GMR
The Original

A MULTI-TASK EXPLAINABLE AI FRAMEWORK FOR PAPILLEDEMA DETECTION AND SYSTEMIC DISEASE RISK PREDICTION USING RETINAL FUNDUS IMAGES

Dr. M. Srinivas[1], Prof. D. Suman[2], Modugu Sandeep[3]

[1]Assistant Professor, Department of Biomedical Engineering, University College Of Engineering (A), Osmania University, Hyderabad – 500007, Email: medisrinivas@osmania.ac.in
[2]Department of Biomedical Engineering, University College of Engineering (A), Osmania University, Hyderabad - 500007, E-Mail: dabbu_suman@osmania.ac.in
[3]Department of Biomedical Engineering, University College of Engineering (A), Osmania University, Hyderabad – 500007, Email: modugusandeep820@gmail.com

ABSTRACT
Papilledema is optic-disc swelling associated with raised intracranial pressure and requires timely recognition. This work develops a multi-task explainable artificial intelligence framework that classifies retinal fundus images while extracting vessel and optic-disc information for systemic-risk analysis. The dataset preparation pipeline converts the original Normal, Pseudopapilledema, and Papilledema organization into a binary Normal, Papilledema task and applies an 80:20 training-validation split. Six architectures—ResNet50, DenseNet201, ViT-base, ViT-CapsuleNetwork, Xception, and ConvNeXt-Tiny—are evaluated using accuracy, precision, recall, F1-score, AUC, confusion matrices, and ROC curves. Xception provides the strongest verified result with 0.996 accuracy, precision, recall, and F1-score and 1.000 AUC. The structural-analysis branch enhances the green channel, generates vessel masks and skeletons, estimates vessel diameter, detects the optic-disc edema region, derives artery and vein diameters and the artery-to-vein ratio, and demonstrates EfficientUNet optic-disc segmentation. Grad-CAM explains the classifier, while a Flask and SQLite interface supports authenticated image upload and integrated result presentation. The system is intended as a research decision-support prototype rather than an autonomous clinical diagnosis tool.

KEYWORDS: Papilledema, retinal fundus images, deep learning, Xception, Vision Transformer, vessel segmentation, oculomics, Grad-CAM, systemic disease risk, Flask.

1. INTRODUCTION

The interplay of genetics, skeletal-muscle properties, body fat composition, training history, nutrition and recovery are Papilledema refers to swelling of the optic disc caused by raised intracranial pressure. Because the optic nerve head is visible through fundus photography, retinal images provide a non-invasive means of observing disc-margin blurring, vascular obscuration, hyperemia, and surrounding retinal changes. Delayed recognition can contribute to visual and neurological harm, yet manual assessment may be difficult when appearances overlap with pseudopapilledema, optic-disc drusen, tilted discs, and other optic nerve abnormalities [1],[3].

Fundus photography captures the optic disc, retinal vessels, macula, and surrounding tissue in a single image. This allows one image to support several computational tasks: binary disease classification, vessel and optic-disc segmentation, extraction of vascular measurements, disease-risk estimation, and visual explanation. The availability of such complementary outputs is important because a classification label alone does not expose the retinal evidence that influenced the decision [4],[7].

The proposed PAPILLENS framework compares six deep-learning architectures under a common training pipeline. The strongest classifier is integrated with image enhancement, vessel-mask extraction, skeletonization, optic-disc analysis, exploratory oculomics measurements, Grad-CAM explanation, and a Flask-based interface. The aim is to provide a transparent end-to-end research prototype while clearly separating verified model outputs from clinical claims that require external validation.

Related Work

Chang et al. [1] developed a DenseNet-based tri-branch system to distinguish pediatric true papilledema from pseudopapilledema using multicentre fundus photographs. Branco et al. [2] applied machine learning to classify and quantify papilloedema severity changes using clinical-trial images. Lin et al. [3] reported the BONSAI deep-learning system for pediatric papilledema identification. These studies demonstrate that optic-disc photographs contain learnable patterns, but their targets, populations, and validation protocols differ from the present binary Normal,Papilledema task.

Retinal vascular analysis has also received considerable attention. Liu et al. [4] used heterogeneous-feature cross-attention for vessel segmentation and hypertensive-retinopathy quantification. Almarri et al. [5] investigated GAN-oriented vessel segmentation, while Liu et al. [6] combined double skip connections and deep supervision for fine-vessel extraction. Faria

et al. [7] linked explainable fundus classification with modern vessel-segmentation approaches. These studies motivate the vessel mask, skeleton, diameter, and systemic-risk modules used in the present framework.

Recent reviews emphasize that internal accuracy alone is insufficient for clinical adoption. Łajczak et al. [8], Rambabu et al. [9], and Karkhur et al. [11] identified heterogeneity in datasets, reference standards, specificity, external validation, and reporting quality. Li et al. [10] further showed that general vision-language models do not consistently match specialised optic-disc classifiers. Consequently, the present work reports its measured notebook results while retaining a decision-support disclaimer.

