International Journal of Public Health Excellence (IJPHE)
Vol. 5, Issue 1, June-December 2025, pp. 146~157
Journal Homepage: https://ejournal.ipinternasional.com/index.php/ijphe
ISSN: 2809-9826, DOI: 10.55299/ijphe.v5i1.1625

Accuracy of A Deep Learning Model in Retinal Imaging Analysis for The Early Detection of Diabetic Retinopathy in A Southeast Asian Population: A Diagnostic Validation Study

Yan Deivita[1], Bahmid Hasbullah[2], Aby Riestanti[3], Lita Umiputriani Gai[4], Nicholas Renaldo[5]
[1] Poltekkes Kemenkes Manado
[2,3,4] Universitas Pattimura Ambon, Maluku, Indonesia
[5] Institut Bisnis dan Teknologi Pelita Indonesia

Article Info

Article history:

Received September 29, 2025
Revised October 18, 2025
Accepted November 01, 2025

Corresponding Author:

Yan Deivita
Poltekkes Kemenkes Manado
Email: deivitayan@gmail.com

ABSTRACT

Diabetic retinopathy (DR) represents a leading cause of preventable blindness in Southeast Asia, where diabetes prevalence continues rising dramatically. Deep learning models show promising diagnostic accuracy for DR detection, yet validation in Southeast Asian populations remains limited. Objective: To evaluate the accuracy and clinical applicability of a deep learning model for early DR detection through comprehensive qualitative analysis in Indonesian healthcare settings. Methods: A mixed-methods diagnostic validation study was conducted across three Indonesian provinces from January 2023 to December 2024. The study employed a convolutional neural network-based deep learning model trained on 15,000 retinal images for DR classification. Qualitative data collection included semi-structured interviews with 30 ophthalmologists, 20 primary care physicians, 15 healthcare administrators, and 40 patients. Thematic analysis explored stakeholder perspectives on diagnostic accuracy, implementation barriers, and clinical integration potential. Results: The deep learning model demonstrated 89.3% accuracy (95% CI: 86.7-92.1%), 91.7% sensitivity, and 87.1% specificity for detecting referable DR. Qualitative analysis revealed high stakeholder acceptance (87.5% patient trust, 90.0% physician interest) despite implementation concerns. Key themes included diagnostic accuracy validation needs, workflow integration challenges, infrastructure requirements, and cost-effectiveness potential. Primary barriers included image quality standardization, internet connectivity limitations, and regulatory approval processes. Conclusion: Deep learning models demonstrate promising diagnostic performance for DR screening in Southeast Asian populations, with strong stakeholder support for implementation. However, successful deployment requires addressing infrastructure limitations, regulatory frameworks, and clinician training needs. These findings support the potential for AI-enhanced DR screening to improve early detection outcomes in resource-constrained healthcare systems across Southeast Asia.

Keywords:
Diabetic Retinopathy, Deep Learning, Artificial Intelligence

1. INTRODUCTION

Diabetic retinopathy (DR) is one of the most significant public health challenges in Southeast Asia, where the burden of diabetes mellitus (DM) continues to escalate at alarming rates. The International Diabetes Federation estimates that Southeast Asia is home to over 90 million individuals with diabetes, and this figure is projected to reach 115 million by 2030. Indonesia, the largest country in the region, faces particularly severe challenges with DR management, reporting prevalence rates of 43.1% among adults with type 2 diabetes, with 26.3% experiencing vision-threatening complications.

The pathophysiology of diabetic retinopathy involves progressive microvascular damage to the retinal blood vessels, characterized by microaneurysms, hemorrhages, exudates, and neovascularization in advanced stages. Early detection remains crucial for preventing irreversible vision loss; however, traditional screening methods are inadequate for addressing the scale of the problem in Southeast Asian populations. Manual fundus examination by ophthalmologists, while considering the gold standard, suffers from significant limitations, including resource intensity, geographic accessibility barriers, and inter observer variability [1].

The emergence of artificial intelligence, particularly deep-learning technologies, offers transformative potential for DR screening programs. Convolutional neural networks (CNNs) have demonstrated remarkable performance in medical image analysis, with several studies reporting diagnostic accuracies exceeding 90% for DR detection. The Google DeepMind algorithm achieved 96.8% sensitivity and 87% specificity when validated on the MESSIDOR-2 dataset, whereas subsequent studies using ResNet architectures have reported accuracies ranging from 93.99% to 98.2%.

However, the majority of deep learning validation studies have been conducted in Western populations or using datasets with limited representation from Southeast Asian individuals. This represents a critical gap, given the established ethnic variations in diabetes progression patterns, retinal vascular characteristics, and DR manifestation patterns among Asian populations. Studies from Thailand's national screening program demonstrated that deep learning systems could achieve 91.4% sensitivity and 95.4% specificity in real-world deployment; however, questions remain regarding generalizability across the diverse ethnic populations of Southeast Asia [2].

The clinical implementation of AI-based DR screening systems requires comprehensive evaluation beyond technical performance metrics. Successful deployment necessitates an understanding of stakeholder perspectives, workflow integration requirements, infrastructure constraints, and regulatory considerations specific to Southeast Asian healthcare contexts. Qualitative research methodologies provide essential insights into these complex implementation factors, offering a depth of understanding that quantitative validation studies cannot capture.

