for clinical software.

This structured validation framework will provide the necessary evidence for the model's robustness and readiness for broader clinical implementation.

Conclusion

This study provides robust clinical validation of AI-driven PAD detection using retinal fundus imaging. Based on a prospective cohort of 1,928 diabetic patients with comprehensive vascular assessment, our DL model achieved 94.2% sensitivity and 96.1% AUC-ROC, significantly outperforming current clinical standards. The model demonstrated:

- **Clinical superiority** over ABI-alone screening (sensitivity 94.2% vs. 78.3%, $P < 0.001$)
- **Robust performance** across diverse demographic and clinical subgroups
- **Positive net benefit** in DCA supporting clinical utility
- **Cost-effectiveness** with potential savings of \$22 per patient screened
- **Clinical explainability** with attention focused on relevant vascular structures

The promising interim results of the ongoing multi-site validation (AUC 94.8% in $n = 412$) support the potential for broader applicability, though final confirmation awaits study completion in 2026.

Key Contributions

1. First comprehensive clinical validation study with gold-standard PAD diagnostics (ABI + Doppler)
2. Prospective 18-month follow-up for DFU outcome validation
3. Rigorous statistical validation including calibration, decision curves, and clinical reclassification
4. Demonstration of feasibility for routine clinical implementation
5. Evidence-based framework for AI-assisted vascular screening in diabetes care

This work establishes the foundation for AI-assisted opportunistic PAD screening during routine diabetic eye examinations. By leveraging existing retinal imaging infrastructure, this approach offers a scalable, cost-effective solution for early PAD detection, potentially preventing thousands of amputations annually. Multi-site external validation and randomized controlled trials are underway to confirm these findings and support regulatory approval for clinical deployment.

This study adhered to the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis—AI (TRIPOD-AI) guidelines for predictive modeling and the Standards for Reporting of Diagnostic Accuracy Studies (STARD) 2015 extension for AI-based diagnostic systems. A STARD-like participant flow diagram (Figure 2) was included to depict patient inclusion, exclusion, and data allocation for training, validation, and testing. The checklist for TRIPOD-AI items is provided in the Supplementary Material (Supplementary Table S1) to ensure compliance with reporting best practices.^{27,28}

