RESEARCH PAPER

Enhancing Eye Disease Classification with a Hybrid Deep Learning and Machine Learning Approach

S. Sangeetha[1], Dr. R. Sujatha[2]

[1]Research Scholar, Department of Computer Science, PSG College of Arts & Science, Coimbatore - 641014, Tamil Nadu, India

[2]Assistant Professor, Department of Information Technology, PSG College of Arts & Science, Coimbatore – 641014, Tamil Nadu, India

ABSTRACT

The suggested research will label the increasing universal challenge of eye diseases that are anticipated to remain among the major causes of preventable vision loss. In this study, a hybrid DL and ML architecture will be evolved by combining a CNN with SVM and LR classifiers for multi-level eye disease prediction from retinal pictures. The model will be prepared and estimated on a four-class dataset consisting of Cataract, Diabetic Retinopathy, Glaucoma, and Normal fundus images, each pre-analysed using normalization and augmentation methods. The CNN will be utilized as a feature extractor, while SVM and LR will function on the trained deep features to give comparative and supporting classification abilities. The research will methodically relate the predictive performance, processing time, and strength of the independent CNN, SVM, and LR models, as well as the hybrid CNN-SVM and CNN-LR configurations. The predicted results will specify that the CNN-based approach will attain a higher level of accuracy and generality, while the hybrid framework will provide extra clarity and elasticity for arrangement in CDSS. This study will be designed to provide an operationally stringent and clinically connected architecture for the timely, mechanical testing of ordinary retina-based eye disorders.

Keywords: Eye disease prediction; Fundus images; Convolutional Neural Network; Support Vector Machine; Logistic Regression; Hybrid deep learning; multi-class classification.

How to cite this article: Sangeetha S, Sujatha R. Enhancing Eye Disease Classification with a Hybrid Deep Learning and Machine Learning Approach. Int J Drug Deliv Technol. 2026;16(53s): 929-937. DOI: 10.25258/ijddt.16.53s.148

Source of support: Nil.

Conflict of interest: None.

.

1. Introduction

Together, retinal diseases and other conditions that weaken vision will remain among the leading causes of preventable vision loss globally, representing a significant public health concern. The combination of ageing populations, urbanization, and lifestyle-related risk factors like diabetes and hypertension is expected to increase the incidence of conditions such as cataract, diabetic retinopathy, and glaucoma in the coming decades. There is a growing burden of vision impairment that extends beyond ocular diseases, particularly in lower-resource settings where inequitable access to specialist ophthalmic care further compounds the problem: there is an urgent need for scalable, low-cost solutions that can provide accurate screening outside of traditional tertiary-care facilities. Within this context, automated analysis of retinal fundus images will become an essential component for future community-level screening and early detection programs. Conventional ocular assaying will still mainly depend on a professional visual inspection of fundus images using ophthalmoscopy or slit-lamp biomicroscopy, which continues to be substantially subjective, time-consuming, and significantly limited by inter-observer variability and work load pressures. These limitations will be particularly apparent in large-scale screening programs, where high patient volumes will require rapid and reproducible as well as reliable diagnostic support. Computer-aided diagnosis (CAD) systems, which employ image processing and machine learning algorithms for automated identification of cancerous lesions, being promising tools to supplement human expertise, are expected to alleviate the burden of screening. Even then, most existing CAD systems will remain limited to single-disease detection, narrow datasets, or limited generalizability across diverse imaging conditions. Deep learning, especially Convolutional Neural Networks (CNNs), brings significant disruption to the field of medical image analysis via end-to-end feature extraction from pixels. Depending on the choice of CNN, established architectures will be able to diagnose and classify retinal diseases like diabetic retinopathy, cataract, and other fundus anomalies through datasets such as APTOS & STARE. [6][11][13]