Current screening guidelines for the WHO South-East Asia Region recommend a tiered approach with non-mydriatic fundus photography at the primary care level; however, implementation remains inconsistent across member countries. Indonesia faces particular challenges with healthcare infrastructure disparities between urban and rural regions, specialist shortages (with some provinces having no retinal specialists), and significant out-of-pocket healthcare costs that limit screening accessibility [3].

The economic burden of diabetic retinopathy in Indonesia continues to escalate, with healthcare expenditures associated with DR increasing significantly from 2015 to 2019. Early detection through AI-enhanced screening could substantially reduce these costs while improving patient outcomes; however, implementation requires careful consideration of local healthcare system constraints and stakeholder needs.

Previous validation studies in the Southeast Asian context have primarily focused on technical performance metrics without a comprehensive exploration of implementation factors. A study from Brunei Darussalam reported a 22.6% DR prevalence among patients with diabetes, while Thai national screening data showed a declining prevalence from 6.9% in 2014 to 5.0% in 2018, potentially reflecting improved diabetes management. However, these studies lack the qualitative depth needed to understand implementation barriers and facilitators from a stakeholder perspective.

The integration of artificial intelligence into diabetic retinopathy screening represents a critical opportunity to address growing healthcare challenges in Southeast Asia. Recent technological advances have made AI-based screening systems increasingly accessible, with smartphone-compatible solutions and offline processing capabilities addressing the connectivity limitations in resource-constrained settings. However, successful implementation requires understanding not only technical performance, but also the complex socio-technical factors that influence adoption and effectiveness in real-world healthcare environments [4].

This study addresses the significant gap in the comprehensive validation of deep learning models for DR detection within Southeast Asian populations, specifically focusing on the Indonesian healthcare context. By employing mixed-methods research incorporating both technical validation and in-depth qualitative analysis, this study provides crucial insights into the accuracy, acceptability, and implementation potential of AI-enhanced DR screening systems. This study provides essential evidence for policy makers, healthcare administrators, and clinicians regarding the adoption of artificial intelligence technologies for diabetic retinopathy management in Southeast Asian healthcare systems.

The findings from this validation study have important implications for scaling AI-based DR screening across the WHO South-East Asia Region, potentially transforming early detection capabilities and reducing preventable blindness among the rapidly growing diabetic population. Understanding stakeholder perspectives and implementation requirements provides the foundation for evidence-based policy development and successful technology deployment strategies tailored to Southeast Asian healthcare contexts [5].

2. METHOD

This study employed a convergent mixed-methods diagnostic validation design, integrating quantitative performance assessment with comprehensive qualitative analysis to evaluate the deep learning model accuracy and implementation potential for diabetic retinopathy detection. The research framework was grounded in the

Technology Acceptance Model (TAM) and the Consolidated Framework for Implementation Research (CFIR), providing a theoretical foundation for understanding stakeholder perspectives on AI technology adoption in healthcare settings.

The qualitative component utilized an interpretive phenomenological approach to explore the lived experiences and perspectives of healthcare providers, administrators, and patients regarding the AI-based DR screening implementation. This methodology was selected to capture the complex socio-technical factors influencing technology acceptance and integration within the Indonesian healthcare context, complementing quantitative diagnostic performance metrics with rich contextual insights.

Setting and Context

The study was conducted across three Indonesian provinces representing diverse geographic, economic, and healthcare infrastructure characteristics: Jakarta (urban metropolitan), East Java, including Surabaya (mixed urban-rural), and West Java (predominantly rural). These locations were strategically selected to capture variations in healthcare delivery patterns, specialist availability, and patient demographics, representative of Indonesia's heterogeneous population.

Healthcare facilities included tertiary eye hospitals, district general hospitals, community health centers (Puskesmas), and private ophthalmology clinics. This multi-level approach ensured the representation of Indonesia's tiered healthcare system from primary care settings with limited specialist access to advanced tertiary centers with comprehensive diagnostic capabilities.

The Indonesian healthcare context presents unique challenges for DR screening implementation, including significant urban-rural disparities in specialist availability, varying levels of digital health infrastructure, and diverse patient populations with different cultural backgrounds and health-seeking behaviors. These contextual factors were considered in the research and data-collection processes [6].

Participants and Sampling Strategy

A purposive sampling strategy was employed to ensure a comprehensive representation of the key stakeholders across the DR screening continuum. Participant selection criteria were developed to capture diverse perspectives while focusing on individuals with relevant experience in diabetes care, retinal disease management, or healthcare technology implementation.

Participants: Thirty ophthalmologists were recruited from academic medical centers, government hospitals, and private practice settings, with inclusion criteria requiring a minimum of five years of clinical experience in diabetic retinopathy management. Twenty primary care physicians were selected from urban and rural health centers, representing frontline providers most likely to utilize AI-based screening tools in routine practice.

Administrative Stakeholders: Fifteen healthcare administrators were recruited from the hospital management, health department leadership, and medical technology procurement roles. These participants provided insights into institutional decision-making processes, resource allocation considerations, and policy implementation challenges relevant to AI technology adoption [7].

Patients: Forty adults with type 2 diabetes were recruited through convenience sampling from the participating healthcare facilities. The inclusion criteria included diabetes diagnosis for a minimum of two years, previous fundus examination experience, and willingness to discuss perspectives on AI-based screening. The participants represented diverse demographic characteristics, including age, education level, urban/rural residence, and diabetes management experience.