Yao et al. [12] proposed RetinalVasNet for microvascular detection, and Li et al. [13] reviewed fundus-based prediction of hypertension-related diseases. Li and Tao [14] introduced a vessel-guided multi-task framework combining vessel segmentation, disease classification, and Grad-CAM, while Hailu et al. [15] compared transfer-learning architectures for retinal disease detection. Earlier work by Milea et al. [16] and Biousse et al. [17] established the potential of deep learning for optic-disc classification. The remaining gap is an integrated application that combines comparative classification, retinal-structure analysis, oculomics, explainability, and web deployment.

Materials and Methods

The system contains an offline model-development branch and an online application branch. Offline processing consolidates the dataset, creates the training-validation split, augments fundus images, trains six classifiers, evaluates their metrics, and stores the best weights. Online processing authenticates the user, accepts a retinal image, executes classification and structural analysis, estimates systemic-risk indicators, generates Grad-CAM, and renders all outputs in a browser.

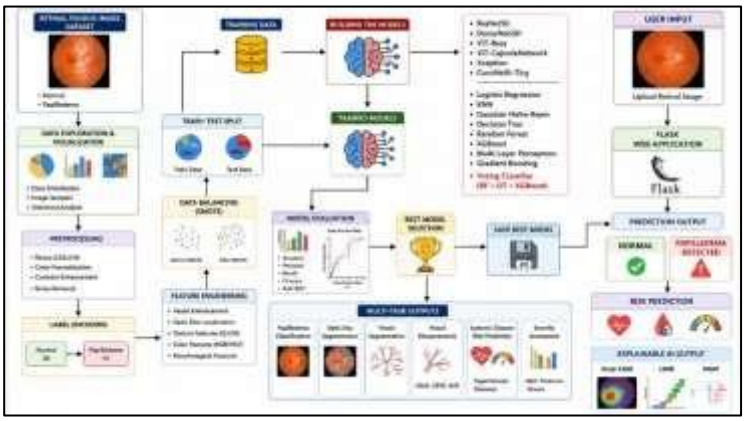

Fig. 1. System Architecture

A) Dataset Collection

The project uses the public Identification of Pseudo Papilledema dataset [26]. The original source contains Normal, Pseudopapilledema, and Papilledema categories. During project preparation, 295 pseudopapilledema images were copied into the Papilledema directory using a psp_ filename prefix, 295 original papilledema images were renamed using a p_ prefix, and the separate pseudopapilledema folder was removed. The resulting binary dataset contains 1,369 images: 779 Normal and 590 Papilledema-class images. An 80:20 split with seed 42 produces 623 Normal and 472 Papilledema training images and 156 Normal and 118 Papilledema validation images.

Dataset link: Kaggle – Identification of Pseudo Papilledema

https://www.kaggle.com/datasets/shashwatwork/identification-of-pseudopapilledema/data

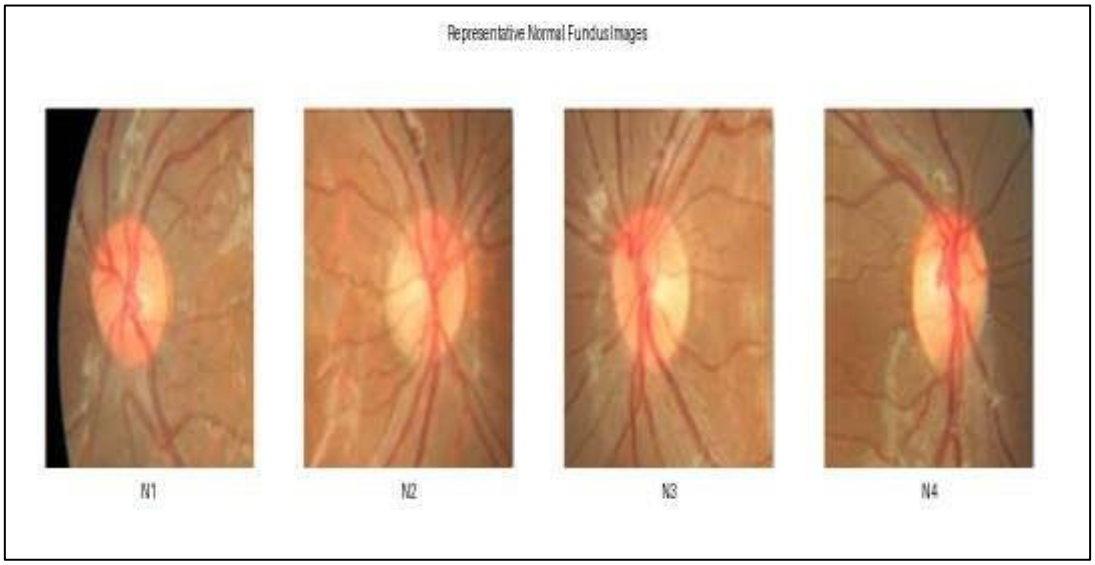

Fig. 2. Representative Normal retinal fundus images.