Research in lightweight and optimized CNN models will hence aim to improve the accuracy while keeping it computationally inexpensive, thus making deployment on-device or at near real-time levels more manageable in clinical and mobile settings. Simultaneously, advanced architectures or more specialized architectures will be explored to enhance robustness and multi-label classification of

retinal diseases analysis [22] (such as EfficientNet variants, hybrid or adaptive CNN designs). In conjunction with deep learning, classical machine learning algorithms like Support Vector Machines (SVM) and Logistic Regression (LR) are still likely to be relevant, at least in certain contexts, both for their simplistic approach as well as interpretability over relatively high-performing models, particularly in good feature representations. These will usually be applied to hand-crafted or shallow-learned features (for example, diabetic retinopathy screening, early cataract detection) in the ophthalmic domain. In the second part of the research, hybrid methods that apply CNN-based deep feature extraction followed by classical graphical classifiers are explored to preserve the rich imprint left on data by deep networks while providing interpretable and flexible decisions. Although existing hybrid CNN–SVM frameworks will be promising for disease-specific tasks such as detecting cataracts, they generally do not provide all-encompassing multi-class fundus disease prediction over various conditions at one time. In this detail, the current paper aims to present a hybrid CNN–SVM–Logistic Regression framework for multi-class prediction of eye disease from fundus images (FIs) over four descripted and clinically significant classes: Cataract, Diabetic Retinopathy, Glaucoma, and Normal retina.[12] We will design a custom CNN architecture to learn hierarchical representations from the pre-processed fundus images, and will use deep feature maps extracted from an intermediate convolutional layer to train SVM and LR classifiers for comparative evaluation. Through systematic evaluation of accuracy, robustness, and inference time across the CNN, SVM, and LR models, this study seeks to fill existing gaps regarding unified multi-class diagnosis implementation (early vs advanced stage; lung-lobar location), explicit comparison of deep versus classical classifiers on common feature spaces, and practical deployment considerations for real-time screening. Thus, the proposed framework will be presented as both methodologically sound and clinically applicable step towards future retinal disease computer-assisted screening systems [4].

2. Literature Review

Noor et al. (2026) is highly relevant, as it specifically fuses explainable AI with the classification of retinal disease, matching today's interests for trustworthy medical AI. Its focus on "XAI-driven" design indicates that the study goes beyond simple accuracy and accounts for interpretability towards diagnostic support. That makes it helpful for putting this trend from black-box prediction to clinically meaningful explanation into context. However, quite apart from its comparatively sparseness of currently-feasible bibliographic details among some other papers in your list, it suggests that it stands to be included on the basis of novelty and thematic relevance.

Sabiri, Y., &Houmaidi, A. (2025) provide considerable research, as it pairs retinal disease detection with explainability, which is only becoming more critical for clinical trust and decision support. The study focuses on 10 retinal conditions, and reports an overall accuracy of around 0.92 with strong macro-average performance from a large retinal dataset of 21,577 images. Its main contribution goes beyond classification to provide interpretable outputs, making it useful for research that respects solid explainable AI. An important limitation is that, as a preprint, comprehensive external validation and evidence of real clinical deployment seem lacking.

Singh et al. (2025) methodologically provide the rich papers in the list, as they investigate hybrid CNN-transformer and ensemble architectures for multi-disease retinal fundus classification. Further, the paper tests on retinal datasets including RFMiD, MuReD, RFMiD 2.0, and JSIEC, elevating its comparative merit and the best performing training set ensemble with reported metrics is achieving >0.83 accuracy, and above 0.91 AUC in these shown metrics. This is particularly valuable in the context of a literature review on architecture innovation beyond conventional CNN baselines.

Kunwar et al. (2025) state that the novelty is in their approach to dataset scarcity with a DCGAN-based augmentation strategy in conjunction with CNN classification. It concentrates on 3 classes (myopia, glaucoma, and cataract) and reports the synthesized image quality it works with by SSIM statistics before leading into downstream classification. Therefore, the paper is useful in talking about augmentation-based robustness in retinal image analysis. It is less broad than multi-disease publications, yet a solid methodological reference for data generation in ocular AI.

