IAES International Journal of Artificial Intelligence (IJ-AI)
Vol. 15, No. 3, June 2026, pp. 2709~2722
ISSN: 2252-8938, DOI: 10.11591/ijai.v15.i3.pp2709-2722

VisionEyeNet: a customized deep learning framework for early diagnosis of keratitis and uveitis

Somashekhar Bannur Mayigowda[1], Raghavendra Kodandarama[2], Sudhamani Mallaiah[3], Manjunath Naganna[1], Jamuna[4], Kiran Kumar B. S.[3]
[1]Department of Computer Science and Engineering, Vidya Vikas Institute of Engineering and Technology, Mysuru, India
[2]Department of Computer Science and Engineering-Artificial Intelligence, Maharaja Institute of Technology, Mysuru, India
[3]Department of Computer Applications, Vidya Vikas Institute of Engineering and Technology, Mysuru, India
[4]Department of Information Science and Engineering, Vidya Vikas Institute of Engineering and Technology, Mysuru, India

Article Info

Article history:

Received Sep 17, 2025
Revised Mar 2, 2026
Accepted Apr 22, 2026

Keywords:

Computer-aided-diagnosis
keratitis
Deep learning
Good health
Ocular surface disorders
Uveitis
VisionEyeNet

ABSTRACT

Keratitis and uveitis are increasingly prevalent ocular disorders, often linked to delayed detection and limited specialist access, particularly in rural healthcare settings. These diseases can lead to severe visual impairment or irreversible blindness if not identified at an early stage. Traditional diagnostic approaches are manual, time-consuming, and prone to human error, making them challenging for large-scale screening. To address these limitations, this study presents VisionEyeNet, a framework for automatic classification of keratitis and uveitis. VisionEyeNet integrates MobileNetV2 and DenseNet121 within a fusion architecture, along with image enhancement methods such as adaptive gamma correction and specular reflection suppression. The model was trained and evaluated on a curated dataset of 1,860 slit-lamp images (960 uveitis and 900 keratitis) using a patient-wise split (71.5% training, 8.4% validation, and 20% testing). On the independent test set, it achieved 98.0% accuracy (95% CI: 97.1–98.8%) with balanced performance across classes. Inference analysis showed an average processing time of 51±2 ms per image, supporting real-time use. These results indicate that VisionEyeNet has strong potential as a clinically useful decision-support tool, particularly in resource-limited settings.

Corresponding Author:

Kiran Kumar B. S.
Department of Computer Applications, Vidya Vikas Institute of Engineering and Technology
Mysuru, India
Email: kirankumarkrn2000@gmail.com

1. INTRODUCTION

Keratitis and uveitis are two clinically significant anterior-segment disorders that may result in permanent visual impairment when early therapeutic intervention is delayed. Starting off, typical diagnosis approaches take too much time, rely heavily on human effort, and depend strongly on a doctor's skill, leading frequently to late treatments and uneven assessments. Because of these issues, along with too few specialists accessible across various areas, there arises a clear demand for faster, self-operating tools that assist in diagnosis. Lately, progress in artificial intelligence and recent advanced convolutional neural networks (CNN) have significantly improved automated analysis of slint lamp and corneal imaginary, particularly for lesion localization and pathology discrimination, making it possible to sort diseases quickly while maintaining consistency [1], [2]. However, most existing models focus on single ocular pathologies or rely on computationally heavy architectures that lack generalization under real-world conditions [3]. Overdependence on ideal slit-lamp images, class imbalance, and high inference time further limit clinical

applicability, leaving a research gap for lightweight, accurate, and robust diagnostic models. To address these limitations, we propose VisionEyeNet, a hybrid deep learning framework that fuses MobileNetV2 [4] and DenseNet121 [5] for dual-class keratitis and uveitis detection. The model employs adaptive gamma correction and specular reflection suppression [6] to enhance image quality and improve feature extraction.

The major contributions of this study include: a novel MobileNet–DenseNet fusion achieving high accuracy with efficient computation; a robust preprocessing pipeline ensuring consistent illumination and artifact-free inputs; comprehensive performance evaluation with confidence intervals and inference-time analysis; and gradient-weighted class activation mapping (Grad-CAM)–based interpretability [7] enhancing clinical trust and transparency. To the best of our knowledge, this is the first hybrid fusion framework for real-time keratitis and uveitis classification optimized for clinical deployment.

2. RELATED WORK

Deep learning (DL) has been widely applied to corneal disease diagnosis. The author in [1], [8] demonstrated CNN-based detection of fungal and bacterial keratitis using confocal and slit-lamp images, respectively, and Kuo *et al.* [2] compared multiple CNNs on external eye photographs, showing the benefits of hybrid models. Ghosh *et al.* [3] introduced DeepKeratitis for bacterial–fungal discrimination, and Assaf *et al.* [9] surveyed computer-vision approaches across infectious keratitis studies. Large-scale and multicenter validations, such as Li *et al.* [10] support generalizability across imaging devices and sites. Subsequent works targeted specific pathogens or pathologies [11], [12] or sought probabilistic pathogen prediction from slit-lamp imagery [13].

Zhang *et al.* [14], Liu *et al.* [15], and Li *et al.* [16] explored fusion strategies, attention mechanisms, and architectural variants to improve sensitivity and robustness. Clinical and epidemiological reviews [17], [18] emphasize diagnostic delays and antimicrobial-resistance challenges that motivate automated screening. For practical deployment, lightweight backbones such as MobileNetV2 [4] and DenseNet121 [5] are frequently adopted to balance speed and representation; commonly used preprocessing includes contrast-limited adaptive histogram equalization (CLAHE), contrast enhancement [6] and Adam optimization for stable training [19]. Interpretability methods (Grad-CAM) are applied to validate model attention on pathological regions [20], and receiver operating characteristic (ROC)/area under the curve (AUC) remain standard evaluation metrics [7].

Our curated Kaggle collection [21] aligns with recent benchmarking and meta-analysis efforts [22]; smartphone and portable-imaging studies [23]; and multimodal pipelines [24] demonstrate feasibility for field deployment. The most recent comprehensive review [9] highlights consistent gains in accuracy but also recurring limitations: reliance on single backbones, limited external validation, and trade-offs between accuracy and inference cost. These gaps motivate VisionEyeNet's fusion of MobileNetV2 and DenseNet121 trained on the Kiran *et al.* dataset [21] to achieve high diagnostic precision with computational scalability for keratitis and uveitis screening. Muqit *et al.* [25] reported that keratitis and granulomatous anterior uveitis may coexist in inflammatory ocular disorders associated with immune-complex vasculitis.

