

Development and Validation of a Novel Deep Learning-Based Model for Detection of Diabetic Kidney Disease from Retinal Imaging Using a Weighted Loss Method

Seskoati Prayitnangsih¹, Othe Ahmad Syarifuddin¹, Fauzan Kurniawan Dhani², Hera Dwi Novita¹, Nur Samsu³, Muhammad Bayu Sasongko⁴, Candra Dewi⁵, Novanto Yudistira⁵

¹Department of Ophthalmology, Faculty of Medicine, Universitas Brawijaya, Dr. Saiful Anwar General Hospital, Malang, East Java, Indonesia;

²Department of Urology, Faculty of Medicine, Universitas Brawijaya, Dr. Saiful Anwar General Hospital, Malang, East Java, Indonesia; ³Department of Internal Medicine, Division of Nephrology, Faculty of Medicine, Universitas Brawijaya, Dr. Saiful Anwar General Hospital, Malang, East Java, Indonesia;

⁴Department of Ophthalmology, Faculty of Medicine, Universitas Gadjah Mada, Yogyakarta, Special Region of Yogyakarta, Indonesia; ⁵Faculty of Computer Science, Universitas Brawijaya, Malang, East Java, Indonesia

Correspondence: Seskoati Prayitnangsih, Department of Ophthalmology, Faculty of Medicine, Universitas Brawijaya, Jalan Veteran, Malang, 65145, Indonesia, Tel +62 811-3618-881, Email seskoatip@ub.ac.id

Background: Retinal photographs offer great opportunity to early detect systemic disorders related to diabetes, including Chronic Kidney Disease (CKD).

Purpose: To develop and validate a novel deep learning model to detect CKD among diabetic patients.

Patients and Methods: We developed an EfficientNet-B2 Deep Learning (DL) model utilizing a weighted cross-entropy loss function to address class imbalance and distinguish retinal images among healthy controls, patients with isolated type 2 diabetes mellitus (T2DM), and patients with CKD stage 3 due to T2DM. The dataset was partitioned using a strict 80/20 patient-level split to evaluate bilateral eyes while strictly preventing data leakage. Model performance was evaluated using sensitivity, specificity, and area under the curve (AUC), alongside Grad-CAM visualizations for clinical interpretability.

Results: The study included 225 participants. Among the evaluated DL architectures, the EfficientNet-B2 model demonstrated the best performance, achieving an overall AUC of 0.96. The model exhibited a sensitivity of 82%, specificity of 94%, precision of 81%, and an F1-score of 0.80. Class-specific AUCs were 0.99 for healthy controls, 0.90 for T2DM without CKD, and 0.90 for T2DM with CKD stage 3. Grad-CAM heatmaps indicated that the model primarily focused on the peripapillary and macular regions to make predictions.

Conclusion: This study presents a three-class fundus-based DL model, trained with a weighted-loss strategy, to differentiate controls, isolated T2DM, and T2DM with CKD stage 3. Further external and prospective validation is needed before it can be considered for screening/triage use.

Keywords: chronic kidney disease, type 2 diabetes mellitus, deep learning, artificial intelligence, retinal fundus photography

Introduction

Diabetes mellitus (DM) is a leading cause of chronic kidney disease (CKD), with approximately 30–40% of diabetic patients developing diabetic kidney disease (DKD) that can progress to end-stage renal disease.¹ Because CKD often advances silently until late stages, early detection is essential to reduce morbidity and healthcare burden. Current DKD screening primarily relies on albuminuria and estimated glomerular filtration rate (eGFR); however, a clinically important subset of patients develops DKD without albuminuria, underscoring limitations of existing biomarkers and motivating alternative, complementary approaches for earlier identification.^{2,3}

The kidney and retina share closely related microvascular architecture and overlapping pathogenic pathways—including inflammation, oxidative stress, endothelial dysfunction, and microangiopathy—through which chronic hyperglycemia drives systemic microvascular damage.^{1,4,5} Retinal microvascular signs such as diabetic retinopathy, arteriolar narrowing, and venular dilation have been associated with CKD, supporting retinal imaging as a non-invasive window into early renal dysfunction.² Importantly, digital fundus photography is already widely implemented in primary care for diabetic retinopathy screening, making it an accessible platform for scalable DKD risk stratification and community-level screening, with ophthalmology positioned to contribute to broader systemic disease detection.⁶

Recent advances in artificial intelligence (AI) and deep learning algorithms (DLAs) have enabled automated retinal image analysis for predicting renal impairment, with prior studies reporting strong discrimination (AUC up to ~0.91) across diverse populations.^{2,3,7–9} However, key gaps remain: medical imaging datasets are frequently class-imbalanced, potentially biasing learning and reducing performance in minority outcomes, necessitating mitigation strategies such as weighted loss functions.¹⁰ Additionally, while architectures such as Vision Transformer (ViT) and ConvNeXt have set new benchmarks in computer vision, their comparative efficacy for retinal-based CKD/DKD detection remains under-explored.^{11,12} Finally, limited interpretability can impede clinical adoption, highlighting the need for explainable AI (XAI) methods such as Grad-CAM to verify that model attention aligns with physiologically plausible retinal features.^{13,14} Unlike multimodal approaches that incorporate clinical variables, this study evaluates a strictly unimodal fundus-image model defined by diabetes status and eGFR-based renal impairment, while acknowledging potential confounding from age, hypertension, and vascular disease burden. Building on prior work, we develop and validate a novel DLAs-based model for DKD/CKD detection from retinal imaging using a weighted-loss method, compare state-of-the-art architectures, and integrate XAI to enhance transparency and clinical relevance.