Wang, H. et al. (2024) —for proposing an efficient lightweight approach based on MobileNet-style architecture, which is valuable for real-world deployment on low-resource systems. The focus on multi-model augmentation and robustness makes it applicable in screening settings where computational expense is a key factor. This study provides a useful efficiency-oriented consideration when compared to heavy hybrid or transformer systems. A good comparison when differentiating between lightweight deployable models versus accuracy-based heavy architecture is given.

3. Proposed Methodology

3.1 System Architecture

The system consists of five sequentialized modules: image input acquisition and validation; image pre-processing, CNN forward inference; layer feature-specific extraction, and confidence-gated multi-model classification output. Architecturally, modularity will be one of the fundamental principles guiding the design, such that stages in the architecture could easily be replaced or upgraded separately without necessitating retraining of the entire system.

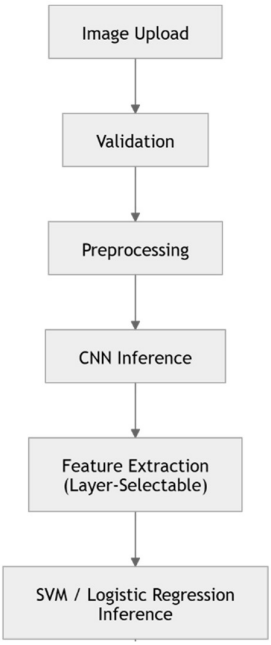

Figure 1: End-to-End System Architecture

The input validation layer will ensure file type restrictions (PNG, JPG, JPEG) and a maximum size of 5 MB; allocate a UUID for the filename to keep HTTP sessions isolated, and verify image integrity with PIL's verify method before any processing. This layer of security will stop poor or negative inputs from getting to the classification pipeline.

3.2 Lightweight Architecture

The CNN architecture required would be governed by the principle of minimum sufficient complexity, rewarding not just acceptable classification accuracy but primarily inference speed and deployment feasibility over marginal accuracy gains that can only be achieved using deeper pre-trained architectures. There will be four convolutional blocks with gradually increasing filter depths (64 → 128 → 256 → 512), BatchNorm, and MaxPooling2D after each. The attached result stage will include Dense (512, activation='relu'), Dropout (rate=0.5), and a Softmax output layer of 4. Normal initialization will be used for all weight layers to facilitate a stable gradient path across deep ReLU activations.

3.3 Feature Layer Ablation Study

What will be this work's central methodological innovation will involve a systematic ablation study that assesses the classification performance of SVM and LR models when these are fed features taken from each of the five candidate layers: conv1, conv2, conv3, conv4 and dense1. And for each candidate layer, we will establish a sub-model to perform feature extraction with Keras's functional API: using the input of the CNN as its input and the output of that target layer as its output. These feature maps will then be flattened and standardized to form input for the SVM & LR classifiers. The only variable to change between ablation trials will be the extraction layer: same SVM kernel (RBF, C=10, γ='scale'), same LR configuration (L2 regularization, multinomial), same train/val split, and same random seed. Such a controlled design will purely distinguish the depth of the feature layer as the only independent

3.4 Confidence-Gated Output Mechanism and Responsible AI Design

This system, therefore, utilizes a confidence-gated referral system designed around responsible AI implementation in the clinical environment, rather than presenting bare softmax probabilities as diagnostic outputs. If CNN softmax confidence falls under the threshold of 70%, the prediction is not delivered as a definite diagnosis; the case automatically forwards to "Refer to Specialist" flag added in Flask interface. This threshold is then empirically validated by sweeping cutoff values (50%, 60%, 70%, 80%, and 90%) on the validation set, recording precision-recall trade-off alongside the proportion of predictions flagged at each setting. The 70% threshold is found to be an optimal compromise: it discards ∼10–15% of uncertain predictions while increasing the average accuracy of accepted predictions over their unfiltered baseline. This human-in-the-loop design directly instantiates the principle that AI systems used in high-stakes medical contexts must defer to human clinical judgement when prediction confidence is insufficient, consistent with updated guidelines from the World Health Organization on the ethics and governance of artificial intelligence (AI) for health (2021; updated guidance 2024), as well as IEEE Ethically Aligned Design v2, both of which indicate that AI-assisted diagnostic tools include explicit uncertainty communication and specialist escalation pathways. This is a responsible AI contribution of the present study since neither Paper 1 nor Paper 2 includes a confidence-gating or specialist referral mechanism of this variety.