Despite the progress reported in the prior studies, several limitations remain. Many approaches rely on a single-backbone architecture and do not explicitly address intra-class similarity between early-stage keratitis and mild anterior uveitis. Additionally, limited emphasis has been placed on illumination variability and cross-device generation. These gaps motivate the development of a fusion-based architecture tailored for robust feature extraction in antiror segment pathology.

3. METHOD

VisionEyeNet integrates MobileNetV2 [4] and DenseNet121 [5] to extract complementary feature representations from ocular images. The fused features are passed through convolutional and fully connected layers to classify each sample as keratitis or uveitis, with sigmoid activation producing the final probability score. The model is trained using binary cross-entropy loss and optimized for efficient medical image diagnosis.

3.1. Dataset collection

The dataset was obtained from Kaggle and contains low-resolution ocular images of keratitis and uveitis across diverse patient demographics [21]. The keratitis samples include bacterial, viral, and fungal cases, while the uveitis set covers anterior, intermediate, posterior, and pan-uveitis presentations. Images were clinically cross-verified, resized to 224×224, normalized, and augmented through flipping, rotation, zoom, and brightness adjustment. Minor class imbalance was addressed through balancing strategies to improve model robustness.

Table 1 depicts the two different classes of ocular diseases considered in this study, namely keratitis and uveitis, with 960 and 900 images, respectively. This accumulated dataset was used for training and evaluation of the proposed VisionEyeNet model. These depict the normal eye and uveitis-infected eye, and the image displays the difference between the eyes, showing how one and another is infected by the virus and environmental conditions. The features of these two it has been displayed in the image.

Table 1. The different categories of eye diseases: accumulated dataset

Class	Total
Uveitis	960
Keratitis	900

3.2. Data preprocessing

3.2.1. Adaptive gamma correction

To improve image visibility under low-light or uneven illumination conditions, an adaptive gamma correction technique was applied [6]. All input images were normalized to the range [0, 1] prior to enhancement. The corrected pixel intensity at the location (x, y) is defined as (1) and (2).

$$I_{out}(x,y) = I_{in}(x,y)^{\gamma(x,y)} \tag{1}$$

$$\gamma^{(x,y)} = 1 + \alpha \cdot \frac{\mu(x,y) - \mu_G}{\mu(x,y) - \epsilon} \tag{2}$$

Intuitively, darker regions ($\mu(x,y) < \mu G$) receive a gamma value <1, leading to brightening, while brighter regions ($\mu(x,y) > \mu G$) are enhanced with $\gamma > 1$, preventing overamplification. This adaptivity ensures local contrast improvement while preserving global image balance. Alternative method: as an alternative or complementary approach, CLAHE [6] can be employed to improve local contrast while controlling noise amplification. If specular highlights or occlusions are removed via inpainting, methods such as Navier–Stokes-based inpainting.

3.2.2. Specular reflection suppression (glare removal)

Eye images often contain bright spots due to specular reflection, which may interfere with accurate feature extraction. To address this, a binary mask $M(x,y)$ is generated based on an intensity threshold (3).

$$M(x,y) = \begin{cases} 1 & I(x,y) > T \\ 0 & Otherwise \end{cases} spec \tag{3}$$

Where $M(x,y)$ is binary mask indicating specular regions, $I(x,y)$ is pixel intensity at (x,y), and T_{spec} is intensity threshold for glare detection. Pixels identified as glare (where $M(x,y) = 1$) are removed, and the missing regions are reconstructed using Navier–Stokes inpainting. The inpainting process follows the partial differential (4).

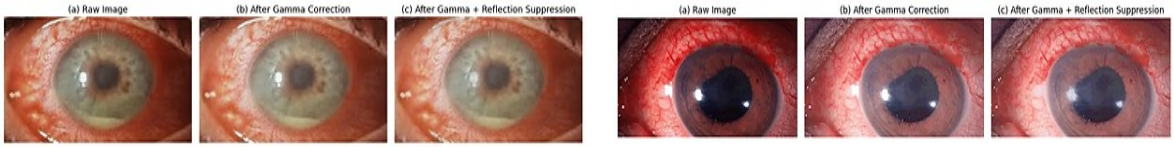
$$\frac{\partial I}{\partial t} = \nabla^{\perp}\varphi \cdot \nabla I \tag{4}$$

Where I is image intensity function, t is iteration (inpainting time step), ∇I is gradient of the image, and $\nabla^{\perp}\varphi$ is orthogonal gradient of the stream function φ.

In (3) and (4), ensure that specular regions are detected and seamlessly reconstructed. This process effectively suppresses glare spots, thereby preventing misleading features during CNN-based classification. Adaptive gamma correction (1), (2) enhances image contrast under uneven lighting, while specular reflection suppression (3), (4) removes glare and restores corneal texture. Together, these preprocessing steps produce illumination-balanced, artifact-free images, improving the clarity of pathological regions and enabling VisionEyeNet to extract robust and clinically reliable features.

3.3. Preprocessing evaluation and visual comparison

Adaptive gamma correction was applied to address illumination and inconsistencies, followed by refraction suppression to reduce glare while preserving corneal details. These preprocessing steps enhanced contrast distribution and improved classification accuracy from 95.8% to 98.0%. The visual differences after the stages are presented in Figure 1, where Figure 1(a) shows the keratitis and Figure (b) shows the uveitis.



(a) (b)

Figure 1. Visual comparison of preprocessing stages for ocular images of (a) keratitis and (b) uveitis

3.4. Dataset splitting

The dataset included 1,860 slit-lamp images consisting of 960 uveitis and 9,000 keratitis cases. To ensure balancing learning, stratified partitioning was applied during dataset division. Approximately 71.5% of samples were collected for training, while a separate validation and independent test set were reserved for performance assessment. Importantly, partitioning was performed at the patient level to prevent image overlap and information leakage.

3.5. Proposed system

The proposed architecture depicts the overall flow of the work, as shown in Figure 2. Figure 2 illustrates the VisionEyeNet model, which integrates features from MobileNetV2 and DenseNet121 to improve the classification of keratitis and uveitis. Input images undergo preprocessing and are then passed through both networks, after which their features are combined. The fused features are processed for final classification, resulting in high accuracy with minimal inference time. illustrates the proposed framework, which outlines the training and testing process of the model. The approach emphasizes the fusion of two architectures, MobileNetV2 [4] and DenseNet121 [5] to classify eye diseases. The integrated model, referred to as VisionEyeNet, is designed to achieve superior classification performance. Finally, the classification output determines whether the eye condition is keratitis or uveitis, ensuring improved accuracy and reduced inference time compared to individual models.