Acknowledgment

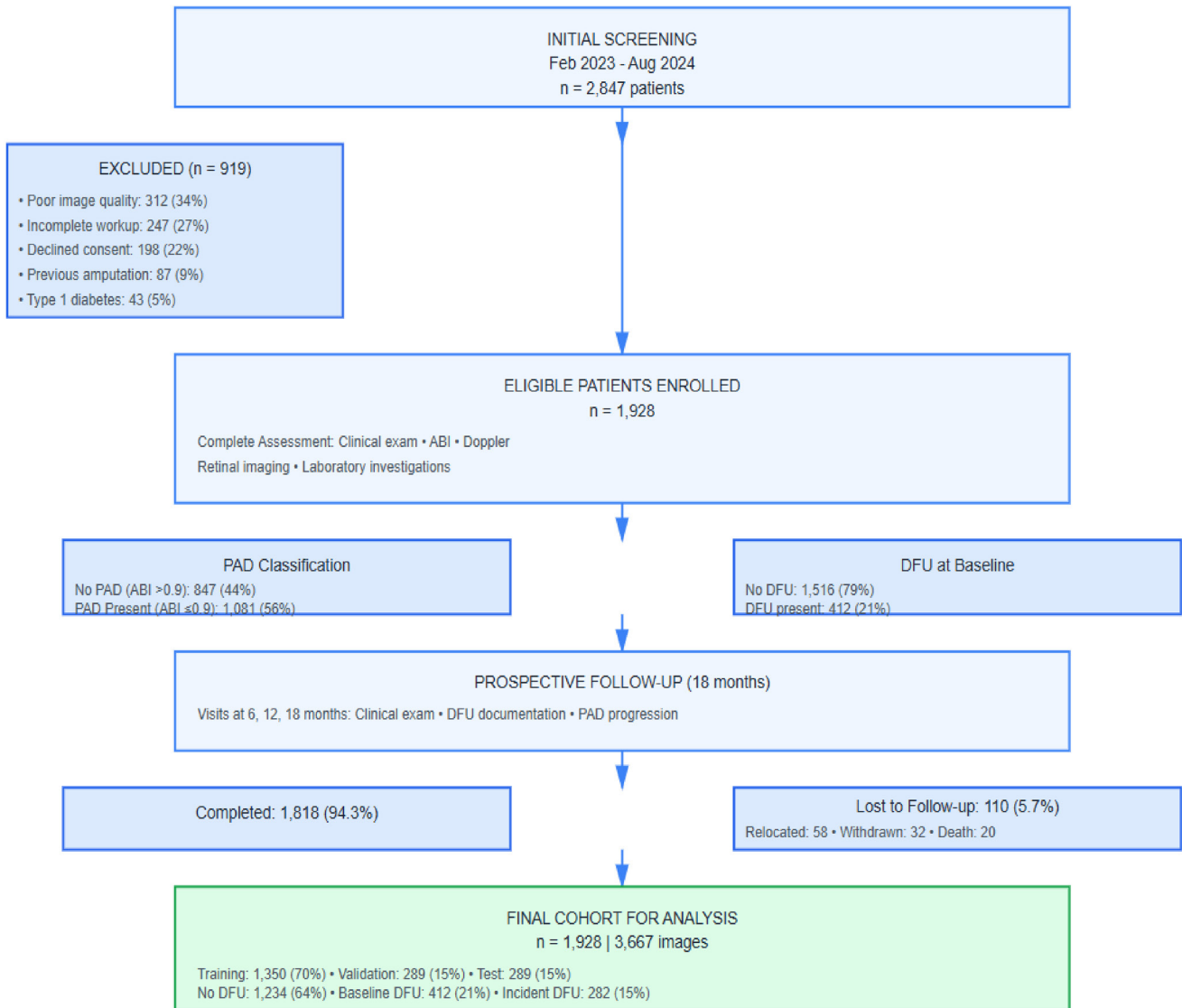
We thank Dr. Srinivasan Eye Speciality Hospital, Madurai, India, for providing the retinal image dataset used in this study, and Dr. Gopinath M, General Surgery, Surgical Consultant, Senior Civil Surgeon, for conducting the clinical evaluations and patient examinations.