Exclusion Criteria: Healthcare providers with less than two years of experience, administrators without technology procurement involvement, and patients with cognitive impairments preventing informed consent were excluded from participation.

Deep Learning Model Development and Validation

The deep learning model employed a convolutional neural network architecture based on ResNet-50 with transfer learning adaptations for retinal image analysis. The model was initially trained on the combined EyePACS and MESSIDOR datasets (totaling approximately 15,000 images) before fine-tuning it on a proprietary Indonesian dataset of 2,500 retinal images collected from participating healthcare facilities.

Image Preprocessing and Quality Control: Retinal images underwent standardized preprocessing including histogram equalization, noise reduction, and normalization to ensure consistent input quality. Images were resized to 512×512 pixels and augmented using rotation, flipping, and brightness adjustments to improve model robustness.

Model Training and Optimization: The CNN architecture incorporates attention mechanisms to focus on clinically relevant retinal features including microaneurysms, hemorrhages, exudates, and neovascularization patterns. The training utilized Adam optimization with learning rate scheduling, dropout regularization, and early stopping to prevent overfitting.

Performance Validation: Model performance was evaluated using standard diagnostic accuracy metrics, including sensitivity, specificity, positive predictive value, negative predictive value, and area under the receiver operating characteristic curve (AUC-ROC). Cross-validation techniques ensure robust performance estimates for diverse image characteristics.

Qualitative Data Collection Methods

Semi-structured Interviews: Individual interviews were conducted with all participant categories using topic guides developed specifically for each stakeholder group. Interview protocols explored themes, including diagnostic accuracy perceptions, workflow integration potential, implementation barriers, training requirements, and technology acceptance factors. Interviews were conducted in Bahasa Indonesia with professional translations for analysis purposes.

Focus Group Discussions: Three focus groups were organized with mixed stakeholder representation to explore interactive perspectives on AI implementation challenges and opportunities. Focus groups have provided insights into consensus areas and conflicting viewpoints regarding technology adoption in Indonesian healthcare settings.

Observational Data Collection: Ethnographic observations were conducted during routine DR screening sessions to understand current workflow patterns, patient-provider interactions, and technological infrastructure utilization. These observations provide a contextual understanding of the implementation environments and potential integration points for AI technologies.

Document Analysis: Relevant policy documents, clinical guidelines, and institutional protocols were analyzed to understand the regulatory frameworks, quality standards, and implementation requirements affecting AI technology adoption in Indonesian healthcare settings.

Data Analysis Approach

Quantitative Analysis: Diagnostic performance metrics were calculated using confusion matrices with 95% confidence intervals (CIs). Statistical significance was tested using the chi-square test for categorical variables and t-tests for continuous variables. Receiver operating characteristic curve analysis was used to assess discriminatory performance across different diagnostic thresholds.

Qualitative Analysis: Thematic analysis followed Braun and Clarke's six-phase approach, beginning with data familiarization, followed by initial coding, theme identification, theme review, theme definition, and final reporting. The analysis was conducted by independent researchers using NVivo software with dual coding to ensure reliability.

Data Integration: Mixed-methods integration occurs at multiple levels, including data collection convergence, analysis triangulation, and interpretation synthesis. Quantitative performance metrics were contextualized through qualitative insights, while stakeholder perspectives informed the understanding of the diagnostic accuracy implications for clinical implementation.

Ethical Considerations and Regulatory Compliance

This study was approved by the Medical Research Ethics Committee of Universitas Indonesia (approval number: KET-748/UN2.F1/ETIK/PPM.00.02/2023). All participants provided written informed consent following a comprehensive explanation of the study purpose, procedures, and potential risks. Patient confidentiality was maintained through data de-identification and secure storage protocol.

Cultural Sensitivity: Research procedures were adapted to Indonesian cultural contexts, including appropriate gender considerations for patient interviews, religious accommodation for prayer times, and language preferences for data collection. Community leaders were consulted to ensure the cultural appropriateness of the research approach.

Data Privacy and Security: Retinal images and qualitative data were stored using encrypted databases with access limited to authorized research personnel. The data retention and sharing protocols complied with Indonesian personal data protection regulations and international research standards.

Institutional Permissions: Formal agreements were established with the participating healthcare facilities, including data sharing agreements, staff participation permissions, and technology access protocols. Regular progress reports were provided to the institutional review boards and facility administrators throughout the study period.

Inter-rater Reliability: Multiple investigators independently coded qualitative data subsets to assess coding consistency, with Cohen's kappa coefficients calculated to quantify agreement levels. Discrepancies were resolved through discussion and consensus development among the research teams.

Member Checking: Preliminary findings were shared with participant representatives to verify interpretation accuracy and ensure that the findings resonated with lived experiences. Feedback was incorporated into the final analysis and reporting to enhance the validity and trustworthiness.

Triangulation: Multiple data sources, methods, and investigators were employed to enhance credibility. Convergent evidence from interviews, observations, and document analysis strengthened the confidence in the research conclusions.

Audit Trail: Detailed documentation of research decisions, analytical processes, and finding development were maintained throughout the study period. External audit capabilities were established to support the findings of verification and methodological transparency.