Materials and Methods

This cross-sectional study was conducted at Panti Nirmala and Dr. Saiful Anwar General Hospital, Malang, Indonesia, between November 2024 to January 2025. Ethical approval was granted by the Institutional Ethics Committee of Dr. Saiful Anwar General Hospital, Indonesia (No. 400/252/K.3/102.7/2024). This study was conducted in compliance with the principles and practice of the Declaration of Helsinki.

The participants included healthy individuals from the Medical Check-Up Clinic, who had no known systemic or ocular disease, as well as patients with type 2 diabetes mellitus (T2DM) and CKD stage 3 admitted to the Department of Internal Medicine. All participants were examined using the same fundus imaging protocol. Eligible subjects provided written informed consent prior to enrollment.

Inclusion and Exclusion Criteria

Inclusion criteria comprised individuals aged ≥ 40 years with normal findings confirmed by laboratory tests and fundus examination, patients with T2DM diagnosed for at least one year, and patients with CKD stage 3 who were able to undergo retinal photography. Exclusion criteria included significant ocular pathologies (eg, high myopia, glaucoma, diabetic retinopathy, retinal hemorrhages), visual acuity $\leq 20/400$, or any significant media opacities that precluded high-quality imaging. Patients with visible signs of Diabetic Retinopathy (DR) were explicitly excluded to ensure the model focuses purely on early microvascular alterations induced directly by CKD, independent of overt DR lesions.

Chronic Kidney Disease and Type 2 Diabetes Mellitus

CKD stage 3 was defined as an estimated glomerular filtration rate (eGFR) of 30–59 mL/min/1.73 m² sustained for at least three months. To confirm the chronicity of the condition, a decline in kidney function had to persist for at least 3 months, evidenced by a minimum of two eGFR measurements obtained at least 3 months apart during routine monitoring.¹⁵ T2DM was diagnosed according to standard criteria: glycated hemoglobin (HbA1c) $\geq 6.5\%$, fasting plasma glucose (FPG) ≥ 126 mg/dL, 2-hour plasma glucose (2-h PG) ≥ 200 mg/dL during a 75-g oral glucose tolerance test (OGTT), or random plasma glucose ≥ 200 mg/dL with classic symptoms of hyperglycemia or hyperglycemic crisis.¹⁶ Diagnoses of CKD and T2DM were established by experienced nephrologists, while fundus imaging was performed by

ophthalmologists. Relevant laboratory data were collected during hospitalization or outpatient visits concurrent with fundus examination.

Demographic Data and Retinal Images Processing

Demographic and comorbidity data for all participants were obtained from medical records, along with ophthalmic parameters including best corrected visual acuity (BCVA) and converted to the logarithm of the minimum angle of resolution (LogMAR) for statistical analysis, intraocular pressure (IOP), axial length, pupil diameter, anterior and posterior segment. For each subject in the primary test dataset, bilateral non-mydratic digital fundus photographs were captured using a Topcon TRC-NW400 camera by a single experienced ophthalmologist. High-resolution images centered on the optic disc were analyzed to quantify retinal vascular parameters. To prevent information leakage from correlated bilateral fundus photographs, all dataset was split at the patient level into a training set (80%) and a test set (20%). Fundus photographs were reviewed by an experienced ophthalmologist as part of image quality control and dataset curation, and images showing retinal hemorrhages or exudates (and other ocular pathologies per the study exclusion criteria) were excluded prior to dataset splitting and model training. Each participant contributed up to two images (right and left eye), but both eyes from the same participant were assigned exclusively to either the training or the test set and were never split across sets. Stratified sampling was used to maintain proportional class representation across splits.

AI Architecture Development

We formulated a three-class (multiclass) classification task to differentiate controls, T2DM, and T2DM with CKD stage 3 from retinal fundus images. Several state-of-the-art deep learning (DL) backbones were evaluated under identical experimental settings, including ResNet (eg, ResNet-34, ResNet-50), EfficientNet-B2, Vision Transformer (ViT) Tiny and Small, and ConvNeXt Tiny and Small. A classification head consisting of a fully connected layer with three output units ($n \times 3$) was appended to each backbone generate multiclass predictions. EfficientNet-B2 was selected as the primary architecture due to its compound scaling strategy, which balances network depth, width, and resolution to achieve strong accuracy and is well-suited for medical imaging tasks with limited datasets and computational constraints.

To address class imbalance in dataset, a weighted cross-entropy loss function was employed during training. The class weight for each category was calculated as:

$$w_c = \frac{N}{n_c \times C}$$

Where N is the total number of samples, n_c is the number of samples in class c , and C is the total number of classes. This weighting scheme penalizes misclassification of under-represented classes and reduces bias toward the majority class.

Model optimization was performed using the AdamW optimizer (batch size of 16) and initial learning rate of 5×10^{-5} , decayed using a cosine annealing schedule over 30 epochs. Prior to model input, original fundus image (3152×3000 pixels) were resized to 640×640 pixels using bicubic interpolation to maintain high-frequency clinical details. Subsequently, the images underwent Z-score normalization using the standard ImageNet mean ($\mu = [0.485, 0.456, 0.406]$) and standard deviation ($\sigma = [0.229, 0.224, 0.225]$) constants. This scaling procedure ensures that the input feature distribution is aligned with the pre-trained weights of the DL backbones, facilitating faster convergence and improved feature extraction.

To mitigate overfitting, on-the-fly data augmentation (random zoom and rotation with a probability of 0.8) was applied only during training and was not used for validation/testing or figure generation. No content-altering image manipulation was performed. For interpretability, Grad-CAM was computed from the trained network using the final convolutional feature layer (last backbone block) as the target layer; saliency maps were generated for the predicted class, upsampled to the input resolution, and visualized as semi-transparent heatmap overlays on the corresponding fundus photographs without modifying the underlying image content. All experiments were conducted on a workstation equipped with four NVIDIA GeForce RTX 3070 Ti (8 GB) GPUs.