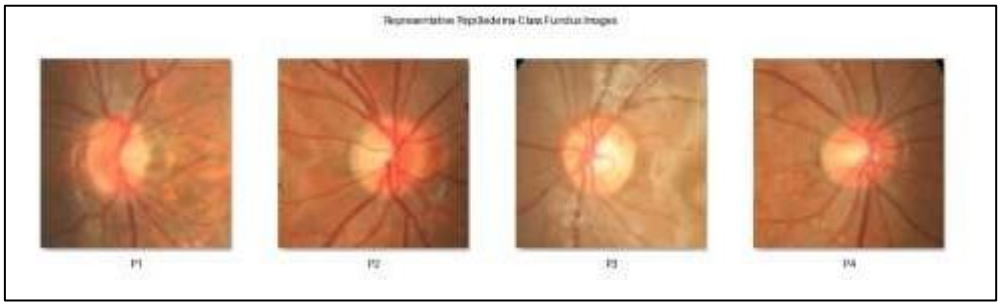

Fig. 3. Representative Papilledema-class retinal fundus images.

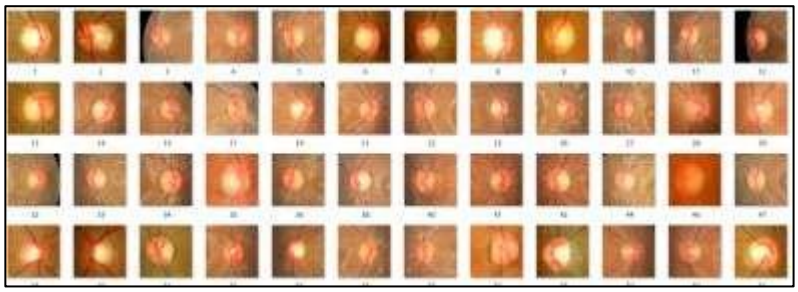

Fig. 4. Overview of Identification of Pseudo papilledema dataset

B) Visualization

Dataset visualization verifies class distribution and image quality before training. The project displays training and validation counts, representative samples from both classes, confusion matrices, ROC curves, training and validation accuracy and loss, and consolidated model-comparison graphs. These views reveal the moderate class imbalance, the visual variability of optic-disc appearances, and the model-specific error patterns that are not visible from accuracy alone. The sample grids also show substantial variation in illumination, field of view, optic-disc scale, vessel visibility and background pigmentation. Reviewing these differences before training is important because a model may otherwise learn camera-specific borders, brightness patterns or acquisition artefacts instead of medically relevant optic-disc characteristics. The visual inspection therefore complements numerical class counts and supports the use of augmentation during model development.

The binary preparation used in this project groups the source Pseudopapilledema images with the Papilledema class. This produces a practical two-class experiment for Normal-versus-Papilledema screening, but it also changes the meaning of the positive class. The dataset figures are consequently presented to document the prepared training task clearly and to avoid implying that the deployed classifier independently differentiates true papilledema from pseudopapilledema.

C) Pre-Processing

Training images are transformed using a 384-pixel random resized crop, horizontal flipping, rotation, colour jitter, tensor conversion, ImageNet normalization, and random erasing. These operations increase variation in orientation, illumination, colour, and local content. The fixed validation subset is kept separate so that the reported metrics represent unseen images from the same prepared dataset. The segmentation branch enhances the green channel, applies local contrast enhancement and smoothing, and uses thresholding and morphology to obtain vessel and optic-disc candidates.

D) Training and Testing

The six classification architectures are trained for ten epochs using a common training function. For every epoch, the notebook records training loss, training accuracy, validation loss, and validation accuracy, and stores updated weights whenever the validation result improves. Final predictions and class probabilities are collected from the validation loader, after which weighted precision, weighted recall, weighted F1-score, AUC, confusion matrices, classification reports, and ROC curves are generated. This shared procedure enables evidence-based selection of the deployment model rather than assuming that a more complex architecture will perform better.

Using the same split, transformation pipeline and evaluation procedure provides a controlled comparison across residual, densely connected, transformer, capsule-hybrid, depthwise-separable and modern convolutional architectures. The training curves are reviewed together with validation metrics so that rapid memorization, unstable convergence or persistent underfitting can be identified. The saved deployment model is selected from verified validation behaviour, while the remaining models are retained as comparative evidence rather than being combined into an unsupported ensemble.