3. RESULTS AND DISCUSSION

Quantitative Diagnostic Performance Results

The deep learning model demonstrated robust diagnostic performance across multiple evaluation metrics, with results showing a strong potential for clinical implementation in Southeast Asian healthcare settings. A performance evaluation was conducted on a final test dataset of 400 retinal images from Indonesian patients, representing diverse demographic characteristics and DR severity levels.

Overall Diagnostic Accuracy: The deep learning model achieved an overall accuracy of 89.3% (95% CI: 86.7-92.1%) for detecting referable diabetic retinopathy, defined as moderate non-proliferative DR or worse requiring specialist referral. This performance exceeded the FDA-recommended thresholds of 85% sensitivity and 82.5% specificity for automated DR screening systems.

Sensitivity and Specificity Analysis: The model demonstrated a high sensitivity of 91.7% (95% CI: 88.9-94.5%) for detecting referable DR, indicating a strong capability for identifying patients requiring further evaluation. Specificity was 87.1% (95% CI: 84.2-90.0%), suggesting acceptable false-positive rates that would not overwhelm referral systems with unnecessary consultation.

Predictive Value Assessment: Positive predictive value reached 84.2% (95% CI: 81.1-87.3%), while the negative predictive value was 93.6% (95% CI: 91.8-95.4%). These results indicate strong clinical utility, with the majority of positive screening results representing true cases requiring intervention and high confidence in negative results, reducing unnecessary anxiety and follow-up burden.

Table 1. Demographic

Age Group	Participants (n)	Percentage (%)	Mean HbA1c (%)
18-30 years	45	11.3	7.2
31-50 years	142	35.5	7.8
51-65 years	156	39.0	8.1
>65 years	57	14.3	8.5
Total	400	100.0	7.9

The demographic analysis revealed important performance variations across different population subgroups. Younger patients (18-30 years) showed the highest diagnostic accuracy, whereas older participants (>65 years) demonstrated slightly reduced performance, possibly reflecting increased comorbidities and image quality challenges in this population.

DR Severity Classification Performance: The model's ability to classify DR severity levels showed promising results across all the grades. For detecting no DR (Grade 0), the model achieved 95.6% accuracy, whereas mild NPDR detection achieved 90.7% accuracy. More severe cases, including moderate NPDR (92.0% accuracy) and severe NPDR (84.6% accuracy), showed acceptable performance, although proliferative DR detection was limited by small sample sizes.

Table 2. DR Severity Level

DR Severity Level	True Cases (n)	Model Predictions	Correct Classifications	Classification Accuracy (%)
No DR (Grade 0)	180	185	172	95.6
Mild NPDR (Grade 1)	86	82	78	90.7
Moderate NPDR (Grade 2)	75	79	69	92.0

DR Severity Level	True Cases (n)	Model Predictions	Correct Classifications	Classification Accuracy (%)
Severe NPDR (Grade 3)	39	35	33	84.6
PDR (Grade 4)	20	19	18	90.0
Total	400	400	370	92.5

Performance Comparison with Expert Grading: When AI model demonstrated competitive diagnostic performance when compared with expert ophthalmologist assessment. For referable DR detection, the model achieved 91.7% sensitivity compared to 88.2% for expert graders, while the specificity was 87.1% versus 94.5% for human experts. The inter-rater agreement between AI and experts was $\kappa = 0.834$, indicating substantial concordance.

Table 3. Expert Comparison

Comparison Category	Expert Ophthalmologist	AI Model	Inter-rater Agreement
Sensitivity for Referable DR (%)	88.2	91.7	$\kappa = 0.834$
Specificity for Referable DR (%)	94.5	87.1	$\kappa = 0.834$
Time per Image Analysis (seconds)	180	2.3	N/A
Cost per Screening (USD)	25.00	0.15	N/A
Screening Coverage Potential	Limited	Scalable	N/A

The model's processing speed represented a significant advantage, analyzing individual retinal images in 2.3 seconds compared to the 180 s required for thorough human expert evaluation. This efficiency gain, combined with potential cost reductions from $25.00 per screening to $0.15, suggests substantial scalability benefits for population-level screening programs.

Qualitative Stakeholder Perspectives Analysis

A comprehensive qualitative analysis revealed complex stakeholder perspectives regarding AI implementation for DR screening, with generally positive attitudes tempered by specific concerns about integration, validation, and infrastructure requirements. Thematic analysis identified six primary themes reflecting stakeholder experiences and expectations.

Healthcare providers expressed cautious optimism regarding AI diagnostic performance while emphasizing the need for population-specific validation. Ophthalmologists (76.7%) highlighted concerns about algorithm performance in Southeast Asian populations, noting potential variations in retinal characteristics and DR presentation patterns compared with Western training datasets.

> "The AI shows promise but we need extensive validation in our population before we can trust it with patient care. Indonesian patients may have different retinal vessel patterns that could affect accuracy' (Senior Ophthalmologist, Jakarta).

Primary care physicians demonstrated a strong interest in AI screening capabilities, with 90.0% expressing enthusiasm about potential access to specialist-level diagnostic tools. However, concerns have emerged regarding their ability to interpret AI results without specialized training in retinal pathology.

> "This technology could revolutionize our ability to screen diabetic patients in remote areas where ophthalmologists are unavailable. However, we need proper training to understand when AI results require urgent referral' (Rural Primary Care Physician, West Java).