Statistical Analysis, Performance Matrix and Attribution Maps

Participant characteristics were summarized as number (percentage) for categorical variables, mean (standard deviation, SD) for normally distributed data, and median [interquartile range, IQR] for non-parametric data. Comparisons between the model and validation datasets were performed using one-way ANOVA or the Mann–Whitney *U*-test, as appropriate. A *P* value of <0.05 was considered statistically significant. Statistical analyses and model evaluations were performed using SPSS (version 26).

The model was trained and inferred at the image level (each fundus photograph as an input), producing a 3-class probability vector per eye. Because CKD status is defined at the participant level, the primary performance evaluation was conducted at the patient level. For participants with bilateral images, the eye-level probability vectors were aggregated into a single patient-level probability vector by taking the mean across both eyes (participants with a single image used that eye's probability), and the patient-level predicted class was defined as the argmax of the aggregated probabilities. AUROC and all classification metrics were computed on these patient-level predictions; eye-level results are reported as secondary analyses to describe misclassification patterns. For multiclass AUROC, a one-vs-rest strategy was used and reported as overall and class-specific AUROCs. Model performance was assessed using standard classification metrics, including sensitivity, specificity, precision, F1-score, and the area under the receiver operating characteristic curve (AUROC). Ninety-five percent confidence intervals (95% CI) were calculated to assess the precision of these estimates. Given the black-box nature of DL models, explainability was prioritized to ensure clinical trustworthiness and facilitate adoption by ophthalmologists.

Gradient-weighted Class Activation Mapping (Grad-CAM) was employed to visualize the model's decision focus, highlighting clinically relevant retinal regions including the optic disc and macula. Grad-CAM leverages the gradients flowing into the final convolutional layer to produce class-discriminative localization maps, enabling visualization of which retinal regions most strongly influence the model's predictions. To further enhance interpretability, Guided Grad-CAM was implemented by combining Grad-CAM with pixel-space gradient information, producing high-resolution, class-specific attribution maps that provide deeper insight into model decision processes. The generated attribution maps were systematically analyzed and compared with established retinal biomarkers known to be associated with CKD and T2DM, including peripapillary vessel morphology (vessel caliber, tortuosity, and branching patterns), and retinal vessel parameters (arteriovenous nicking and vessel narrowing). Each attribution map was visually inspected by experienced ophthalmologists to verify that the highlighted regions corresponded to anatomically and pathophysiologically relevant areas, ensuring that the model bases its predictions on clinically meaningful features rather than spurious correlations. This interpretability framework not only strengthens clinical confidence in AI predictions but also provides a foundation for understanding how retinal microvascular changes correlate with systemic diseases such as CKD and T2DM.¹³

Result

Of the 267 patients initially screened, 225 were included in the final analysis. Exclusions were due to obstructed or unclear retinal images ($n = 20$), retinal hemorrhages and/or exudates ($n = 10$), glaucoma ($n = 7$), and prior retinal surgery or laser treatment ($n = 5$), as detailed in the study flowchart (Figure 1). Among the included participants, 132 were classified as normal controls, 38 as T2DM, and 55 as T2DM with CKD stage 3. Demographic and clinical characteristics of the study population are summarized in Table 1. Significant differences were observed in glucose status and kidney function across the three groups, while demographic variables and other ocular characteristics did not differ significantly.

Data Splitting Strategies and Weighted Loss

Before comparing model architectures, an ablation study using EfficientNet B2 was conducted to determine the optimal evaluation framework (Table 2). The results demonstrate that a strict Patient-Level Split combined with a weighted loss function yielded the best performance, achieving an AUC of 0.96, sensitivity of 82%, specificity of 94%, precision of 81%, and an F1-score of 0.80. Removing the weighted loss in this setup slightly reduced the sensitivity to 80% and the F1-score to 0.79. Conversely, both Image-Level Split configurations resulted in lower overall metrics, with F1-scores dropping to 0.78 (unweighted) and 0.75 (weighted). Consequently, the Patient-Level Split with weighted loss was selected as the primary, most rigorous strategy for all subsequent analyses to effectively handle class imbalance and prevent data leakage.

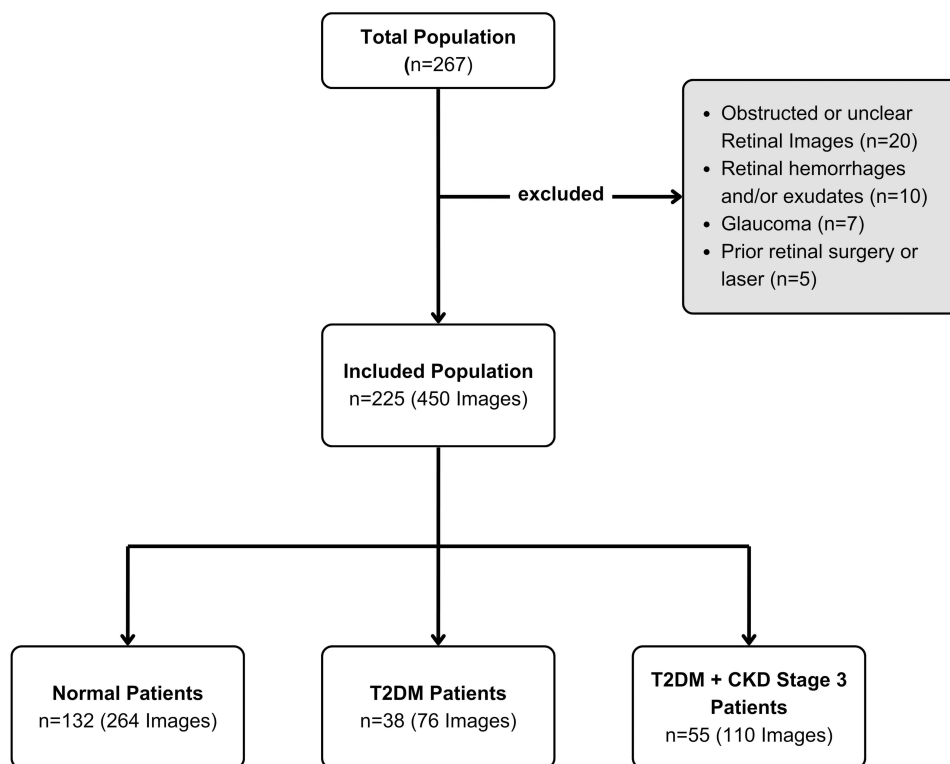


Figure 1 Research Study Participant Flowchart.