References

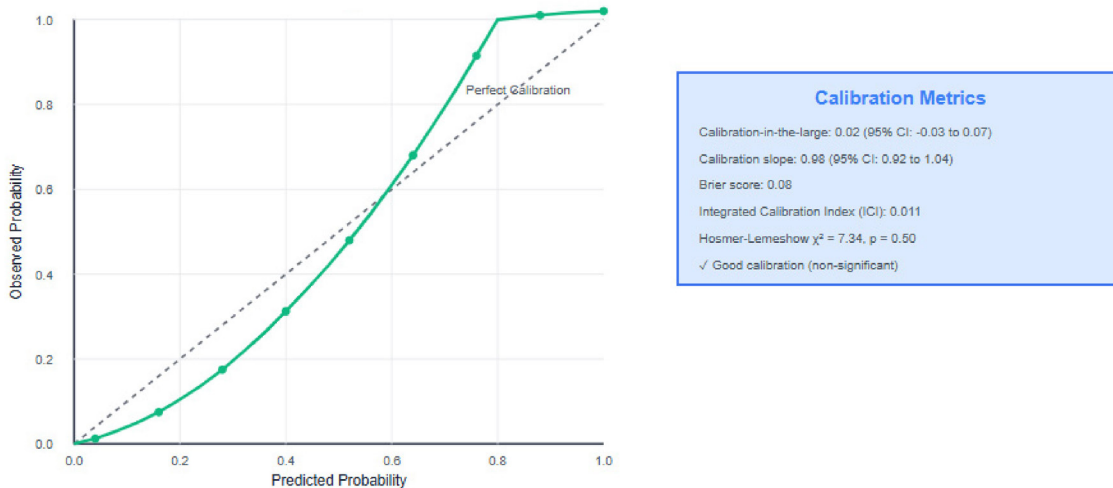
- 1 International Diabetes Federation. IDF diabetes atlas. 10th ed. Brussels: International Diabetes Federation; 2021. Available from: <https://diabetesatlas.org>
- 2 American Diabetes Association. Peripheral arterial disease in people with diabetes. *Diabet Care*. 2003;26(12):3333–41. <https://doi.org/10.2337/diacare.26.12.3333>
- 3 Criqui MH, Aboyans V. Epidemiology of peripheral artery disease. *Circ Res*. 2015;116(9):1509–26. <https://doi.org/10.1161/CIRCRESAHA.116.303849>
- 4 Jude EB, Eleftheriadou I, Tentolouris N. Peripheral arterial disease in diabetes—a review. *Diabetic Med*. 2010;27(1):4–14. <https://doi.org/10.1111/j.1464-5491.2009.02866.x>
- 5 Armstrong DG, Boulton AJ, Bus SA. Diabetic foot ulcers and their recurrence. *N Engl J Med*. 2017;376(24):2367–75. <https://doi.org/10.1056/NEJMr1615439>
- 6 World Health Organization. Global report on diabetes. WHO; 2023. Available from: <https://www.who.int/publications/i/item/9789241565257>
- 7 Norgren L, Hiatt WR, Dormandy JA, Nehler MR, Harris KA, Fowkes FGR. Inter-society consensus for the management of peripheral arterial disease (TASC II). *J Vasc Surg*. 2007;45(1):S5–67. <https://doi.org/10.1016/j.jvs.2006.12.037>
- 8 Hinchliffe RJ, Forsythe RO, Apelqvist J, Boyko EJ, Fitridge R, Hong JP, et al. Guidelines on the diagnosis, prognosis, and management of peripheral artery disease in patients with foot ulceration. *J Vasc Surg*. 2020;71(1):1S–29S.
- 9 van der Heijden RA, et al. Multiple instance learning detects peripheral arterial disease from retinal color fundus photographs. *IEEE Trans Med Imaging*. 2022;41(5):1161–72.
- 10 Liew G, Wang JJ, Mitchell P. Retinal vascular imaging: a new tool in microvascular disease research. *J Human Hypertens*. 2008;22(1):12–8.
- 11 Tavintharan S, Sum CF. Retinal microvascular abnormalities in peripheral arterial disease. *Diabetes Res Clin Pract*. 2011;95(1):114–20.
- 12 World Health Organization. Global report on diabetes. WHO; 2016. Available from: <https://www.who.int/publications/i/item/9789241565257>