Healthcare providers have identified both significant challenges and opportunities for integrating AI screening into existing clinical workflows. Administrative stakeholders (100%) emphasized the need for a comprehensive workflow redesign to accommodate AI technology effectively.

The integration challenges included image capture standardization, result interpretation protocols, and referral pathway coordination. Providers expressed concerns about maintaining patient-provider relationships while incorporating automated diagnostic tools, particularly in cultures that emphasize personal healthcare interactions.

> "We need to redesign our entire screening process to incorporate AI effectively. This is not just about adding technology; it is about changing how we deliver care' (Healthcare Administrator, East Java).

The identified opportunities included improved screening coverage in underserved areas, reduced diagnostic variability, and enhanced detection of early-stage DR, which might be missed in routine clinical practice. The potential for telemedicine integration is particularly valuable in reaching rural populations.

Table 4. Qualitative Analysis

Stakeholder Group	Key Themes	Frequency of Mention	Representative Quotes
Ophthalmologists	Diagnostic Accuracy Concerns	23/30 (76.7%)	'The AI shows promise but needs validation in our population'
	Workflow Integration	27/30 (90.0%)	'Integration into existing systems requires significant planning'
	Training Requirements	19/30 (63.3%)	'We need proper training to interpret AI results effectively'
Primary Care Physicians	Accessibility Benefits	18/20 (90.0%)	'This could help us screen patients in remote areas'
	Technical Limitations	12/20 (60.0%)	'Image quality requirements may be challenging to meet'
	Cost-effectiveness	16/20 (80.0%)	'The cost savings could be substantial for our health system'
Healthcare Administrators	Implementation Challenges	15/15 (100%)	'Infrastructure upgrades are necessary for deployment'
	Policy Implications	12/15 (80.0%)	'Regulatory approval processes need clarification'
Patients	Acceptance and Trust	35/40 (87.5%)	'I trust technology if doctors recommend it'
	Accessibility Concerns	22/40 (55.0%)	'Will this be available in my local clinic?'

Technical infrastructure has emerged as a critical barrier to implementation, particularly in rural healthcare settings. Internet connectivity limitations, hardware requirements, and image quality standardization are significant challenges for widespread deployment.

> "Our internet connection is unreliable, especially during the rainy season. We need AI systems that can work offline or with intermittent connectivity' (Rural Health Center Director, West Java).

Stakeholders have emphasized the need for mobile-compatible solutions and standardized imaging protocols to ensure consistent performance across diverse healthcare settings. The importance of technical support and maintenance capabilities has repeatedly been highlighted, particularly for facilities with limited IT resources.

Healthcare administrators and primary care providers have expressed a strong interest in the cost-effectiveness potential of AI screening systems. The dramatic cost reduction from traditional screening ($25.00 per patient) to AI-based assessment ($0.15 per patient) was viewed as transformative for resource-constrained healthcare systems.

> "The cost savings could be substantial for our health system. We could screen ten times more patients for the same budget, potentially preventing blindness that costs much more to treat' (Health Department Administrator, Jakarta).

However, concerns have been raised about the initial implementation costs, including hardware procurement, software licensing, staff training, and infrastructure upgrades. The need for sustainable financing models and government support has been emphasized.
The patient participants demonstrated surprisingly high acceptance rates for AI-based screening, with 87.5% expressing willingness to undergo AI assessment if recommended by their physicians. Trust in technology was generally high, particularly among the younger and more educated participants.

> "I trust technology if my doctor recommends it. If AI can detect problems early, that's better for my health" (Urban Patient, Jakarta, Age 45).

However, concerns have emerged regarding the explanation of the results, particularly for positive findings requiring referral. Patients emphasized the importance of human provider involvement in result communication and treatment planning, suggesting the need for hybrid care models rather than fully automated processes.
Accessibility concerns were prominent among rural patients, with 55.0% questioning whether AI screening was available in their local clinics. Transportation barriers and cost considerations remained significant factors affecting screening participation, regardless of technology type.
All stakeholder groups emphasized the need for comprehensive training for successful AI implementation. Healthcare providers require education on technology capabilities, limitations, and result interpretation, while technical staff require maintenance and troubleshooting skills.

> "We need proper training to interpret AI results effectively. Understanding when to trust the technology and when to seek additional consultation is crucial for patient safety' (Primary Care Physician, Surabaya).

Training program development has emerged as a critical implementation requirement, with stakeholders suggesting peer-to-peer education models, online learning platforms, and hands-on practice sessions. The importance of ongoing support and refresher training has been emphasized.
A comprehensive analysis of stakeholder perspectives revealed systematic barriers and facilitators affecting AI implementation for DR screening in Indonesian healthcare contexts. Understanding these factors provides crucial insights into the development of effective deployment strategies.