The Performance Metrics

The performance of the three-class (multiclass) models in differentiating controls, T2DM, and T2DM with CKD stage 3 is summarized in Table 3. Among the evaluated architectures, EfficientNet B2 demonstrated the best overall performance, achieving an AUC of 0.96 with a sensitivity of 82%, specificity of 94%, precision of 81%, and an F1-score of 0.80. ResNet 18 showed the second-best performance, achieving a sensitivity of 78%, specificity of 93%, precision of 76%, F1-score of 0.75, and an AUC of 0.94. This was followed by ResNet 34, which yielded slightly lower metrics, including a sensitivity of 75%, specificity of 92%, precision of 72%, F1-score of 0.72, and an AUC of 0.92. Meanwhile, Convnext demonstrated the lowest overall performance with a sensitivity of 68%, precision of 66%, and F1-score of 0.66, although it maintained a relatively high specificity of 92% and an AUC of 0.94. Taken together, these results suggest that EfficientNet B2 provides the most favorable balance between sensitivity, specificity, and overall discriminative power across all groups.

Table 1 Study Demographic and Patient Characteristics

Characteristics	Control (n=132)	T2DM (n=38)	T2DM + CKD Stage 3 (n=55)	p-value ^a
Male n (%)	56 (42.42%)	15 (39.47%)	36 (65.45%)	0.14
Female n (%)	76 (57.58%)	23 (60.53%)	19 (34.55%)	0.09
Age (years) mean ± SD	58.27 ± 5.98	60.26 ± 6.25	61.75 ± 8.9	0.37
Body Weight (kg) mean ± SD	68.31 ± 18.25	70.43 ± 16.37	67.73 ± 13.33	0.31
Height (cm) mean ± SD	159.87 ± 8.56	158.92 ± 7.76	160.75 ± 8.6	0.7
BMI. (kg/m ²)	26.12 ± 6.76	27.80 ± 5.88	26.66 ± 4.47	0.82
Waist Circumference (cm) mean ± SD	95.17 ± 11.67	97.84 ± 14.62	97.06 ± 12.1	0.55

(Continued)

Table 1 (Continued).

Characteristics	Control (n=132)	T2DM (n=38)	T2DM + CKD Stage 3 (n=55)	p-value ^a
Blood Pressure				
Systole. mean ± SD	129.55 ± 11.32	133.84 ± 21.43	141.23 ± 23.61	0.34
Diastole. mean ± SD	75.42 ± 11.3	79.42 ± 11.3	81.67 ± 12.43	0.65
Glucose Status				
FPG (mg/dL) mean ± SD	98.65 ± 12.41	120.85 ± 42.41	122.70 ± 28.76	0.00*
OGTT mg/dL mean ± SD	156.15 ± 26.98	192.23 ± 68.54	188.20 ± 67.87	0.02*
HbA1c. (%) mean ± SD	5.67 ± 0.43	7.51 ± 1.39	7.61 ± 1.47	0.00*
Kidney Function				
eGFR mean ± SD	88.78 ± 7.36	86.25 ± 12.89	41.35 ± 9.2	0.00*
uACR	90.52 ± 13.2	90.52 ± 13.2	462.90 ± 428.53	0.00*
Visual Acuity				
RE; LogMAR mean ± SD (Snellen Equivalent)	0.45 ± 0.22 (approx. 20/56.3)	0.55 ± 0.27 (approx. 20/85.42)	0.55 ± 0.28 (approx. 20/86.82)	0.01
LE (logMAR) mean ± SD (Snellen Equivalent)	0.41 ± 0.22 (approx. 20/51.4)	0.51 ± 0.27 (approx. 20/79.16)	0.57 ± 0.30 (approx. 20/93.47)	0.00
Pupil Diameter				
RE (mm)	4.50 ± 1.1	4.60 ± 1.2	4.06 ± 0.9	0.87
LE (mm)	4.72 ± 0.9	4.78 ± 1.1	4.07 ± 1.2	0.97
Intra Ocular Pressure				
RE (mmHg)	16.35 ± 4.8	16.59 ± 5.1	16.06 ± 4.3	0.78
LE (mmHg)	16.33 ± 3.9	16.32 ± 3.4	16.08 ± 4.5	0.89
Axial Length				
RE	22.84 ± 0.77	22.94 ± 0.67	23.14 ± 0.66	0.9
LE	22.75 ± 0.71	22.95 ± 0.73	22.95 ± 0.65	0.84

Notes: ^aOne way anova test, Mann Whitney U-Test. *Significant Result.

Abbreviations: BMI Body Mass Index, CKD Chronic Kidney Disease, FPG Fasting Plasma Glucose, OGTT Oral Glucose Tolerance Test, eGFR Estimated Glomerular Filtration Rate, uACR Urine AlbuminCreatinine Ratio, RE Right Eye, LE Left Eye, T2DM Type 2 Diabetes Mellitus.