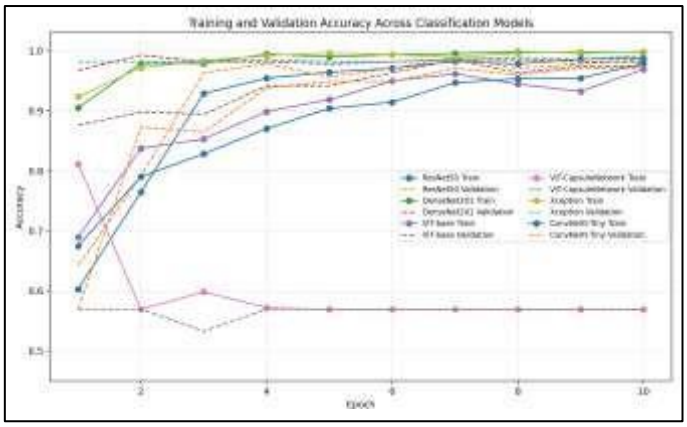

Fig. 5. Training and validation accuracy across the six classification models.

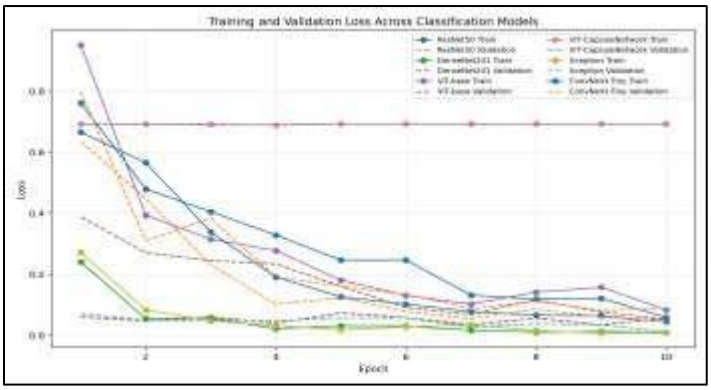

Fig. 6. Training and validation loss across the six classification models.

E) Deep-Learning Classification Algorithms

ResNet50 [18] uses residual blocks in which a learned residual mapping is added to the block input. The skip connection stabilizes gradient propagation through a deep feature extractor and supports learning of optic-disc and peripapillary patterns. The final fully connected layer is replaced by a two-class output layer.

DenseNet201 [19] connects each layer to later layers within a dense block. This design encourages feature reuse and combines low-level edge and vessel information with higher-level disc representations. ViT-base [20] divides the 384-pixel image into patches, projects the patches into tokens, adds positional information, and applies global self-attention. The ViT-CapsuleNetwork extends transformer features through capsule-style grouping and routing [23]. In the reported experiment, however, it remains close to the majority-class baseline, illustrating that the hybrid requires stronger optimization or more training data. Xception [21] replaces standard convolutions with depthwise spatial filtering followed by pointwise channel mixing. ConvNeXt-Tiny [22] retains convolutional inductive bias while adopting modern block design, large kernels, normalization, and training choices inspired by transformers.

F) Vessel, Optic-Disc, and Oculomics Analysis

The runtime analysis begins with green-channel extraction because retinal vessels generally have stronger contrast in that channel. Contrast enhancement, Gaussian smoothing, thresholding, and morphological opening and closing generate a binary vessel mask. Skeletonization reduces the mask to one-pixel-wide centre lines, and branch-point removal supports cleaner segment-wise measurement. The distance transform assigns each vessel pixel its distance to the nearest background pixel; twice the distance sampled on the skeleton estimates local diameter.

Valid diameters are summarized through minimum, maximum, median, mean, and standard deviation. The notebook treats the thinner lower 30% as artery candidates and the thicker upper 30% as vein candidates, then divides their mean diameters to obtain an exploratory artery-to-vein ratio. A separate optic-disc routine enhances the image, applies thresholding and elliptical morphology, and retains the largest connected region to estimate edema spread. EfficientUNet provides an additional learned optic-disc mask, but no ground-truth Dice or IoU scores are supplied.

G) Explainability and Flask Deployment

Grad-CAM [24] weights convolutional feature maps using gradients for the predicted class and generates a heatmap showing influential retinal regions. The Flask application validates registration fields, stores account records in SQLite, authenticates users, accepts image uploads, and invokes preprocessing, classification, vessel analysis, disease-risk estimation, severity assessment, and Grad-CAM in sequence. The result page presents the predicted class, confidence, severity, masks, vascular measurements, estimated hypertension and diabetes risk, and the explanation heatmap. Exact clinical thresholds for the internal risk rules were not supplied and are therefore not reproduced.

The application-level integration is deliberately modular. Classification, vessel analysis, optic-disc analysis, risk estimation, severity assessment and explanation are executed as connected components, allowing each output to be inspected independently. This design is useful for research because an incorrect classification can be examined alongside the vessel mask, optic-disc region and heatmap instead of being accepted as an unexplained final label.

The web interface should be interpreted as a presentation and decision-support layer. It does not replace neurological examination, neuroimaging, lumbar puncture, optical coherence tomography or specialist review where clinically indicated. Confidence, severity and systemic-risk values are displayed to organize the prototype outputs, but they require calibration on independent clinical datasets before they can be interpreted as medical probabilities or treatment recommendations.