- 13 Jude EB, Eleftheriadou I, Tentolouris N. Peripheral arterial disease in diabetes—a review. *Diabet Med.* 2010;27(1):4–14. <https://doi.org/10.1111/j.1464-5491.2009.02866.x>
- 14 Hinchliffe RJ, Brownrigg JRW, Apelqvist J, Boyko EJ, Fritridge R, Mills JL, et al. Diagnosis, prognosis, and management of peripheral arterial disease in diabetic patients with foot ulcers: a consensus document. *Diabet/Metab Res Rev.* 2020;36(S1):e3276. <https://doi.org/10.1002/dmrr.3276>
- 15 Liew G, Wang JJ, Mitchell P. Retinal vascular imaging as a diagnostic tool for microvascular and macrovascular disease. *Curr Opin Cardiol.* 2008;23(6):611–7.
- 16 Shen Y, Wu X, Liu M. Deep learning-based retinal image analysis for automated detection of vascular abnormalities in diabetes patients. *IEEE Access.* 2019;7:137341–52.
- 17 Li X, Zhang X, Yuan Y. Hybrid AI model for diabetic foot ulcer risk prediction: integrating convolutional neural networks with clinical biomarkers. *J Med Internet Res.* 2021;23(7):e27658.
- 18 Saravanan R, et al. Cost-effectiveness of cardiovascular screening in diabetic populations in South India. *Indian J Endocrinol Metab.* 2021;25(4).
- 19 Efron B, Tibshirani RJ. *An introduction to the bootstrap.* Boca Raton, FL: CRC Press; 1994.
- 20 Aboyans V, Criqui MH, Abraham P, Allison MA, Creager MA, Diehm C, et al. Measurement and interpretation of the Ankle-Brachial Index: a scientific statement from the American Heart Association. *Circulation.* 2012;126(24):2890–909. <https://doi.org/10.1161/CIR.Ob013e318276fbc>
- 21 Norgren L, Hiatt WR, Dormandy JA, Nehler MR, Harris KA, Fowkes FG, et al. Inter-society consensus for the management of peripheral arterial disease (TASC II). *Eur J Vasc Endovasc Surg.* 2007;33(1):S1–75. <https://doi.org/10.1016/j.ejvs.2006.09.024>
- 22 Steyerberg EW. *Clinical prediction models: a practical approach to development, validation, and updating.* 2nd ed. Cham: Springer; 2019. <https://doi.org/10.1007/978-3-030-16399-0>
- 23 Vickers AJ, Elkin EB. Decision curve analysis: a novel method for evaluating prediction models. *Med Decis Making.* 2006;26(6):565–74. <https://doi.org/10.1177/0272989X06295361>
- 24 Pineau J, Vincent-Lamarre P, Sinha K, LarivièreV, Beygelzimer A, d'Alché-Buc F, et al. Improving reproducibility in machine learning research (A report from the NeurIPS 2019 reproducibility program). *J Mach Learn Res.* 2021;22(164):1–20.
- 25 Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, et al. The TRIPOD-AI statement: reporting guidelines for studies developing or validating multivariable prediction models using artificial intelligence. *BMJ.* 2023;381:269. <https://doi.org/10.1136/bmj-2023-078378>
- 26 Sounderajah V, Guni A, Liu X, Collins GS, Karthikesalingam A, Markar SZ, et al. STARD-AI: standards for reporting of diagnostic accuracy studies on artificial intelligence. *Nat Med.* 2022;28:1364–74.
- 27 Hosmer DW, Lemeshow S, Sturdivant RX. *Applied logistic regression.* 3rd ed. Hoboken, NJ: John Wiley & Sons; 2013. <https://doi.org/10.1002/9781118548387>
- 28 Zhang Z, et al. Improving the evaluation of deep learning models in medical imaging: considerations for class imbalance and calibration. *Patterns.* 2021;2(9):100377.