Table 2 Ablation Study: Impact of Data Splitting Strategies and Weighted Loss

Model	Sensitivity (%)	Specificity (%)	AUC	Precision (%)	F1-Score
Patient-level Split - (Weighted)	82	94	0.96	81	0.80
Patient-level Split - (Unweighted)	80	94	0.96	78	0.79
Image-level Split - (Weighted)	79	93	0.95	79	0.78
Image-level Split - (Unweighted)	77	93	0.93	76	0.75

Abbreviation: AUC, Area Under Curve.

Table 3 Performance Metrics of the Three-Class (Multiclass) Models for Differentiating Control Patients, Patients with T2DM, and Patients with CKD Stage 3

Model	Sensitivity (%)	Specificity (%)	AUC	Precision (%)	F1-Score
EfficientNet B2	82	94	0.96	81	0.80
ResNet 18	78	93	0.94	76	0.75
ResNet 34	75	92	0.92	72	0.72
Convnext	68	92	0.94	66	0.66

Abbreviation: AUC, Area Under Curve.

ROC Analysis

The receiver operating characteristic (ROC) curves for the EfficientNet-B2 model are shown in Figure 2. The model achieved area-under-the-curve (AUC) values of 0.93 for the T2DM group (Class 0), 0.97 for the CKD stage 3 group (Class 1), and 1.00 for the control group (Class 2). These results demonstrate the model's strong ability to distinguish among disease categories, with particularly high accuracy in identifying CKD stage 3. The consistently high AUC values indicate that EfficientNet-B2 performs robustly in multiclass classification using retinal images.

The Confusion Matrix

Table 4 presents a secondary eye-level (per fundus photograph) confusion matrix for the EfficientNet-B2 model to summarize misclassification patterns across the three clinical groups, whereas primary performance metrics are based on patient-level predictions aggregated from bilateral eyes. The model correctly classified 49 out of 54 healthy control images, 17 out of 22 concurrent T2DM with CKD stage 3 images, and 11 out of 14 isolated T2DM images. Notably, the model demonstrated excellent specificity for the control group, as it rarely misclassified true disease cases as normal (0 for both T2DM and CKD).

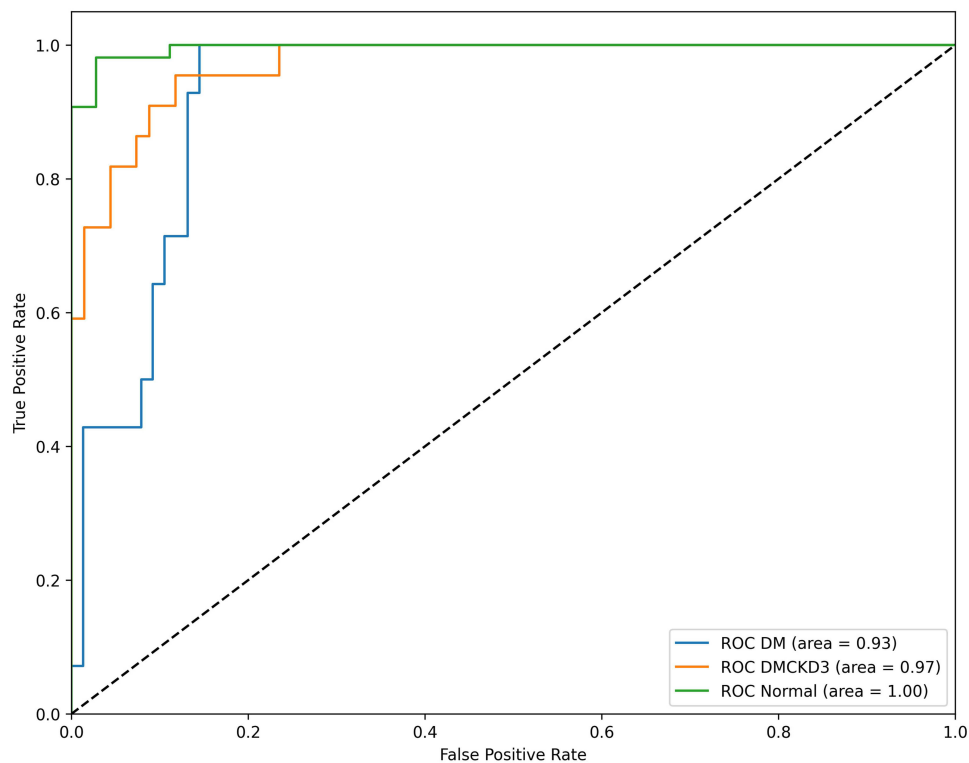


Figure 2 Receiver operating characteristic (ROC) curves of the EfficientNet-B2 model, showing area under the curve (AUC) values of 0.93 for T2DM (Class 0), 0.97 for CKD stage 3 (Class 1), and 1.00 for controls (Class 2).

Table 4 Eye-Level (per-Fundus Image) Confusion Matrix of the EfficientNet-B2 Model for Differentiating Controls, T2DM, and CKD Stage 3 (Test Set)

		Predicted Class		
		T2DM	CKD Stage 3	Control
True Class	T2DM	11	3	0
	CKD Stage 3	5	17	0
	Control	4	1	49

Abbreviations: CKD, Chronic Kidney Disease; T2DM, Type 2 Diabetes Mellitus.

The majority of misclassifications occurred between the disease categories: 5 true CKD images were incorrectly predicted as T2DM, and 3 true T2DM images were predicted as CKD. These specific misclassifications likely reflect the overlapping and subtle nature of early microvascular phenotypes shared between advancing diabetes and early-stage chronic kidney disease.