Experimental Results

Accuracy measures the proportion of correctly classified samples. Precision measures the fraction of predicted Papilledema samples that are correct, recall measures the fraction of actual Papilledema samples detected, and F1-score balances precision and recall. The formulas are written as follows:

Table 1. Classification Performance of Evaluated Models

Model	Acc.	Prec.	Recall	F1	AUC
ResNet50	0.982	0.982	0.982	0.982	0.998
DenseNet201	0.978	0.978	0.978	0.978	0.999
ViT-base	0.985	0.986	0.985	0.985	0.997
ViT-Caps	0.569	0.324	0.569	0.413	0.898
Xception	0.996	0.996	0.996	0.996	1.000
ConvNeXt-T	0.971	0.971	0.971	0.971	0.997

Xception achieves the strongest verified performance with 0.996 accuracy, precision, recall, and F1-score and 1.000 AUC. Its confusion matrix contains 155 correctly classified Normal images, 118 correctly classified Papilledema images, one Normal false positive, and no Papilledema false negatives. ViT-base correctly classifies 153 Normal and 117 Papilledema images, while ResNet50 identifies 155 Normal and 114 Papilledema images. DenseNet201 produces three errors in each class, and ConvNeXt-Tiny correctly classifies 151 Normal and 115 Papilledema images.

The ViT-CapsuleNetwork behaves differently from the other models. Its weighted precision is 0.324 and F1-score is 0.413, even though the reported AUC is 0.898. The gap between AUC and thresholded classification metrics shows that ranking ability alone does not guarantee a useful operating point. The comparative graph therefore reports accuracy, precision, recall, F1-score, and AUC together instead of presenting accuracy as the only criterion.

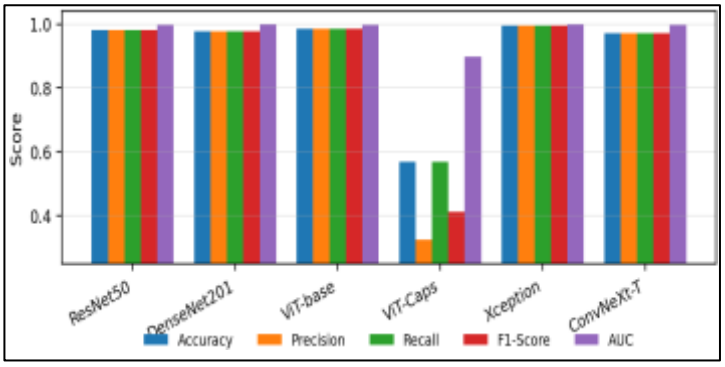

Fig. 7. Accuracy, precision, recall, F1-score, and AUC comparison.

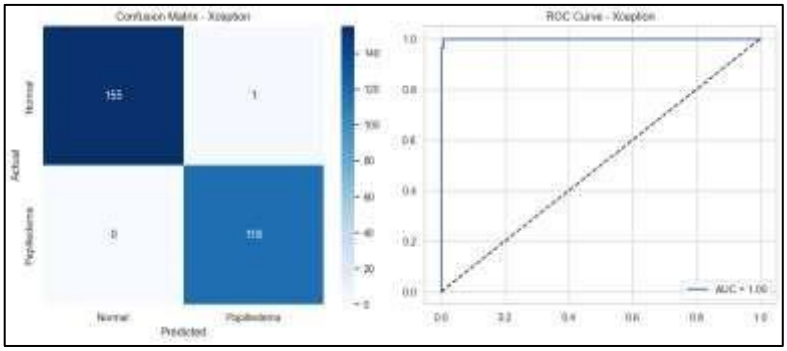

Fig. 8. Xception confusion matrix and ROC curve.

The segmentation notebook provides qualitative structural analysis and verified measurements for an illustrated Papilledema image. The average vessel diameter is 7.25 pixels, disc-edema spread is 16.92%, average artery diameter is 2.11 pixels, average vein diameter is 8.17 pixels, and AVR is 0.258. The hemorrhage-candidate stage reports 74 connected components with a combined candidate area of 23,607 pixels. These values are exploratory image-derived measurements and are not substitutes for validated clinical biomarkers.

Table 2. Selected Segmentation-Derived Oculomics Measurements

Biomarker	Verified value
Average vessel diameter	7.25 pixels
Disc-edema spread	16.92%
Average artery diameter	2.11 pixels
Average vein diameter	8.17 pixels
Artery-to-vein ratio	0.258

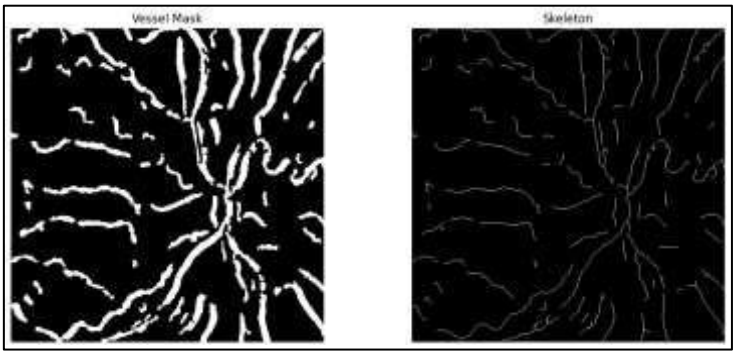

Fig. 9. Retinal vessel mask and skeletonized vascular network.