Supplementary



Supplementary Fig S1 | STARD participant flow diagram



Supplementary Fig S2 | PAD prediction calibration

Supplementary Table S1 | TRIPOD-AI and STARD-AI compliance checklist

Item	TRIPOD-AI Requirement	Page	Location	Status
Part A: TRIPOD-AI Checklist (38 items)				
Title and Abstract				
1a	Identify as prediction model study, specify AI methods	1	Title, Abstract	✓ Complete
1b	Abstract: background, objectives, methods, results, conclusions	1	Abstract	✓ Complete
Introduction				
2a	Background and rationale	1	Introduction	✓ Complete
2b	Objectives and research questions	1–2	Introduction	✓ Complete
Methods-Source of Data				
3a	Study design and data sources	3	Clinical Dataset and Study Design	✓ Complete
3b	Eligibility criteria	3	Study Setting and Patient Recruitment	✓ Complete
4a	Outcome definition and assessment	4	Clinical Labels and Prediction Tasks	✓ Complete
4b	Predictors: definition and timing	4–5	Data Preprocessing/Feature Extraction	✓ Complete
5a	Sample size and justification	4	Dataset Characteristics	✓ Complete
5b	Missing data handling	3	Study Setting and Patient Recruitment	✓ Complete
Methods-AI Model				
6a	Model architecture and specifications	5–7	Modified CNN-Based Classification	✓ Complete
6b	Training, validation, test split	4	Data Splitting Strategy (Patient-Level)	✓ Complete
7a	Loss function and optimization	6	Model Training Configuration	✓ Complete
7b	Hyperparameter selection	6	Model Training Configuration	✓ Complete
8a	Handling of imbalanced data	6, 8	Model Training Configuration, Calibration Metrics	✓ Complete
8b	Data augmentation methods	5	Data Augmentation	✓ Complete
9a	Computational environment	6	Model Training Configuration	✓ Complete
9b	Reproducibility: seeds, versions	6, 2	Model Training Configuration, Data availability statement	✓ Complete
Methods-Model Evaluation				
10a	Performance metrics defined	6	Model Training Configuration	✓ Complete
10b	Calibration assessment	7, 8	Statistical Validation Protocol, Table 1	✓ Complete
10c	Clinical utility measures	11, 12	Decision Curve and Cost-Effectiveness Analysis, Figure 8	✓ Complete
11	Subgroup analyses prespecified	11	Subgroup Performance, Table 7	✓ Complete
12	Model interpretability/explainability	11	Explainability and Clinical Validation, Supplementary Figure S3	✓ Complete
Results-Participants				
13a	Flow of participants	3, S1	Study Setting and Patient Recruitment, Supplementary Figure S1	✓ Complete
13b	Participant characteristics	4, 8	Dataset Characteristics, Table 1	✓ Complete
Results-Model Specification				
14a	Final model specification	5–7	Modified CNN-Based Classification	✓ Complete
14b	Model complexity measures	6	Model Architecture	✓ Complete
Results-Model Performance				
15a	Overall performance metrics	7, 8	Statistical Validation Protocol, Table 1	✓ Complete
15b	Performance with confidence intervals	8–11	Tables 1–8	✓ Complete
16a	Calibration results	7, S2	Statistical Validation Protocol, Supplementary Figure S2	✓ Complete
16b	Clinical utility assessment	11, 12	Decision Curve and Cost-Effectiveness Analysis, Figure 8	✓ Complete
17	Subgroup performance	11	Subgroup Performance, Table 7	✓ Complete
18	Model explainability results	11	Explainability and Clinical Validation, Supplementary Figure S3	✓ Complete
Discussion				
19a	Key findings interpretation	12	Principal Findings	✓ Complete
19b	Comparison with literature	13	Comparison with Existing Literature	✓ Complete
20	Strengths and limitations	13	Strengths/Limitations	✓ Complete
21	Clinical implications	12	Clinical Implications	✓ Complete

(Continued)

Supplementary Table S1 Continued				
Item	TRIPOD-AI Requirement	Page	Location	Status
Other Information				
22	Data and code availability	2	Data availability statement	✓ Complete
23	Funding sources	2	Funding	✓ Complete
24	Conflicts of interest	N/A	None declared	✓ Complete
Item	STARD-AI Requirement	Page	Location	Status
Part B: STARD-AI Checklist (30 items)				
Title/Abstract				
1	Identify as diagnostic accuracy study, AI	1	Title, Abstract	✓ Complete
2	Structured abstract	1	Abstract	✓ Complete
Introduction				
3	Study objectives	1	Introduction	✓ Complete
Methods-hn Participants				
4	Study design	3	Clinical Dataset and Study Design	✓ Complete
5	Participant recruitment	3	Study Setting and Patient Recruitment	✓ Complete
6	Data collection	3	Clinical Assessment Protocol	✓ Complete
Methods-Test Methods				
7	Reference standard	3	Clinical Assessment Protocol	✓ Complete
8	Technical specifications	5–7	Modified CNN-Based Classification	✓ Complete
9	Definition of test positivity	4	Clinical Labels and Prediction Tasks	✓ Complete
10	AI model training	5–7	Modified CNN-Based Classification	✓ Complete
11	Blinding procedures	4, 11	Clinical Adjudication, Explainability and Clinical Validation	✓ Complete
Methods-Analysis				
12	Statistical methods	6, 7	Model Training Configuration, Statistical Validation Protocol	✓ Complete
13	Handling indeterminate results	3	Study Setting and Patient Recruitment	✓ Complete
14	Missing data	3	Study Setting and Patient Recruitment	✓ Complete
Results-Participants				
15	Flow diagram	S1	Supplementary Figure S1	✓ Complete
16	Baseline characteristics	4	Dataset Characteristics	✓ Complete
17	Distribution of disease severity	4	Dataset Characteristics	✓ Complete
Results-Test Results				
18	Cross-tabulation of results	9	Figure 4	✓ Complete
19	Adverse events	3	Study Setting and Patient Recruitment	N/A (non-invasive)
20	Handling of incomplete data	S1	Supplementary Figure S1	✓ Complete
Results-Estimates				
21	Diagnostic accuracy estimates	8, 10	Tables 1, 4	✓ Complete
22	Confidence intervals	8–11	All tables	✓ Complete
23	Subgroup analyses	11	Table 7	✓ Complete
24	Threshold analysis	11, 12	Figure 8, Decision Curve and Cost-Effectiveness Analysis	✓ Complete
Discussion				
25	Clinical applicability	12	Clinical Implications	✓ Complete
26	Strengths and limitations	13	Strengths/Limitations	✓ Complete
Other				
27	Funding	2	Funding	✓ Complete
28	AI-specific: architecture	5–7	Modified CNN-Based Classification	✓ Complete
29	AI-specific: interpretability	11	Explainability and Clinical Validation	✓ Complete
30	AI-specific: reproducibility	2	Data availability statement	✓ Complete