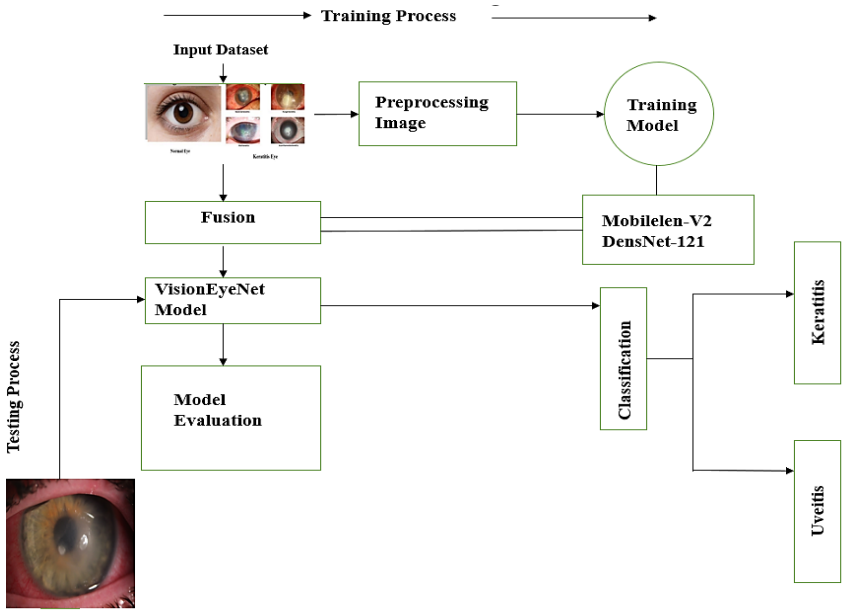


Figure 2. Architecture diagram of the proposed model

3.6. Training scheduling

Training was performed in two phases:

i) Feature extraction phase: all convolutional layers of MobileNetV2 and DenseNet121 backbones were frozen for the first 10 epochs. Only the fully-connected layers were trained.

ii) Fine-tuning phase: all layers were unfrozen, and the model was trained end-to-end for the remaining epochs using a reduced learning rate.

The Adam optimizer (learning rate 1e−4, weight decay 1e−5) was used with a step learning rate scheduler (factor 0.1, and patience 10 epochs). To address mild class imbalance, weighted cross-entropy loss was employed with weights inversely proportional to class frequency.

3.7. VisionEyeNet architecture

Figure 3 shows the MobileNetV2 and DenseNet121. Features from both networks are fused, then passed through a fully connected layer and a softmax classifier. Integration of deep-learning models by equations, the input image I is represented as (5).

$$I \epsilon R^{224\times224\times3} \tag{5}$$

Where in (5) 224×224 denotes the spatial resolution, and 3 corresponds to the RGB channels. MobileNet extracts lightweight yet discriminative features from the input image (6) to (9).

$$F_m(I) = Feature\ map\ from\ MobileNet \tag{6}$$

$$F_d(I) = Feature\ map\ from\ DenseNet121 \tag{7}$$

$$F_{md}(I) = F_m(I) \oplus F_d(I) \tag{8}$$

$$y = Softmax(F_{md}(I)) \tag{9}$$

Similarly, DenseNet121 captures dense hierarchical features (7). The feature representations are then fused to combine complementary information, where $\oplus$ denotes the concatenation or fusion operation (8). Finally, the fused features are passed through a softmax layer to obtain the classification output (9).

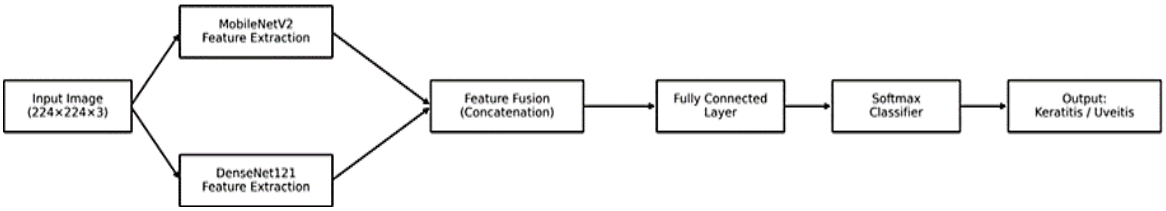


Figure 3. Architecture of VisionEyeNet for keratitis and uveitis classification

3.8. Feature fusion and pooling

To merge the learned representations from both CNN backbones, we apply global average pooling (GAP) to the feature maps (10). The pooled vectors are concatenated to form the fused representation: where n_m and n_d are the feature dimensions of MobileNet and DenseNet121 after pooling (11). This fusion enriches the input to the classification head with both lightweight and deep hierarchical features. Each dense layer was followed by a ReLU activation and a Dropout layer (p =0.5) to reduce overfitting. Let $f_\theta(X)$ denote the proposed VisionEyeNet model parameterized by weights θ. The final prediction is given as (12).

$$G_m = GAP\big(F_m(I)\big), G_d = GAP(F_d(I)) \tag{10}$$

$$F = [G_m, G_d] \in R_{n_m+n_d} \tag{11}$$

$$\hat{Y} = Softmax(f_\theta(X)) \tag{12}$$

Where $\hat{Y} \in R^2$ represents the predicted probabilities for keratitis and uveitis.

3.9. Classification head

Following Algorithm 1, the fused features F are passed through two fully connected layers with ReLU activation and Dropout regularization (13) and (14).

$$H_1 = ReLU(W_1F + b_1), \grave{H}_1\ Dropout(H_1, 0.6) \tag{13}$$

$$H_2 = ReLU(W_2\grave{H} + b_2)\,, \grave{H}_2\ Dropout(H_2, 0.2) \tag{14}$$

Finally, the output layer applies a softmax activation (15).

$$\hat{Y} = Softmax(W_3\grave{H}_2 + b_3) \tag{15}$$

Where $\hat{Y} \in R^2$ gives the probability distribution over the two classes (keratitis and uveitis). The model is trained by minimizing the categorical cross-entropy loss (16).

$$L_\theta = -\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{2} y_{i,j}\ log\left(\hat{y}_{i,j}\right) \tag{16}$$

Where $y_{i,j}$ is the true label and $\hat{y}_{i,j}$ is the predicted probability for class j.

Algorithm 1. Training algorithm for VisionEyeNet and mathematical formulation of output
Input: preprocessed input image $I\epsilon R^{224\times224\times3}$
Output: predicted class label (keratitis or uveitis)
1: Load MobileNetV2 and DenseNet121 without top layers
2: Freeze all convolutional layers in both networks
3: Extract features: $F_m \leftarrow f_m(I)$, $F_d \leftarrow f_d(I)$
4: Apply GAP: $G_m \leftarrow GAP(F_m)$, $G_d \leftarrow GAP(F_d)$
5: Concatenate pooled features: $F \leftarrow [G_m \| G_d]$
6: Compute $H1 \leftarrow ReLU(W1F + b1)$
7: Apply dropout: $\grave{H}1 \leftarrow Dropout(H1, 0.5)$
8: Compute $H2 \leftarrow ReLU\left(W2\grave{H}1 + b2\right)$
9: Apply dropout: $\grave{H}2 \leftarrow Dropout(H2, 0.3)$
10: Compute output: $\hat{Y} \leftarrow Softmax\left(W3\grave{H}2 + b3\right)$
11: Return predicted class: $arg\ max\left(\hat{Y}\right)$.