The binary vessel mask preserves the main retinal vascular tree, while skeletonization converts the segmented regions into centre lines for topology and width analysis. The notebook also removes branch points through local-neighbour counting so that vessel segments can be measured with fewer junction-related distortions. These images provide visual evidence that the structural-analysis stage is operating on the vascular network rather than only returning a classification label.

Fig. 10. Oculomics parameters and systemic disease-risk presentation.

The oculomics result page reports vessel diameter, optic-disc edema spread, artery diameter, vein diameter, and AVR together with estimated hypertension and diabetes risk. The displayed risk values are outputs of the connected application module; because the project files do not provide clinically calibrated thresholds or an external risk dataset, these estimates are presented as research indicators rather than diagnostic probabilities.

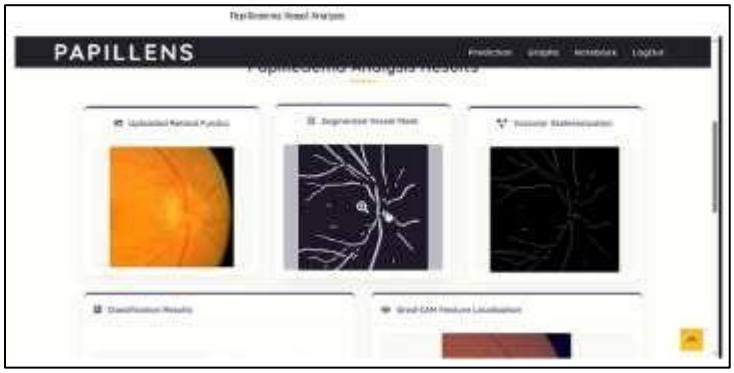

Fig. 11. PAPILLENS retinal vessel-mask and vascular-skeleton outputs.

The vessel-analysis view places the uploaded fundus image beside the segmented vessel mask and the one-pixel-wide vascular skeleton. This arrangement allows the user to verify that the structural branch has isolated the main retinal vascular tree before diameter and artery-to-vein measurements are interpreted.

Fig. 12. PAPILLENS classification result and Grad-CAM feature-localization output.

The classification view reports the predicted class, confidence and severity together with the Grad-CAM heatmap. Displaying the explanation next to the numerical prediction helps determine whether the classifier is concentrating on the optic-disc and peripapillary region rather than unrelated background structures.

The frontend demonstrates the complete request-response workflow. A user signs in, selects a fundus image, and receives a combined page containing the uploaded image, segmented vessel mask, vascular skeleton, predicted class, confidence, severity, Grad-CAM heatmap, oculomics parameters, and systemic-risk profiles. The result layout allows classification and structural evidence to be reviewed together instead of being distributed across separate notebooks.

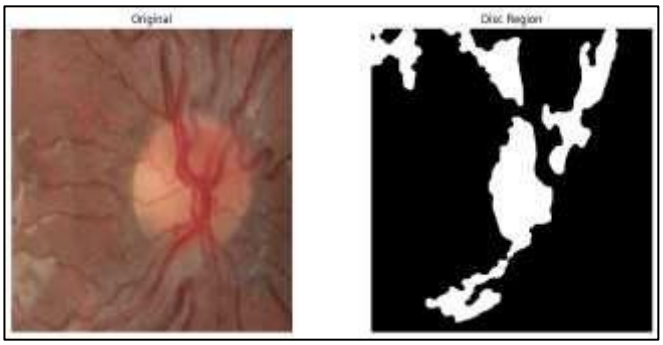

Fig. 13. Original fundus image and detected optic-disc region.

The optic-disc analysis compares the original fundus image with the binary region retained after green-channel enhancement, smoothing, Otsu thresholding, elliptical morphology, and connected-component selection. The selected region supports the reported disc-area and edema-spread measurements, while the EfficientUNet output provides a separate learned mask for qualitative comparison.

The heuristic disc-region output and the EfficientUNet demonstration serve different purposes. The connected-component procedure provides an interpretable image-processing path for calculating an approximate edema-spread value, whereas the learned segmentation model illustrates how an encoder-decoder network can delineate the disc region. Because expert reference masks are not included in the supplied experiment, the two outputs are treated as qualitative structural evidence and are not assigned unsupported Dice or intersection-over-union scores.