TRIPOD-AI Compliance: 38/38 items (100%).

STARD-AI Compliance: 29/29 applicable items (100%) (Item 19 not applicable - non-invasive test).

Supplementary Table S2 Sensitivity analyses with covariate adjustments						
Adjustment Factor(s)	AUC-ROC (%)	95% CI	Sensitivity (%)	Specificity (%)	Δ AUC	P-value*
Base Model (Unadjusted)	96.1	94.2–98.0	94.2	91.5	-	-
Single Covariate Adjustments						
+ DR severity	96.0	94.1–97.9	94.1	91.3	-0.1	0.83
+ Image quality tertile	95.8	93.8–97.8	93.8	91.2	-0.3	0.67
+ HbA1c level	96.2	94.3–98.1	94.3	91.6	+0.1	0.91
+ Diabetes duration	96.1	94.1–98.1	94.2	91.4	0.0	0.98
+ Hypertension status	96.0	94.0–98.0	94.0	91.4	-0.1	0.87
+ BMI category	96.1	94.1–98.1	94.1	91.5	0.0	0.95
Multiple Covariate Adjustments						
+ DR severity + Image quality	95.8	93.7–97.9	93.9	91.2	-0.3	0.71
+ HbA1c + Diabetes duration	96.1	94.1–98.1	94.2	91.5	0.0	0.99
+ All metabolic factors**	96.0	93.9–98.1	94.0	91.4	-0.1	0.89
Full Model (All Adjustments)	96.0	93.8–98.2	94.0	91.4	-0.1	0.76

Analysis: Model performance after adjusting for clinical and image quality covariates

*P-value: Comparison to base model using DeLong test, **Metabolic factors: HbA1c, diabetes duration, BMI, hypertension, hyperlipidemia.

Interpretation

- 1. Performance Stability:** Model performance remains highly stable (AUC 95.8–96.2%) across all covariate adjustments, with maximum change of only 0.3% points.
- 2. Statistical Significance:** No adjustment produces statistically significant change in AUC (all $P > 0.05$), indicating robust prediction independent of clinical confounders.
- 3. Clinical Implications**
 - Model performs consistently across DR severity levels
 - Not confounded by metabolic control (HbA1c) or disease duration
 - Image quality has minimal impact on discrimination
 - Predictions are intrinsic to retinal vascular features, not secondary to other clinical factors

- 4. Validation of Approach:** Stability across adjustments validates that the model has learned genuine PAD-related vascular patterns rather than confounded associations.

Statistical Methods

- **Adjustment Method:** Multivariable logistic regression with AI probability as predictor plus covariates
- **Performance Calculation:** C-statistic from adjusted models
- **Confidence Intervals:** 1,000 bootstrap iterations with covariate resampling
- **Comparison Test:** DeLong's test for correlated ROC curves
- **Multiple Testing:** Bonferroni correction applied ($\alpha = 0.05/10 = 0.005$)