Attribution Map Results

The attribution maps generated using Grad-CAM from the EfficientNet-B2 model highlight condition-specific retinal regions that contributed most to the classification (Figure 3). In normal control eyes (A1–A3), the Grad-CAM heatmaps display a broad and diffuse activation across the entire image, indicating that the AI globally scans the retina for potential abnormalities. The model confirms the “Normal” classification because it detects no anomalies. For isolated T2DM patients (B1–B3), the model detects subtle, early microvascular alterations primarily around the optic disc (Zone A). Lacking visible diabetic retinopathy, these less pronounced features make differentiation challenging, which explains the lowest classification accuracy for this group (Table 4). Conversely, in patients with concurrent T2DM and CKD stage 3 (C1–C3), the model highlights established, widespread microvascular changes predominantly localized in Zone B, the most representative region for CKD-related damage. Ultimately, while these heatmaps must be interpreted cautiously as visual correlations requiring further clinical validation, this interpretability framework provides valuable clinical insights into the model’s behavior.

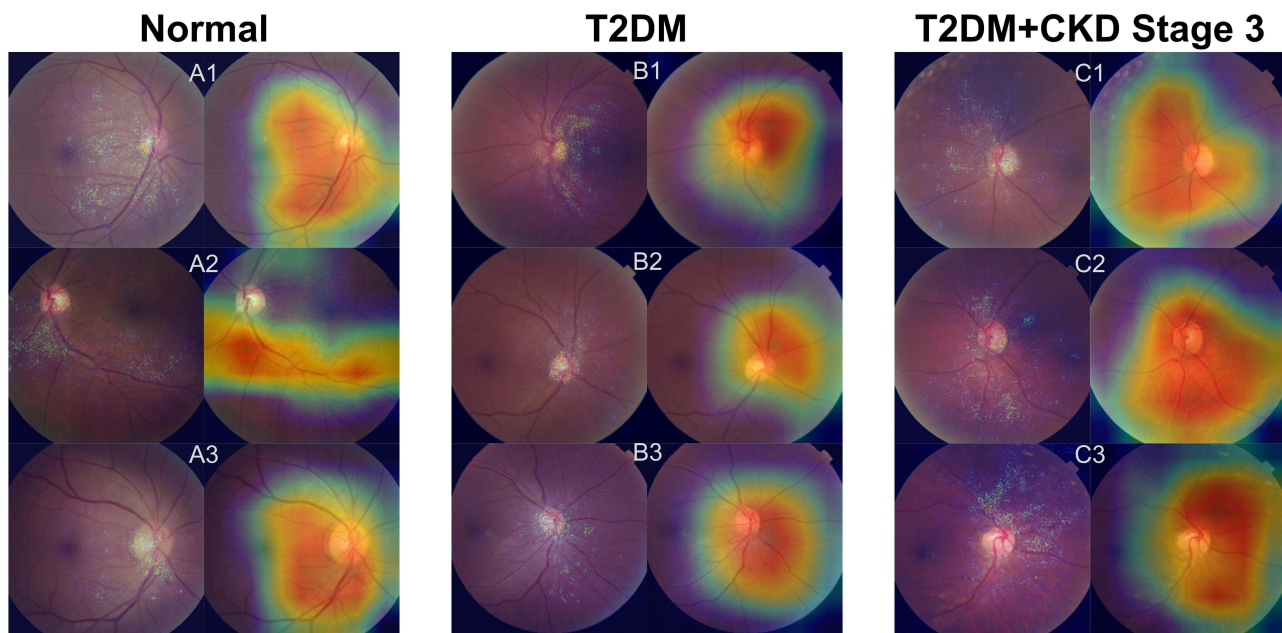


Figure 3 Gradient-weighted Class Activation Mapping (Grad-CAM) visualizations for model interpretability. The images are categorized by diagnostic class: A1–A3, Normal control eyes; B1–B3, Eyes from patients with isolated Type 2 Diabetes Mellitus (T2DM); and C1–C3, Eyes from patients with T2DM and Chronic Kidney Disease (CKD) stage 3. Any peripheral darkening represents acquisition-related shadow/vignetting from non-mydriatic imaging and should not be interpreted as retinal haemorrhage.

Discussion

Chronic kidney disease often progresses silently over an d is therefore well-suited for screening-based preventive strategies.^{2,15} We targeted CKD stage 3 because stages 1–2 are often non-specific, with eGFR values that may overlap normal age-related decline. Moreover, stages 1–2 frequently require additional evidence of kidney damage (eg, albuminuria/structural abnormalities), which may be absent or inconsistently captured, increasing label noise in a unimodal fundus-only setting. Stage 3 represents a more definitive and clinically actionable decline, providing a clearer diagnostic target for algorithmic detection.²

Integrating DLAs into retinal imaging can enhance detection of CKD and T2DM by quantifying subtle retinal microvascular alterations linked to systemic pathology. This oculomics perspective supports the retina as a non-invasive biomarker of systemic vascular health and may aid earlier identification of microvascular complications.¹⁷ Using non-invasive retinal fundus photographs, our three-class framework demonstrated strong discrimination of controls, T2DM, and T2DM with CKD stage 3, supporting the feasibility of fundus-only risk stratification based on microvascular signatures captured in routine retinal imaging.² Our AI model minimizes operator bias by enabling automated prediction without the need for expert annotation and can detect subtle microvascular abnormalities that are not routinely quantified in clinical practice. Previous studies clinical variables have shown strong performance for AI-based CKD detection (AUC 0.911 [95% CI, 0.886–0.936], 83% accuracy, and 83% sensitivity).² In the present study, our fundus-only model achieved comparable performance, with a sensitivity of 82%, specificity of 94%, and an AUC of 0.96.