The classification results are derived from an internal 80:20 split of 1,369 prepared images and may not represent performance on new cameras, hospitals, age groups, or disease severities. The original dataset contains a separate Pseudopapilledema category, but the current project consolidates it into the Papilledema class; consequently, the trained model does not solve the clinically important three-way differential diagnosis. The segmentation examples are visually demonstrated without paired ground-truth masks or reported Dice and IoU values. Likewise, systemic-risk and severity outputs require external calibration, and the current SQLite implementation stores passwords without production-grade hashing. These boundaries must remain explicit when interpreting the prototype.

A further validation concern is the possibility of source-related similarity between images. Filename management preserves the origin of consolidated samples, but independent patient-level and device-level partitioning would be preferable when metadata are available. External evaluation should report sensitivity and specificity at clinically selected thresholds, calibration error, confidence intervals and subgroup performance rather than relying only on a single internal accuracy value.

Structural measurements are expressed in pixels because a physical retinal scale was not supplied. Vessel diameter, edema spread, artery and vein estimates and AVR are therefore useful for comparing the demonstrated image-processing stages, but they cannot be interpreted directly as standardized clinical measurements. Future validation should use calibrated image resolution, expert vessel labels, optic-disc boundaries and clinically confirmed systemic outcomes.

CONCLUSION

This study presents a multi-task explainable framework for Papilledema detection and retinal structure analysis. Six deep-learning classifiers were compared under a common binary workflow, and Xception produced the strongest verified validation result. The project extends beyond classification by generating vessel masks, vascular skeletons, optic-disc analysis, vessel-diameter statistics, artery and vein estimates, AVR, hemorrhage candidates, severity information,

systemic-risk estimates, and Grad-CAM explanations. Flask and SQLite integrate these outputs into an authenticated browser interface. The system demonstrates the value of combining prediction with visible retinal evidence, but it remains a research prototype and should not be interpreted as an autonomous clinical diagnostic system.

Future work should validate the classifier on multicentre and device-diverse datasets, preserve Pseudopapilledema as a separate diagnostic class, evaluate subgroup performance and probability calibration, and measure optic-disc and vessel segmentation with expert ground truth. The exploratory artery-vein separation and systemic-risk rules should be replaced by clinically validated models. Security should be strengthened through password hashing, protected sessions, upload validation, and encrypted communication. Prospective comparison with neuro-ophthalmologists and integration with OCT, clinical symptoms, and longitudinal follow-up would further support safe translation.

The next technical stage should convert the connected application modules into a genuinely joint multi-task learning framework. A shared retinal encoder could support papilledema classification, vessel segmentation and optic-disc segmentation, while task-specific heads preserve the distinct objectives. Learned vessel and disc masks could then act as spatial priors for classification, and uncertainty estimates could flag low-confidence images for specialist review instead of forcing a definitive prediction.

For deployment research, the application should also record model version, preprocessing configuration, inference time and audit information for each result. Secure password storage, session protection, file-type validation, size limits, de-identification and controlled retention of uploaded images are necessary before testing with sensitive clinical data. These improvements would strengthen reproducibility and safety without changing the framework's intended role as supervised decision support.

REFERENCES

[1] M. Y. Chang et al., "Artificial Intelligence to Differentiate Pediatric Pseudopapilledema and True Papilledema on Fundus Photographs," Ophthalmology Science, vol. 4, no. 4, p. 100496, 2024, doi: 10.1016/j.xops.2024.100496.

[2] J. Branco et al., "Classifying and Quantifying Changes in Papilloedema Using Machine Learning," BMJ Neurology Open, vol. 6, no. 1, e000503, 2024, doi: 10.1136/bmjno-2023-000503.

[3] M. Y. Lin et al., "The BONSAI (Brain and Optic Nerve Study with Artificial Intelligence) Deep Learning System Can Accurately Identify Pediatric Papilledema on Standard Ocular Fundus Photographs," Journal of AAPOS, vol. 28, no. 1, p. 103803, 2024.

[4] X. Liu, H. Tan, W. Wang, and Z. Chen, "Deep Learning Based Retinal Vessel Segmentation and Hypertensive Retinopathy Quantification Using Heterogeneous Features Cross-Attention Neural Network," Frontiers in Medicine, vol. 11, p. 1377479, 2024, doi: 10.3389/fmed.2024.1377479.

[5] B. Almarri et al., "Redefining Retinal Vessel Segmentation: Empowering Advanced Fundus Image Analysis with the Potential of GANs," Frontiers in Medicine, vol. 11, p. 1470941, 2024, doi: 10.3389/fmed.2024.1470941.

[6] Q. Liu et al., "A Fundus Vessel Segmentation Method Based on Double Skip Connections Combined with Deep Supervision," Frontiers in Cell and Developmental Biology, vol. 12, p. 1477819, 2024, doi: 10.3389/fcell.2024.1477819.