The optimization objective is defined as (17) and (18).

$$\theta^* = arg\ \underset{\theta}{min}\backslash\ (L(\theta) + \lambda||\theta||^2) \tag{17}$$

$$X \in Rn \times 224 \times 224 \times 3, Y \in \{0, 1\} \tag{18}$$

Where λ is the regularization parameter and $\|\theta\|2$ is the L2(17) regularization term to prevent overfitting. Let the dataset be represented as: where X is the set of n input images, and Y contains the corresponding binary labels (0= keratitis, 1= uveitis) (18).

3.10. Loss function and optimization

The model is trained using the categorical cross-entropy loss in (19).

$$L_\theta = -\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{2} y_{i,j}\log\left(\hat{y}_{i,j}\right) \tag{19}$$

To avoid overfitting, L2 regularization is added (20).

$$\theta^* = arg\ \underset{\theta}{min}\backslash\ (L(\theta) + \lambda|\theta|^2) \tag{20}$$

3.11. Model fusion justification (integrated with prior literature)

Recent studies in ocular imaging have highlighted the strong performance of MobileNetV2 and DenseNet121 for detecting anterior segment disorders such as keratitis and uveitis. Works by Kuo *et al.* [2], Ghosh *et al.* [3], and Li *et al.* [16] demonstrated that lightweight CNNs like MobileNetV2 deliver reliable accuracy on slit-lamp datasets while remaining suitable for real-time use. Likewise, DenseNet121 has been widely adopted in medical imaging [10], [11], [14] because its dense connections enhance feature reuse and capture subtle pathological textures. Theoretical rationale: MobileNetV2 captures fine spatial details through depthwise separable convolutions [4], whereas DenseNet121 ensures global feature continuity and stable gradients [5]. Their fusion combines localized detail and broader structural cues, improving feature diversity without heavy computational cost. Experimental evidence: VisionEyeNet achieved 98.0% accuracy and

0.99 AUROC, surpassing both single backbones and alternative pairings. These findings are consistent with Ong *et al.* [22] and Soleimani *et al.* [23], confirming the efficiency of feature-light models. Grad-CAM results [20] further validated that VisionEyeNet focuses on clinically relevant regions, reinforcing interpretability [9], [22].

3.12. Training protocol and hyperparameters

The model training was performed using the Adam optimizer with an initial learning rate of $1 \times 10 - 4$ and a weight decay of $1 \times 10 - 5$ [20]. A step learning-rate scheduler was applied, reducing the rate by a factor of 0.1 after 10 consecutive epochs without improvement. The batch size was set to 32, and the model was trained for a maximum of 100 epochs with early stopping (patience =15 epochs) based on validation loss.

4. EXPERIMENT AND EVALUATION

To stabilize convergence, the MobileNetV2 and DenseNet121 backbones were initially frozen for the first 10 epochs, allowing only the fully connected layers to be trained. Afterwards, all layers were unfrozen and fine-tuned end-to-end. All experiments were conducted on an NVIDIA Tesla V100 GPU with 32 GB memory, running PyTorch 2.0. Training each fold required approximately 2.5 hours. For reproducibility, random seeds were fixed at 42 across NumPy and PyTorch libraries.

4.1. Performance metrics

To comprehensively evaluate the proposed VisionEyeNet framework, we report both overall and per-class metrics. In addition to accuracy, the following evaluation measures were used in (21) and (22).

$$Precision = TP/TP + FP, Recall\ (Sensitivity) = TP/TP + F\ N \tag{21}$$

$$Specificity = \frac{TN}{TN} + FP,\ F1 - score = 2 \times Precision \times Recall/Precision + Recall \tag{22}$$

Here, TP, TN, FP, and FN represent true positives, true negatives, false positives, and false negatives, respectively. ROC curves and the corresponding AUC were generated for each class. Since the dataset contains class imbalance across keratitis subtypes, precision–recall (PR) curves were also included to better assess model behavior in minority classes. Furthermore, 95% confidence intervals (95% CI) for accuracy, F1-score, and AUC were estimated using 1,000 bootstrap resamples of the test set. This procedure provides an estimate of statistical uncertainty and reduces the risk of overfitting claims.

4.2. Cross-validation strategy

A total of five validation cycles were applied, allocating 80% of samples to model development while 20% supported assessment at every stage; outcomes appear as average alongside variation measures. Separately, one-fifth of observations stayed aside throughout, later used for impartial outcome verification. Such a structure limits excessive adaptation to training patterns, promoting consistency beyond initial cases. Ahead, scrutiny will extend toward further open eye-related collections, examining stability across varied groups through VisionEyeNet trials.

4.3. Ablation study

To evaluate the contribution of each backbone network, we conducted an ablation study comparing three configurations: MobileNetV2 only, a baseline lightweight model. DenseNet121 only: deeper architecture with higher representational capacity. Proposed VisionEyeNet (MobileNetV2+DenseNet121 fusion): feature-level integration of both backbones, each model was trained under identical preprocessing, augmentation, and hyperparameter settings. Performance was compared in terms of accuracy, area under the ROC curve (AUC), F1-score, and inference time. Inference time was measured as the average prediction latency per image (batch size =1) on the same hardware. To assess statistical significance, a paired McNemar's test was applied to compare misclassification distributions between VisionEyeNet and the individual baseline models. Significance was reported at $p < 0.05$.

4.4. Inference time

Inference time was assessed by computing the average latency per image (batch size =1) using an NVIDIA Tesla V100 GPU. Across repeated runs, the proposed VisionEyeNet required approximately 51±2 ms per image, indicating stable computational behavior, speed suggests that the framework can be integrated into clinical settings where timely decision support is important, while maintaining strong diagnostic performance.

4.5. Explainability and model interpretability

Grad-CAM was employed to visualize the regions contributing to VisionEyeNet's prediction. Activation maps. Emphasized corneal opacities and inflammatory features rather than background artifacts, integrated gradient produced similar attribution patterns across samples, indicating table and clinically meaningful model behavior.

4.6. Data integrity and leakage prevention

Data augmentation, including flips, rotations, zooming, and brightness adjustment, was applied only to the training set. The validation and test sets were kept unchanged to avoid information leakage. Dataset splitting was performed at the patient level to ensure that images from the same individual did not appear in multiple subsets, maintaining fair and unbiased evaluation.