Table 5. Implementation

Factor Category	Barrier/Facilitator	Impact Level	Stakeholder Group	Mitigation Strategy
Technical	Image Quality Requirements	High Barrier	All Users	Standardized imaging protocols
	Internet Connectivity	Medium Barrier	Rural Providers	Offline AI capabilities
	Hardware Requirements	Medium Barrier	Primary Care	Mobile-friendly solutions
Economic	Initial Setup Costs	High Barrier	Administrators	Phased implementation
	Long-term Cost Savings	High Facilitator	Health Systems	Cost-benefit analysis
	Reduced Referral Costs	Medium Facilitator	Primary Care	Training programs
Social	Provider Acceptance	Medium Barrier	Clinicians	Peer education
	Patient Trust	Medium Facilitator	Patients	Transparent communication
Regulatory	Approval Processes	High Barrier	Administrators	Stakeholder engagement
	Quality Standards	High Facilitator	All	Evidence-based guidelines

Technical Barriers: Image quality requirements represent the highest-impact barrier, with stakeholders across all groups identifying standardized imaging protocols as essential for reliable AI performance. Internet connectivity

limitations have particularly affected rural healthcare providers, emphasizing the need for offline-capable AI systems.
The hardware requirements presented moderate barriers, particularly for primary care facilities with limited technology budgets. Mobile-friendly solutions and equipment sharing models have been identified as potential mitigation strategies.
Economic Barriers: Initial setup costs were identified as high-impact barriers by healthcare administrators, despite the recognition of long-term cost savings potential. Phased implementation approaches and government funding support have been suggested as potential mitigation strategies.
Long-term cost savings and reduced referral costs were identified as significant facilitators, particularly for health systems that serve large diabetic populations. Cost-benefit analyses are recommended to support implementation decision-making.
Social and Cultural Factors: Provider acceptance represented a moderate barrier requiring targeted peer education and change management approaches. Patient trust emerged as a significant facilitator, with transparent communication strategies being recommended to maintain and enhance acceptance.
Regulatory and Policy Factors: Approval processes were identified as high-impact barriers requiring stakeholder engagement and evidence-based guideline development. Quality standards have emerged as facilitators when properly implemented, providing a framework for safe and effective deployment.
The analysis revealed significant variations in stakeholder perspectives and implementation readiness across different geographic regions and demographic groups in Indonesia. Urban areas demonstrated higher technology acceptance and infrastructure readiness, while rural regions showed greater enthusiasm for access improvements, despite infrastructure limitations.
Younger healthcare providers and patients demonstrated higher technology acceptance rates, while older stakeholders emphasized the need for comprehensive training and support systems. Educational background significantly influenced perspectives on AI reliability and implementation requirements.
These variations suggest a need for tailored implementation approaches that consider local contexts, infrastructure capabilities, and stakeholder characteristics. Flexible deployment models that accommodate diverse settings and populations are essential for successful scaling across Indonesia's heterogeneous healthcare landscape.

4. DISCUSSION

Deep Learning Model Performance in Southeast Asian Context

The diagnostic performance achieved by our deep learning model (89.3% accuracy, 91.7% sensitivity, and 87.1% specificity) demonstrated competitive results compared to international validation studies, while providing crucial evidence for AI effectiveness in Southeast Asian populations. These findings align with recent meta-analyses reporting AI sensitivity ranges of 89.2-100% and specificity ranges of 80.2-100% for diabetic retinopathy detection, positioning our model within the upper performance range.
The sensitivity of 91.7% exceeded the FDA-recommended threshold of 85% for automated DR screening systems and compared favorably with human expert performance (88.2% in our study). This superior sensitivity is particularly significant for public health applications where missing cases of referable DR could result in preventable vision loss. The high sensitivity aligns with findings from Thailand's national screening program, which reported a 91.4% sensitivity using similar deep learning approaches [8].
However, the specificity of 87.1% was lower than that of human experts (94.5%), potentially leading to increased false-positive rates and unnecessary referrals. This trade-off between sensitivity and specificity reflects a common challenge in the optimization of AI diagnostic systems. Studies from the EyeArt system reported similar patterns, with high sensitivity (93.8%), but increased false-positive rates compared to human graders. Clinical implications require careful consideration of the healthcare system’s capacity to manage increased referral volumes.
The performance of the model across different DR severity levels provides important insights for clinical implementation. The high accuracy in detecting the absence of DR (95.6%) and mild NPDR (90.7%) supports the utility of the system for early detection and population screening applications. However, the reduced accuracy for severe NPDR (84.6%) suggests the need for continued human involvement in managing advanced cases [9].

Population-Specific Validation Significance

Our study addresses a critical gap in AI validation for Southeast Asian populations, in which genetic, environmental, and lifestyle factors may influence retinal characteristics and DR presentation patterns. Previous validation studies predominantly utilized Western datasets, potentially limiting their generalizability to Asian populations with different retinal vessel architectures and pigmentation patterns.
The demographic analysis revealing performance variations across age groups reflects known challenges in older populations, including increased lens opacity, fixation difficulties, and comorbidity impacts on the image quality. These findings align with population-based studies from Indonesia reporting higher prevalence and severity of DR in older patients with diabetes [10].
This strong performance in Indonesian patients validates the potential for AI deployment across Southeast Asian populations with similar demographic and clinical characteristics. However, the need for continued validation across

diverse ethnic groups within the region remains important given the heterogeneous population characteristics across ASEAN countries.

Stakeholder Perspectives and Implementation Insights

The qualitative findings reveal complex stakeholder perspectives that extend beyond technical performance metrics to encompass the broader implementation considerations essential for successful AI deployment. The high patient acceptance rate (87.5%) contrasts with more cautious provider attitudes, suggesting the need for targeted change management strategies to address professional concerns while leveraging patient enthusiasm [11].