Methodologically, we benchmark multiple backbones (ResNet, EfficientNet-B2, ViT, and ConvNeXt) under identical settings and mitigated class imbalance using weighted-loss strategy. EfficientNet-B2 achieved the best overall performance, consistent with its compound scaling design by Razali et al, which improve accuracy with fewer parameters and well-suited for limited, imbalanced medical datasets and resource-constrained deployment.¹⁸ Although DL models are often perceived as “black boxes”, fundus-based networks can plausibly detect both T2DM and CKD because the retina is an accessible microvascular bed that mirrors systemic endothelial dysfunction, inflammation, and microangiopathy shared by these conditions.^{19,20} By learning high-dimensional “oculomic” signatures beyond overt lesions (and in our study, images with visible diabetic retinopathy were excluded), the model can capture subtle vascular remodeling, such as shifts in vessel caliber, tortuosity/branching complexity, arteriovenous crossing patterns, and microvascular rarefaction that may be difficult to quantify by routine clinical grading.^{21,22} This mechanistic plausibility is further supported by our attribution map analysis (Grad-CAM), discussed below. Together, these properties support the potential of DL to translate minute retinal microvascular alterations into scalable screening signals for metabolic and renal impairment.

Grad-CAM provided supportive interpretability by highlighting retinal regions contributing to model predictions in the EfficientNet-B2 architecture.^{13,23} In our study, attribution maps consistently emphasized the retinal vascular network and peripapillary–macular regions, suggesting that the model relied on physiologically plausible microvascular features commonly evaluated in automated retinal image analysis. These findings align with the concept that subtle retinal microvascular alterations may reflect systemic vascular changes associated with metabolic and renal disease.^{8,9}

However, Grad-CAM is a qualitative saliency method and should be interpreted cautiously, as it indicates where the model attends rather than providing a causal explanation. Consistent with the confusion matrix, misclassifications occurred predominantly between the T2DM and T2DM with CKD stage 3 groups, likely reflecting overlapping and subtle microvascular phenotypes in the absence of overt diabetic retinopathy, while controls were rarely confused with disease. These findings highlight the challenge of differentiating early systemic complications using retinal imaging and underscore the need for larger, multi-device external validation to better characterize model failure modes and improve generalizability, while interpretability analyses may help enhance transparency and support clinical adoption.^{7,23}

Beyond technical performance, the translational value of AI-assisted fundus analysis depends on its integration into existing clinical workflows. Given its lightweight architecture, our EfficientNet-B2 may be suitable for deployment in teleophthalmology and portable fundus-camera programs, including on-device (edge) inference where connectivity is limited. This could support CKD risk triage during routine diabetic eye screening and streamline referral for confirmatory renal assessment in resource-limited settings such as Indonesia, where access is limited. Integration into diabetic eye-care pathways may enable ophthalmologists to support CKD risk triage by prompting confirmatory renal testing and referral.

Several limitations should be considered. First, the cross-sectional, single-center design and reliance on a single imaging device may limit external generalizability and introduce domain shift when applied to images acquired using different cameras or protocols. Second, we excluded eyes with visible diabetic retinopathy and other ocular pathologies to minimize overt lesion-driven confounding; therefore, the findings may not fully generalize to broader clinical screening settings where these conditions are frequently present. Third, factors such as age, hypertension, and overall vascular disease burden can influence retinal microvascular appearance independently of kidney function, and these effects cannot be completely separated in a unimodal fundus-only framework. Fourth, class imbalance across the three groups may still affect model behavior despite the weighted-loss strategy. Finally, although predictive performance was strong, DL models remain partly non-transparent; thus, attribution maps should be interpreted carefully as qualitative indicators of where the model focuses, rather than as evidence of causality.

Overall, these findings support retinal imaging as a non-invasive, low-cost adjunct for early CKD risk stratification, leveraging fundus photography that is already routine in diabetes care pathways. A key strength of this work is the development and validation of a retinal image-based algorithm that is feasible for primary care and community settings. Future studies should focus on external validation across diverse populations and imaging devices, prospective assessment of longer-term predictive value, and evaluation of multimodal extensions incorporating relevant clinical variables (eg, blood pressure, HbA1c, duration of diabetes, and laboratory biomarkers) to improve robustness and scalability for screening.

Conclusion

This work is an internal proof-of-concept showing that a multiclass DLAs model using non-invasive retinal fundus photographs, trained with a weighted-loss strategy to mitigate class imbalance, can provide promising discrimination of controls, T2DM, and T2DM with CKD stage 3. External validation across centers and imaging devices, and ideally prospective testing, are needed before clinical deployment. For screening/triage, the model can be calibrated to a sensitivity-prioritized operating point for CKD stage 3, with positive screens routed to confirmatory kidney testing (eGFR/serum creatinine and urine ACR) and referral according to standard care pathways.

Ethical Approval and Consent to Participate

The study was approved by the Ethics Committee of Dr. Saiful Anwar General Hospital (Approval No. 400/252/K.3/102.7/2024) and conducted in accordance with the Declaration of Helsinki.

Acknowledgments

The authors would like to thank Intan Kautsarani, MD, for assistance with manuscript formatting and preparation according to the journal submission guidelines and Panti Nirmala Hospital for providing facilities and support as a data collection site.

Author Contributions

All authors made a significant contribution to the work reported, whether that is in the conception, study design, execution, acquisition of data, analysis and interpretation, or in all these areas; took part in drafting, revising or critically reviewing the article; gave final approval of the version to be published; have agreed on the journal to which the article has been submitted; and agree to be accountable for all aspects of the work.

Funding

This research received funding from the Directorate of Research and Community Service, Universitas Brawijaya (Grant No. 00565/UN10.A0501/B/PT/2024).

Disclosure

The authors declare no competing interest in this study.