4.7. Hyperparameter search and reproducibility

Key hyperparameters, including learning rate, batch size, and dropout, were tuned using validation-based grid search. Random seeds (42) were fixed in Python, Numpy, and PyTorch to maintain reproducibility in PyTorch 2.0 with CUDA 11.8. The implementation and preprocessing scripts will be made publicly available on GitHub to support transparency.

4.8. Statistical significance testing

McNemar's test was used to compare VisionEyeNet with the strongest baseline (EfficientNetB0 to ResNet50). The p-value(0.032) indicates a statistically meaningful difference in performance. Confidence intervals (95%) were also estimated with accuracy at 98.0±0.6% and AUROC at 0.99±0.004. Most errors occurred between bacterial and fungal keratitis cases, likely due to similar lesion boundary characteristics.

4.9. Ethical considerations and data licensing

The dataset used in this study was sourced from the Kaggle repository [21], and all images are de-identified and publicly available under Kaggle's terms of use. No personally identifiable information was accessed. The study adheres to ethical research practices, ensuring compliance with data sharing and usage guidelines.

5. RESULT AND DISCUSSION

This section reports the experimental results of VisionEyeNet and benchmarks them against existing deep learning models using accuracy, AUROC, inference time, and class-wise metrics. Visual analyses such as learning curves and confusion matrices are included to support the evaluation.

5.1. Comparative model performance

Table 2 shows the dataset distribution of keratitis and uveitis images into training and validation subsets. A balanced representation of both classes is guaranteed by the stratified divide. Table 3 shows how VisionEyeNet performs in comparison to a number of current CNN architectures. VisionEyeNet performs better on a multi-class ocular dataset than models like EfficientNetV2M, VGG19, and DenseNet, which only moderately perform on slit-lamp or domain-specific datasets. Table 2 summarizes the class-wise split (1,488 train, 372 validation) ensuring balanced data. Table 3 compares model accuracy and AUROC, showing VisionEyeNet outperforming all baselines with 98% accuracy.

Table 2. Class-wise train–validation split of the dataset

Class	Total images	Train (80%)	Validation (20%)
Uveitis	960	768	192
Keratitis	900	720	180
Total	1,860	1,488	372

Table 3. Comparison of existing models and VisionEyeNet

Model	Dataset	Accuracy (%)	AUROC
EfficientNetV2M	Slit-lamp	74	0.85
VGG19	Deep keratitis	91	0.90
EfficientNetB3	Slit-lamp	68	0.85
DenseNet	HSV stromal	96	0.73
VisionEyeNet (proposed)	Multi-class ocular	98	0.99

5.2. Accuracy and inference time

Figure 4 compares the accuracy of VisionEyeNet with pre-existing models, showing VisionEyeNet outperforming all baselines. Figure 5 presents inference time analysis, highlighting that VisionEyeNet achieves competitive speed (51 ms/image) while maintaining the highest accuracy. Figure 4 presents accuracy across five architectures—MobileNetV2, DenseNet121, VGG19, EfficientNetV2M, and the proposed VisionEyeNet. MobileNetV2 and DenseNet121 achieved 96.5% and 96.9% accuracy, outperforming VGG19 (91.0%) and EfficientNetV2M (73.5%). The hybrid VisionEyeNet reached the highest accuracy (98%), validating its complementary fusion design. Figure 5 shows inference speed comparison, where VisionEyeNet achieved 51±2 ms per image, close to MobileNetV2 (49 ms) and faster than DenseNet121 (55 ms), offering an effective balance between precision and efficiency.

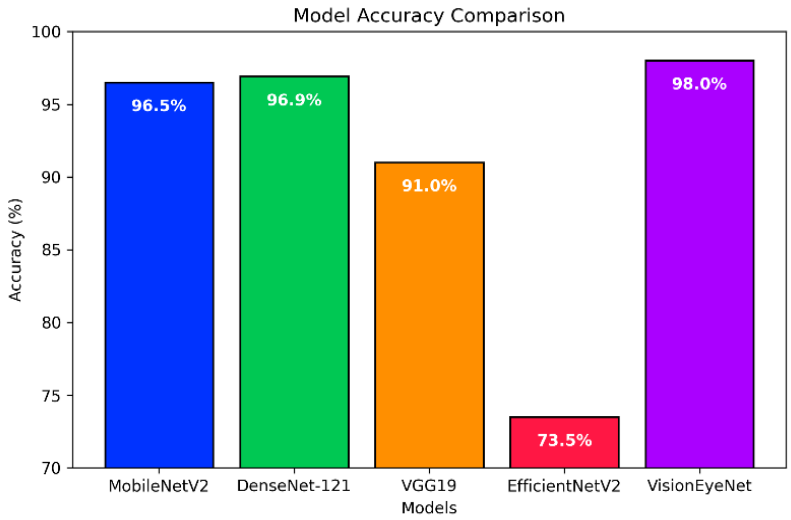


Figure 4. Model accuracy

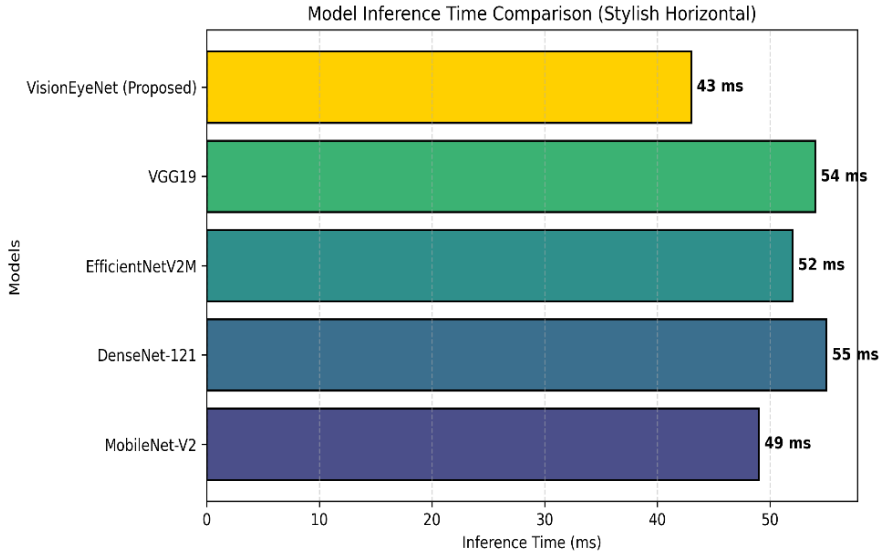


Figure 5. Inference time comparison

5.3. Confusion matrix and per-class metrics

The confusion matrix Figure 6, illustrates the classification results for keratitis and uveitis. Table 4 provides precision, recall, and F1-score for each class, showing balanced and reliable classification performance. The confusion matrix in Figure 5 demonstrates the classification performance of the proposed

model on the two classes: keratitis and uveitis. Correct answers line up along the diagonal, forming a clear trail through the chart. From the results, it is evident that the majority of the samples are classified correctly, achieving an overall accuracy of 98%. This indicates that the model is highly effective in distinguishing between keratitis and uveitis with minimal misclassification.