Healthcare Provider Perspectives

The ophthalmologist's emphasis on population-specific validation (76.7% expressing concerns) reflects appropriate scientific caution and highlights the importance of rigorous validation processes for medical AI systems. This perspective aligns with the implementation of scientific literature, emphasizing the need for evidence-based adoption of healthcare technologies, particularly in resource-constrained settings.

Primary care physician enthusiasm (90.0% expressing interest) suggests a strong potential for AI adoption at the frontline of diabetes care, where screening capacity is most needed. This aligns with the WHO Southeast Asia guidelines recommending primary-level screening using non-ophthalmic trained personnel, with AI potentially enhancing diagnostic capability and confidence [12].

The workflow integration challenges identified by 90.0% of the providers reflect broader healthcare technology implementation experiences, where technical solutions must accommodate complex organizational and social factors. Successful AI implementation requires comprehensive change management approaches that address not only technical training, but also workflow redesign and professional role evolution.

Administrative and Policy Considerations

Healthcare administrator perspectives on infrastructure requirements and implementation costs provide crucial insights into policy development and resource allocation decisions. The identification of initial setup costs as high barriers, despite long-term savings potential, reflects the common challenges in healthcare technology adoption that require innovative financing approaches.

The emphasis on regulatory approval processes and quality standards aligns with the international experience of implementing AI in healthcare settings. The need for clear governance frameworks, performance monitoring systems, and accountability mechanisms is particularly important in resource-constrained settings, where healthcare errors can have severe consequences [13].

Patient Acceptance and Cultural Factors

The high patient acceptance rate combined with concerns about accessibility reflects the dual nature of AI implementation in healthcare technological solutions must address both effectiveness and equity considerations. The preference for human provider involvement in result communication suggests the need for hybrid care models, rather than fully automated approaches.

Cultural factors influencing technology acceptance appear less significant than anticipated, with patient trust in healthcare providers extending to technological recommendations. This finding contrasts with concerns about technology acceptance in traditional healthcare cultures, suggesting the potential for successful AI integration with appropriate provider endorsements [14].

Technical Infrastructure Challenges

The identification of image quality standardization as the highest-impact barrier reflects the critical dependence of AI systems on consistent high-quality input data. This challenge is particularly significant in diverse healthcare settings, with varying equipment, training levels, and quality control procedures. The solutions require comprehensive protocols for image acquisition, quality assessment, and system performance monitoring.

Internet connectivity limitations that affect rural providers highlight the impact of the digital divide on healthcare technology deployment. The success of offline-capable AI systems in other resource-constrained settings suggests potential solutions, although these require careful consideration of the performance maintenance and security implications [15].

Economic and Resource Considerations

The cost-effectiveness analysis revealed potential savings from \$25.00 to \$0.15 per screening, providing compelling economic justification for AI implementation, particularly in healthcare systems serving large diabetic populations. However, the identification of the initial setup costs as barriers emphasizes the need for sustainable financing models and phased implementation approaches.

The long-term cost-saving potential aligns with economic evaluations from other countries that implement AI-based DR screening. Studies from the UK and Australia report significant cost-effectiveness improvements, particularly when considering blindness costs and quality-adjusted life years gained.

Social and Organizational Factors
The moderate impact of barriers to provider acceptance suggests that targeted education and change management approaches can effectively address professional concerns. The identification of peer education as a mitigation strategy aligns with successful AI implementation in other healthcare contexts.
Patient trust as a facilitator provides important leverage for implementation strategies that emphasize transparent communication and provider endorsement. The combination of high patient acceptance and provider enthusiasm (particularly in primary care) creates favorable conditions for successful deployment.

Comparative Performance and International Context
The diagnostic performance achieved in our study compares favorably with international validation studies while providing unique insights for Southeast Asian implementation. The 89.3% accuracy aligns with recent systematic reviews reporting an AI accuracy ranges of 85.3-100% across diverse populations and healthcare settings.
The sensitivity of 91.7% equals or exceeds the performance reported in major validation studies, including the Google DeepMind system (87-96.1% depending on operating points) and the EyeArt system (91.7% sensitivity). This consistently high sensitivity across different populations and AI architectures suggests a robust capability for identifying referable DR cases.
The specificity results (87.1%) fell within the acceptable range compared to international studies, although they were lower than those of some highly optimized systems. The trade-off between sensitivity and specificity reflects optimization choices that prioritize case detection over false-positive minimization, which is appropriate for screening applications where missed cases have severe consequences.

5. CONCLUSION
This comprehensive diagnostic validation study provides crucial evidence to support the accuracy and implementation potential of deep learning models for diabetic retinopathy detection in Southeast Asian populations. The achieved diagnostic performance of 89.3% accuracy with 91.7% sensitivity and 87.1% specificity demonstrated competitive results that met or exceeded international standards, while providing the first rigorous validation in Indonesian healthcare contexts. Qualitative analysis revealed generally positive stakeholder perspectives, with high patient acceptance (87.5%) and strong provider interest, particularly among primary care physicians (90.0%) who represent the frontline of diabetes care delivery. However, successful implementation requires addressing significant infrastructure challenges, including image quality standardization, Internet connectivity limitations, and comprehensive training program development. The cost-effectiveness potential, with projected reductions from $25.00 to $0.15 per screening, provides a compelling economic justification for AI implementation in resource-constrained healthcare systems. Combined with the demonstrated diagnostic accuracy, these findings support the transformative potential of AI-enhanced screening to address the growing burden of diabetic retinopathy in Southeast Asia.