References

- McFarlane P, Cherney D, Gilbert RE, Senior P. Chronic Kidney Disease in Diabetes. *Can J Diabetes*. 2018;42:S201–S209. doi:10.1016/j.jcjd.2017.11.004
- Sabanayagam C, Xu D, Ting DSW, et al. A deep learning algorithm to detect chronic kidney disease from retinal photographs in community-based populations. *Lancet Digit Health*. 2020;2(6):e295–e302. doi:10.1016/S2589-7500(20)30063-7
- Shi S, Gao L, Zhang J, et al. The automatic detection of diabetic kidney disease from retinal vascular parameters combined with clinical variables using artificial intelligence in type-2 diabetes patients. *BMC Med Inform Decis Mak*. 2023;23(1). doi:10.1186/s12911-023-02343-9
- Parchwani D, Upadhyah A. Diabetic nephropathy: progression and pathophysiology. *Int J Med Sci Public Health*. 2012;1(2):59. doi:10.5455/ijmsph.2012.1.59-70
- Tom ES, Saraf SS, Wang FP, et al. Retinal capillary nonperfusion on OCT-angiography and its relationship to kidney function in patients with diabetes. *J Ophthalmol*. 2020;2020. doi:10.1155/2020/2473949
- Abdullah FI, Rahman NA, Kamaluzaman QA, Aris MA, Basri MA. Prevalence of chronic kidney disease and its associated factors among type-2 diabetes mellitus patients at kuantan primary health clinics. *IUM Med J Malaysia*. 2025;24(1):142–150. doi:10.31436/imjm.v24i01.2613
- Betzler BK, Chee EYL, He F, et al. Deep learning algorithms to detect diabetic kidney disease from retinal photographs in multiethnic populations with diabetes. *J Am Med Inf Assoc*. 2023;30(12):1904–1914. doi:10.1093/jamia/ocad179
- Hamzah NAA, Wan Zaki WMD, Wan Abdul Halim WH, Mustafar R, Saad AH. Evaluating the potential of retinal photography in chronic kidney disease detection: a review. *PeerJ PeerJ Inc*. 2024;12(8). doi:10.7717/peerj.17786
- Wen J, Liu D, Wu Q, Zhao L, Iao WC, Lin H. Retinal image-based artificial intelligence in detecting and predicting kidney diseases: current advances and future perspectives. *View*. 2023;4(3). doi:10.1002/VIW.20220070
- Johnson JM, Khoshgofaar TM. Survey on deep learning with class imbalance. *J Big Data*. 2019;6(1). doi:10.1186/s40537-019-0192-5
- Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: transformers for image recognition at scale. Available from: <https://github.com/>. Accessed April 01, 2026.
- Liu Z, Mao H, Wu CY, Feichtenhofer C, Darrell T, Xie S. A ConvNet for the 2020s. Available from: <https://github.com/facebookresearch/ConvNeXt>. Accessed April 01, 2026.
- Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-CAM: visual explanations from deep networks via gradient-based localization. *Proceedings of the IEEE International Conference on Computer Vision*. 2017:618–626. doi:10.1109/ICCV.2017.74.
- Zhang K, Liu X, Xu J, et al. Deep-learning models for the detection and incidence prediction of chronic kidney disease and type 2 diabetes from retinal fundus images. *Nat Biomed Eng*. 2021;5(6):533–545. doi:10.1038/s41551-021-00745-6
- Stevens PE, Ahmed SB, Carrero JJ, et al. KDIGO 2024 clinical practice guideline for the evaluation and management of chronic kidney disease. *Kidney Int*. 2024;105(4):S117–S314. doi:10.1016/j.kint.2023.10.018
- ElSayed NA, Aleppo G, Bannuru RR, et al; Committee ADAPP. 2. diagnosis and classification of diabetes: standards of care in diabetes—2024. *Diabetes Care*. 2024;47(Supplement_1):S20–S42. doi:10.2337/DC24-S002
- Giarratano Y, Pavel A, Lian J, et al. A framework for the discovery of retinal biomarkers in Optical Coherence Tomography Angiography (OCTA). *Lecture Notes in Computer Science*. 2020:155–164. 12069 LNCS. doi:10.1007/978-3-030-63419-3_16.
- Razali MN, Arbaify N, Lin PC, Ismail S. Optimizing multiclass classification using convolutional neural networks with class weights and early stopping for imbalanced datasets. *electronics*. 2025;14(4):705. doi:10.3390/ELECTRONICS14040705
- Abbas Q, Daadaa Y, Rashid U, Sajid MZ, Ibrahim MEA. HDR-efficientnet: a classification of hypertensive and diabetic retinopathy using optimize efficientnet architecture. *Diagnostics*. 2023;13(20):3236. doi:10.3390/diagnostics13203236
- Mandal AC, Phatak A. Optimizing deep learning based retinal diseases classification on optical coherence tomography scans. In: *Optical Coherence Imaging Techniques and Imaging in Scattering Media V*; 2023. doi:10.1117/12.2672249.
- Al-Smadi M, Hammad M, Baker QB, Al-Zboon SA. A transfer learning with deep neural network approach for diabetic retinopathy classification. *Int J Electr Comput Eng*. 2021;11(4):3492–3501. doi:10.11591/ijece.v11i4.pp3492-3501
- Olabanjo O, Wusu A, Mazzara M. Deep Unsupervised Machine Learning for Early Diabetes Risk Prediction Using Ensemble Feature Selection and Deep Belief Neural Networks. doi:10.20944/PREPRINTS202301.0208.V1
- Ikram A, Imran A. ResViT FusionNet Model: an explainable AI-driven approach for automated grading of diabetic retinopathy in retinal images. *Comput Biol Med*. 2025;186. doi:10.1016/j.combiomed.2025.109656

Clinical Ophthalmology

Publish your work in this journal

Clinical Ophthalmology is an international, peer-reviewed journal covering all subspecialties within ophthalmology. Key topics include: Optometry; Visual science; Pharmacology and drug therapy in eye diseases; Basic Sciences; Primary and Secondary eye care; Patient Safety and Quality of Care Improvements. This journal is indexed on PubMed Central and CAS, and is the official journal of The Society of Clinical Ophthalmology (SCO). The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/clinical-ophthalmology-journal>

Dovepress
Taylor & Francis Group