[7] F. T. J. Faria, M. B. Moin, P. Debnath, A. I. Fahim, and F. M. Shah, "Explainable Convolutional Neural Networks for Retinal Fundus Classification and Cutting-Edge Segmentation Models for Retinal Blood Vessels from Fundus Images," arXiv:2405.07338, 2024.

[8] P. M. Łajczak, S. Sirek, and D. Wyględowska-Promieńska, "Unveiling AI's Role in Papilledema Diagnosis from Fundus Images: A Systematic Review with Diagnostic Test Accuracy Meta-Analysis and Comparison of Human Expert Performance," Computers in Biology and Medicine, vol. 184, p. 109350, 2025, doi: 10.1016/j.compbiomed.2024.109350.

[9] L. Rambabu et al., "Detecting Papilloedema as a Marker of Raised Intracranial Pressure Using Artificial Intelligence: A Systematic Review," PLOS Digital Health, vol. 4, no. 9, e0000783, 2025, doi: 10.1371/journal.pdig.0000783.

[10] K. Z. Li, T. T. Nguyen, and H. E. Moss, "Performance of Vision Language Models for Optic Disc Swelling Identification on Fundus Photographs," Frontiers in Digital Health, vol. 7, p. 1660887, 2025, doi: 10.3389/fdgth.2025.1660887.

[11] S. Karkhur et al., "Diagnostic Accuracy of Artificial Intelligence for the Detection of Papilledema on Fundus Images: A Systematic Review and Meta-Analysis," Cureus, vol. 17, no. 12, e99135, 2025, doi: 10.7759/cureus.99135.

[12] Z. Yao et al., "RetinalVasNet: A Deep Learning Approach for Robust Retinal Microvasculature Detection," Frontiers in Molecular Biosciences, vol. 12, p. 1562608, 2025, doi: 10.3389/fmolb.2025.1562608.

[13] P. Li et al., "Research Progress in Deep Learning-Based Fundus Image Analysis for the Diagnosis and Prediction of Hypertension-Related Diseases," Frontiers in Cell and Developmental Biology, vol. 13, p. 1608994, 2025, doi: 10.3389/fcell.2025.1608994.

[14] Q. Li and L. Tao, "A Vessel-Guided Multi-Task Deep Learning Framework with Visual Interpretability for Simultaneous Retinal Vessel Segmentation and Multi-Disease Classification from Fundus Images," Frontiers in Medicine, vol. 13, p. 1799745, 2026, doi: 10.3389/fmed.2026.1799745.

[15] A. A. Hailu et al., "Early Retinal Disease Detection from Fundus Images Using Deep Neural Networks," Scientific Reports, 2026, doi: 10.1038/s41598-026-42059-0.

[16] D. Milea et al., "Artificial Intelligence to Detect Papilledema from Ocular Fundus Photographs," New England Journal of Medicine, vol. 382, pp. 1687-1695, 2020, doi: 10.1056/NEJMoa1917130.

[17] V. Biousse et al., "Optic Disc Classification by Deep Learning versus Expert Neuro-Ophthalmologists," Annals of Neurology, vol. 88, pp. 785-795, 2020.

[18] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770-778.
[19] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, “Densely Connected Convolutional Networks,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 4700-4708.
[20] A. Dosovitskiy et al., “An Image Is Worth 16 × 16 Words: Transformers for Image Recognition at Scale,” in International Conference on Learning Representations, 2021.
[21] F. Chollet, “Xception: Deep Learning with Depthwise Separable Convolutions,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 1251-1258.
[22] Z. Liu et al., “A ConvNet for the 2020s,” in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 11976-11986.
[23] S. Sabour, N. Frosst, and G. E. Hinton, “Dynamic Routing Between Capsules,” in Advances in Neural Information Processing Systems, 2017.
[24] R. R. Selvaraju et al., “Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization,” in Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 618-626.
[25] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” in Medical Image Computing and Computer-Assisted Intervention, 2015, pp. 234-241.
[26] S. Work, “Identification of Pseudo Papilledema,” Kaggle Dataset. [Online]. Available: https://www.kaggle.com/datasets/shashwatwork/identification-of-pseudopapilledema/data
[27] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” in Advances in Neural Information Processing Systems, vol. 25, 2012, pp. 1097-1105.
[28] Z. Zhong, L. Zheng, G. Kang, S. Li, and Y. Yang, “Random Erasing Data Augmentation,” in Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, no. 7, 2020, pp. 13001-13008.
[29] K. Zuiderveld, “Contrast Limited Adaptive Histogram Equalization,” in Graphics Gems IV, P. S. Heckbert, Ed. San Diego, CA, USA: Academic Press, 1994, pp. 474-485.
[30] M. Tan and Q. V. Le, “EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks,” in Proceedings of the 36th International Conference on Machine Learning, 2019, pp. 6105-6114.
Submission note: The DOI remains to be assigned by the journal. The scientific content, figures, tables, results, and references are retained from the supplied manuscript.