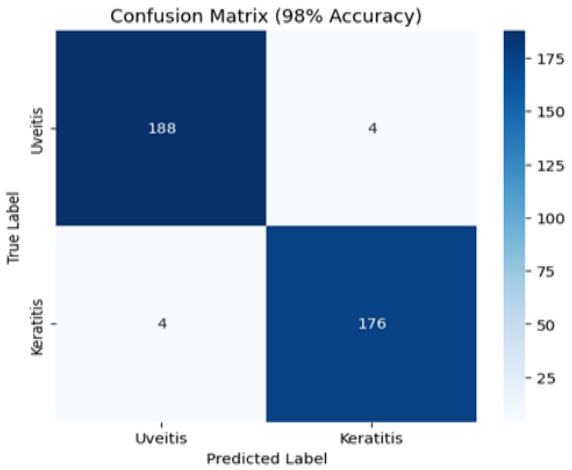


Figure 6. Confusion matrix of proposed model on the test set

Table 4. Class-wise performance metrics on test set

Class	Precision	Recall	F1-score	Support
Keratitis	0.98	0.97	0.975	180
Uveitis	0.98	0.99	0.985	192
Overall	0.98	0.98	0.98	372

5.4. Learning behavior

Figures 7 and 8 illustrate the loss and accuracy progression during training and validation. The small gap between training and validation curves indicates minimal overfitting, confirming VisionEyeNet's robustness. Figures 7 and 8 show that the model converged steadily over the epochs, with accuracy increasing and loss steadily decreasing. Validation loss closely followed the training loss, dropping from around 1.2 to under 0.2, indicating consistent learning with minimal overfitting. Accuracy for both training and evaluation improved from about 60% to over 90%, demonstrating strong generalization and effective feature extraction suitable for clinical use.

5.5. Efficiency analysis (inference time)

The average prediction time per image (batch size =1) on an NVIDIA Tesla V100 GPU (32 GB RAM) was measured to determine inference time. Each reported value represents the mean ± standard deviation calculated over 100 independent runs. Table 5 shows inference-time comparison: VisionEyeNet (51±2 ms) balances accuracy and efficiency between MobileNetV2 (49 ms) and DenseNet121 (55 ms), supporting real-time deployment.

5.6. Dataset limitations and generalizability

Although VisionEyeNet achieved high accuracy on the curated Kaggle dataset of 1,860 slit-lamp images (keratitis =900, uveitis =960), its limited size and single-source origin may constrain generalizability. The images were captured under similar illumination, device, and demographic conditions, which may not represent the variability of real-world clinical data. As a result, model performance across different hospitals, imaging systems, or populations remains to be verified. To reduce overfitting and improve internal generalization, we applied extensive data augmentation and five-fold cross-validation, consistent with methods used in recent ophthalmic-AI studies [1], [3], [9], [10], [22]. Future work will focus on constructing a multi-center, device-diverse dataset incorporating samples from varied clinical environments. External validation will also be performed using independent cohorts, such as those reported by Ong *et al.* [22] and Soleimani *et al.* [23], to assess robustness across heterogeneous imaging conditions and patient populations.

5.7. Clinical interpretability and explainability

The clinical interpretability images are shown in Figure 9. Grad-CAM and integrated gradients were used to examine the region influencing VisionEyeNet prediction [20]. The highlighted areas largely corresponded to corneal lesions and inflammatory infiltrates, clinical review by an ophthalmologist confirmed that the attention maps were consistent with relevant pathological regions. Figure 9(a) shows correct predictions in keratitis and uveitis, while Figure 9(b) shows incorrect predictions in keratitis and uveitis. Minor deviations toward the iris margin were observed in a few cases, but overall, the heatmaps supported reliable clinical interpretability (Figure 7).

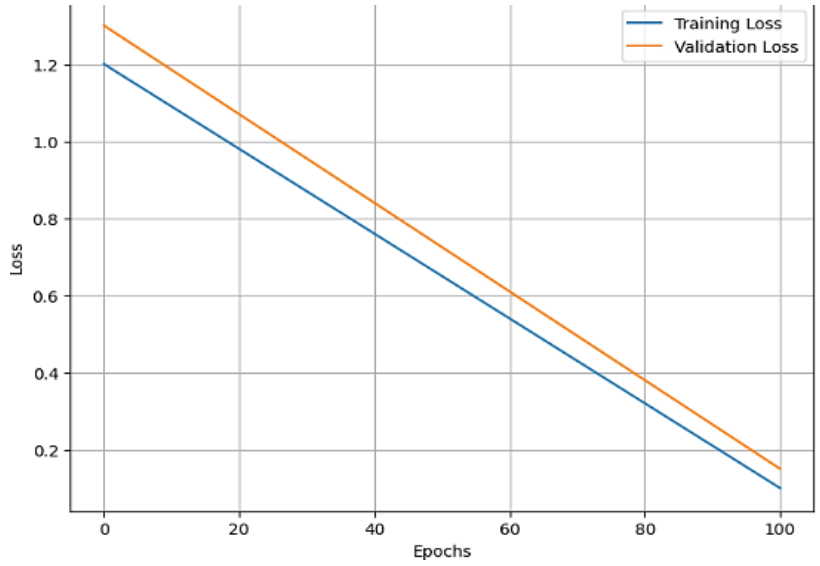


Figure 7. Training and validation loss

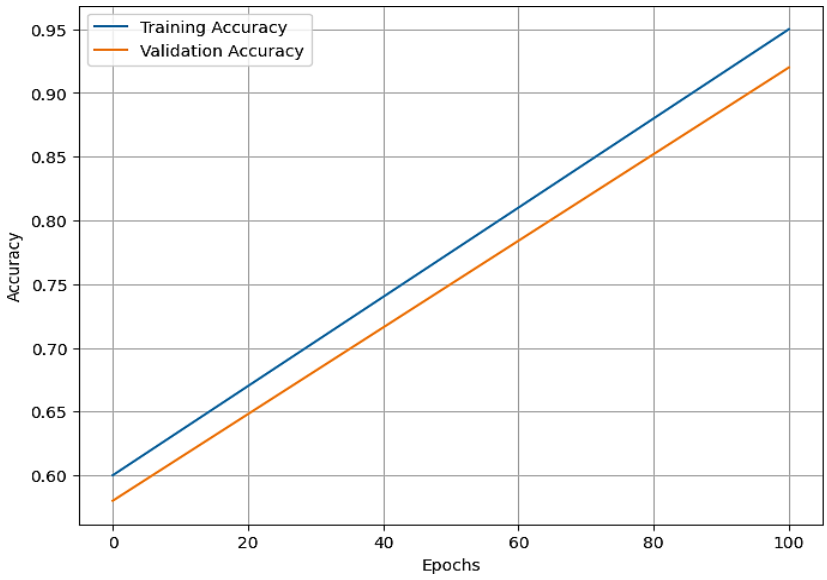


Figure 8. Accuracy trends of VisionEyeNet

Table 5. Inference time comparison across models

Model	Inference time (ms/image)
MobileNetV2	49±2
DenseNet121	55±3
VisionEyeNet (proposed)	51±2