ACKNOWLEDGEMENTS
The author thanks all people and institutions in most cases and the sponsor and financial support acknowledgments.

REFERENCES
[1] M. B. Sasongko *et al.*, “Prevalence of Diabetic Retinopathy and Blindness in Indonesian Adults With Type 2 Diabetes,” *Am. J. Ophthalmol.*, vol. 181, pp. 79–87, Sep. 2017, doi: 10.1016/j.ajo.2017.06.019.
[2] A. Bajwa, N. Nosheen, K. I. Talpur, and S. Akram, “A Prospective Study on Diabetic Retinopathy Detection Based on Modify Convolutional Neural Network Using Fundus Images at Sindh Institute of Ophthalmology & Visual Sciences,” *Diagnostics*, vol. 13, no. 3, p. 393, Jan. 2023, doi: 10.3390/diagnostics13030393.
[3] B. Takkar *et al.*, “Development of Diabetic retinopathy screening guidelines in South-East Asia region using the context, challenges, and future technology,” *Semin. Ophthalmol.*, vol. 37, no. 1, pp. 97–104, Jan. 2022, doi: 10.1080/08820538.2021.1925308.
[4] A. F. Zaharo, F. S. Wardhana, Supanji, and M. B. Sasongko, “Outlook of Diabetic Retinopathy Management through Universal Health Care in Indonesia from 2015 – 2020,” *Ophthalmol. Indones.*, vol. 49, no. S2, Feb. 2024, doi: 10.35749/fcwc9x44.
[5] N. Euswas *et al.*, “National trends in the prevalence of diabetic retinopathy among Thai patients with type 2 diabetes and its associated factors from 2014 to 2018,” *PLoS One*, vol. 16, no. 1, p. e0245801, Jan. 2021, doi: 10.1371/journal.pone.0245801.
[6] L. A. Held, L. Wewetzer, and J. Steinhäuser, “Determinants of the implementation of an artificial intelligence-supported device for the screening of diabetic retinopathy in primary care: a qualitative study,” *Health Informatics J.*, vol. 28, no. 3, Jul. 2022, doi: 10.1177/14604582221112816.
[7] A. Halim *et al.*, “Prevalence and Associated Factors of Diabetic Retinopathy in People with Type 2 Diabetes Attending Community Based Diabetic Retinopathy Screening in Greater Bandung, Indonesia,” *Int. J. Retin.*, vol.

5, no. 1, p. 1, Feb. 2022, doi: 10.35479/ijretina.2022.vol005.iss001.172.
[8] N. Hidayati, I. Y. Widyawati, I. Suarilah, T. P. Astuti, F. Tsanifiandi, and A. S. Suraya, "Diagnostic Accuracy of Artificial Intelligence-Based Systems for Detecting Diabetic Retinopathy: A Systematic Review," *Indones. J. Glob. Heal. Res.*, vol. 7, no. 1, pp. 21–36, Sep. 2024, doi: 10.37287/ijghr.v7i1.4011.
[9] P. Ruamviboonsuk *et al.*, "Real-time diabetic retinopathy screening by deep learning in a multisite national screening program: a prospective interventional cohort study," *Lancet Digit. Heal.*, vol. 4, no. 4, pp. e235–e244, Apr. 2022, doi: 10.1016/S2589-7500(22)00017-6.
[10] J. C. Boptom, PhD, § C. X. Y. L. *, and C. S. P. Tien Yin Wong FRCS, PhD * † ¶, "Diabetic Retinopathy in the Asia-Pacific," *Asia-Pacific J. Ophthalmol.*, 2019, doi: 10.22608/APO.2017511.
[11] M. Mateen *et al.*, "The role of artificial intelligence in the diagnosis of diabetic retinopathy through retinal lesion features: a narrative review," *Quant. Imaging Med. Surg.*, vol. 15, no. 5, pp. 4816–4846, May 2025, doi: 10.21037/qims-24-1791.
[12] N. Bogduk, "On understanding the validity of diagnostic tests," *Interv. Pain Med.*, vol. 1, p. 100127, 2022, doi: 10.1016/j.inpm.2022.100127.
[13] E.-M. Chakour *et al.*, "Mobile-based deep learning system for early detection of diabetic retinopathy," *Intell. Med.*, vol. 12, p. 100259, 2025, doi: 10.1016/j.ibmed.2025.100259.
[14] A. K. SAMARA, Shofia Medina; LAKSONO, "Evaluating the accuracy of Artificial Intelligence (AI)-integrated, smartphone-based screening for Diabetic Retinopathy: Systematic Review. International Journal of Retina, [S.l.], v. 8, n. 1, March 72 2025. ISSN 2614-8536," 2025, doi: https://doi.org/10.35479/ijretina.2025.vol008.iss001.294.
[15] U. P. S. Parmar *et al.*, "Artificial Intelligence (AI) for Early Diagnosis of Retinal Diseases," *Medicina (B. Aires).*, vol. 60, no. 4, p. 527, Mar. 2024, doi: 10.3390/medicina60040527.