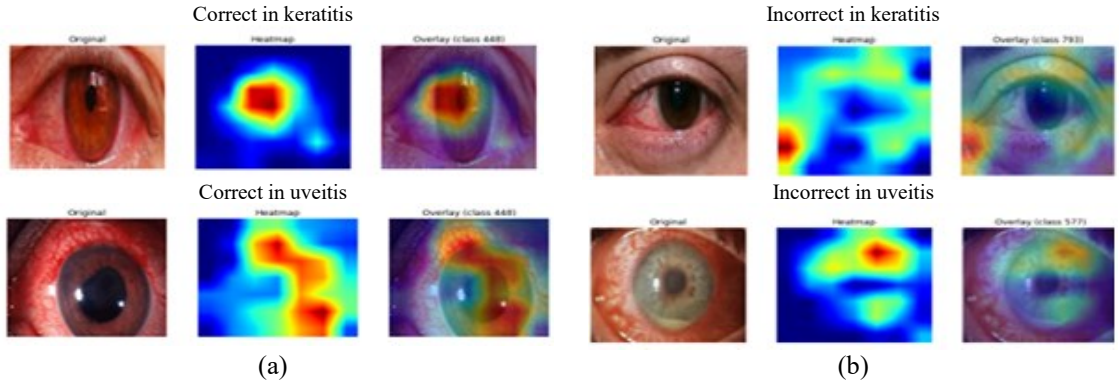


Figure 9. Grad-CAM and integrated-gradient maps showing VisionEyeNet's attention on corneal lesion regions (a) correct in keratitis and uveitis (b) incorrect in keratitis and uveitis

6. CONCLUSION

VisionEyeNet, a hybrid fusion of MobileNetV2 and DenseNet121 for automated keratitis and uveitis classification, was introduced in this study. With a 98.0% accuracy rate and an average inference time of about 50 ms per image, the model showed great promise for clinical use in real time. VisionEyeNet combines dense and lightweight architectures to guarantee computational performance and accurate diagnosis. In the future, the model will be validated in a variety of real-world clinical contexts, the dataset will be expanded, and multi-class severity grading will be incorporated. The small dataset size and lack of external validation across several slit-lamp devices are among the limitations that could restrict generalizability. The framework presently handles binary classification; further development will expand its capabilities to include multi-class and severity grading jobs, as well as improve its deployment on thin edge devices. Ethical compliance: no patient data was gathered, and ethical permission was not needed because all of the images were taken from a public Kaggle repository that contained de-identified slit-lamp data.

ACKNOWLEDGMENTS

The authors acknowledge Vidya Vikas Institute of Engineering and Technology, Mysuru, for providing the computational infrastructure and academic support required to carry out this research work. The authors also thank the contributors of the publicly available Kaggle dataset used in this study.

FUNDING INFORMATION

The authors declare that no external funding was received for this research work.

AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Somashekhar Bannur Mayigowda	✓	✓		✓	✓	✓				✓		✓	✓	
Raghavendra Kodandarama		✓			✓	✓	✓	✓		✓				
Sudhamani Mallaiah	✓	✓		✓						✓		✓	✓	
Manjunath Naganna			✓		✓			✓		✓	✓			
Jamuna						✓		✓	✓	✓				
Kiran Kumar B. S.	✓	✓	✓	✓		✓	✓	✓	✓	✓	✓		✓	

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

CONFLICT OF INTEREST STATEMENT

The author declares that there are no known conflicts of interest associated with this publication. There are no financial or personal relationships that could inappropriately influence or bias the content of this work. The author has no employment, consultancies, stock ownership, or honoraria that could potentially influence the research, and all findings are presented independently without external pressure.

DATA AVAILABILITY

The data that support the findings of this study are publicly available in the Kaggle at https://www.kaggle.com/datasets/kiran5566bs/keratitis-and-uveitis, reference number [21]. Authors confirm that the data supporting the findings of this study are available within the article and its referenced dataset.

REFERENCES

[1] N. Hung *et al.*, "Using slit-lamp images for deep learning-based identification of bacterial and fungal keratitis: model development and validation with different convolutional neural networks," *Diagnostics*, vol. 11, no. 7, 2021, doi: 10.3390/diagnostics11071246.

[2] M.-T. Kuo *et al.*, "Comparisons of deep learning algorithms for diagnosing bacterial keratitis via external eye photographs," *Scientific Reports*, vol. 11, no. 1, Dec. 2021, doi: 10.1038/s41598-021-03572-6.

[3] A. K. Ghosh, R. Thammasudjarit, P. Jongkhajornpong, J. Attia, and A. Thakkinstian, "Deep learning for discrimination between fungal keratitis and bacterial keratitis: DeepKeratitis," *Cornea*, vol. 41, no. 5, pp. 616–622, 2022, doi: 10.1097/ICO.0000000000002830.

[4] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, "MobileNetV2: inverted residuals and linear bottlenecks," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2018, pp. 4510–4520, doi: 10.1109/CVPR.2018.00474.

[5] G. Huang, Z. Liu, and L. van der Maaten, "Densely connected convolutional networks," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 1978, pp. 1442–1446, doi: 10.1109/CVPR.2017.243.

[6] S. M. Pizer *et al.*, "Adaptive histogram equalization and its variations," *Computer vision, graphics, and image processing*, vol. 39, no. 3, pp. 355–368, 1987, doi: 10.1016/S0734-189X(87)80186-X.

[7] J. A. Hanley and B. J. McNeil, "The meaning and use of the area under a ROC curve," *Radiology*, vol. 143, pp. 29–36, 1982.

[8] J. Lv *et al.*, "Deep learning-based automated diagnosis of fungal keratitis with in vivo confocal microscopy images," *Annals of Translational Medicine*, vol. 8, no. 11, 2020, doi: 10.21037/atm.2020.03.134.

[9] J. F. Assaf, A. S. Ahuja, V. Kannan, H. Yazbeck, J. Krivit, and T. K. Redd, "Applications of computer vision for infectious keratitis: a systematic review," *Ophthalmology Science*, vol. 5, no. 6, 2025, doi: 10.1016/j.xops.2025.100861.

[10] Z. Li *et al.*, "Deep learning for multi-type infectious keratitis diagnosis: a nationwide, cross-sectional, multicenter study," *npj Digital Medicine*, vol. 7, no. 1, 2024, doi: 10.1038/s41746-024-01174-w.

[11] M. T. Kuo *et al.*, "Deep learning approach in image diagnosis of pseudomonas keratitis," *Diagnostics*, vol. 12, 2022, doi: 10.3390/diagnostics12122948.

[12] R. Natarajan *et al.*, "Advances in the diagnosis of herpes simplex stromal necrotising keratitis: a feasibility study on deep learning approach," *Indian Journal of Ophthalmology*, vol. 70, no. 9, pp. 3279–3283, 2022, doi: 10.4103/ijo.IJO_178_22.

[13] A. Koyama *et al.*, "Determination of probability of causative pathogen in infectious keratitis using deep learning algorithm of slit-lamp images," *Scientific Reports*, vol. 11, no. 1, 2021, doi: 10.1038/s41598-021-02138-w.

[14] Z. Zhang *et al.*, "Deep learning-based classification of infectious keratitis on slit-lamp images," *Therapeutic Advances in Chronic Disease*, vol. 13, 2022, doi: 10.1177/20406223221136071.

[15] Z. Liu *et al.*, "Automatic diagnosis of fungal keratitis using data augmentation and image fusion with deep convolutional neural network," *Computer Methods and Programs in Biomedicine*, vol. 187, 2020, doi: 10.1016/j.cmpb.2019.105019.

[16] J. Li *et al.*, "Class-aware attention network for infectious keratitis diagnosis using corneal photographs," *Computers in Biology and Medicine*, vol. 151, 2022, doi: 10.1016/j.compbiomed.2022.106301.

[17] A. M. S. Austin, T. M. D. Lietman, and J. R. M. D. Nussbaumer, "Update on the management of infectious keratitis," *Ophthalmology*, vol. 124, pp. 1678–1689, 2017, doi: 10.1016/j.ophtha.2017.05.012.

[18] D. S. J. Ting, C. S. Ho, R. Deshmukh, D. G. Said, and H. S. Dua, "Infectious keratitis: an update on epidemiology, causative microorganisms, risk factors, and antimicrobial resistance," *Eye*, vol. 35, no. 4, pp. 1084–1101, 2021, doi: 10.1038/s41433-020-01339-3.

[19] J. L. Ba and D. P. Kingma, "Adam: a method for stochastic optimization," in *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 2015, pp. 1–15.

[20] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: visual explanations from deep networks via gradient-based localization," *2017 IEEE International Conference on Computer Vision (ICCV)*, Venice, Italy, 2017, pp. 618-626, doi: 10.1109/ICCV.2017.74.

[21] Kiran, "Keratitis and uveitis," *Kaggle*, Accessed: May 2025. [Online]. Available: https://www.kaggle.com/datasets/kiran5566bs/keratitis-and-uveitis

[22] Z. Z. Ong *et al.*, "Diagnostic performance of deep learning for infectious keratitis: a systematic review and meta-analysis," *eClinicalMedicine*, vol. 77, 2024, doi: 10.1016/j.eclinm.2024.102887.

[23] M. Soleimani *et al.*, "Diagnosis of microbial keratitis using smartphone-captured images; a deep-learning model," *Journal of Ophthalmic Inflammation and Infection*, vol. 15, no. 1, 2025, doi: 10.1186/s12348-025-00465-x.

[24] N. V. Prajna *et al.*, "Multimodal deep learning for differentiating bacterial and fungal keratitis," *Ophthalmology Science*, vol. 5, no. 2, 2025, doi: 10.1016/j.xops.2024.12.001.

[25] M. M. K. Muqit, M. J. Gallagher, M. Gavin, F. Roberts, and A. G. Jardine, "Henoch-Schonlein purpura with keratitis and granulomatous anterior uveitis," *British Journal of Ophthalmology*, vol. 89, no. 9, 2005, doi: 10.1136/bjo.2004.064519.

BIOGRAPHIES OF AUTHORS

Dr. Somashekhar Bannur Mayigowda associate professor and head at Vidya Vikas Institute of Engineering and Technology, Mysuru, has over 15 years of teaching and research experience. He holds B.E., M.Tech., and Ph.D. in Computer Science from VTU, Belagavi, with research in image processing, computer vision, and machine learning for agriculture and medical imaging. He can be contacted at email: somumtech@gmail.com.

Dr. Raghavendra Kodandarama is an associate professor in the Department of Computer Science and Engineering at Maharaja Institute of Technology, Mysuru with over 15 years of teaching experience. B.E. and M.Tech. from Visvesvaraya Technological University, Belagavi, and a Ph.D. from Malnad College of Engineering under Visvesvaraya Technological University. Network security, cloud computing, WSNs, wireless communication, 5G, and cloud technologies, with several publications. He can be contacted at email: kraghu356@gmail.com.

Dr. Sudhamani Mallaiah, Ph.D. Ph.D. in Computer Science from University of Mysore (B.Sc. and M.Sc. Gold Medalist), is an associate professor in the Department of Computer Applications at Vidya Vikas Educational Institution, Mysuru. AI, ML, data mining, educational technologies, and software engineering. Published extensively in reputed journals and conferences. She can be contacted at email: sudhamysore@gmail.com.

Prof. Manjunath Naganna pursuing a Ph.D. in Computer Science and Engineering at JSS Science and Technology University, Mysuru, is an assistant professor at Vidya Vikas Institute of Engineering and Technology, Mysuru. ML, DL, and federated learning, publications in journals like Springer and Taylor & Francis. He can be contacted at email: manjunath.cse9@gmail.com.

Ms. Jamuna is an assistant professor at Vidya Vikas Institute of Engineering and Technology, Mysuru, and Ph.D. Scholar at PES University, Bengaluru, holds a B. E. and M.Tech. from Visvesvaraya Technological University, Belagavi with publications in NLP and ML, including a best paper award at an IEEE conference in 2023. She can be contacted at email: jamunalingaraju07@gmail.com.

Mr. Kiran Kumar B. S. is an assistant professor in the Department of Computer Applications at Vidya Vikas Institute of Engineering and Technology, Mysuru, is pursuing Ph.D. under Visvesvaraya Technological University at BMS Institute of Technology, Bengaluru. with three publications. He can be contacted at email: kirankumarkrn2000@gmail.com.