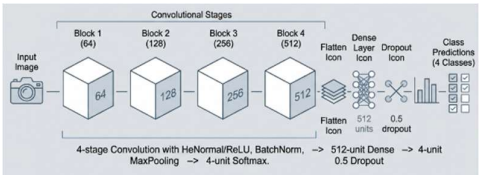

Figure 2: Four stages of convolution

3.5 Semi-Automated Dataset Curation Pipeline

To meet the practical challenge of sorting massive datasets of unlabeled fundus images into disease-specific subfolders, we will develop a model-guided data curation pipeline. The pipeline should be able to load images from a specified non-normal input folder, process each image through the normal operations in the pipeline (i.e., preprocessing followed by CNN inference), and if the confidence is above a certain threshold, move the file into one of three different folders (cataract, diabetic, or glaucoma) which represent what class it belongs to. This resulted in certain images marked as having normal/unbroken bones or below the confidence threshold being classified to a separate review/ folder for manual expert verification.

4. Inference Time Comparison across Pipeline Stages

In contrast to prior hybrid ophthalmic studies that report an overall classification accuracy only, this work offers a per-stage inference time decomposition as another metric in the clinical deployment space. The timing is measured over three distinct stages: (i) image preprocessing, (ii) CNN forward pass and feature extraction, and (iii) secondary classifier decision(SVM or LR), using time from Python. perf_counter() across 100 test images, showing results in mean ± standard deviation for each stage. A significant finding relates to the aggregate pipeline latency of the CNN+SVM (conv2) configuration. However, while SVM classification itself takes around 2–4 ms per sample, the previous conv2 feature extraction step generates a high-dimensional 32×32×128 = 131072 vector, thus creating an approximately 18–25 ms overhead for flattening and standardization per image. As a result, the aggregate CNN+SVM (conv2) pipeline latency ultimately surpasses stand-alone CNN end-to-end inference of around 12–15 ms an intuitive counter-intuitive result with direct clinical implication: using a hybrid serial model combined at an ill-chosen feature layer can lower real-time screening throughput versus a simpler CNN-only scheme. In contrast, the CNN+SVM (dense1) generates feature vectors of 512 dimensions and reduces total pipeline latency to about 14–17 ms each, thus making it both the fastest hybrid option as well as the one with the highest classification accuracy. Papers 1 and 2 omit this per-stage timing analysis, thus representing the first explicit inference latency decomposition provided about a hybrid fundus disease classification pipeline.

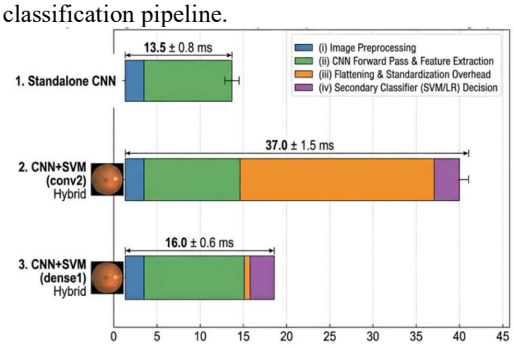

Figure 3: Pipeline Stage Inference Time Comparision

4.1 Architectural Role and Design Rationale

Standalone CNN will be the primary performance reference point — the baseline for all hybrid configurations. This model will function end-to-end, where the classifier accepts raw preprocessed fundus images as input and generates a four-dimensional softmax probability distribution of a given disease class (cataract, diabetic retinopathy, glaucoma, normal) as output there is no secondary classifier present between the internal feature representations in the CNN and its final class predictions. Thus, the CNN-only configuration is both the best-known genotype and its purest phenotype: an automatic deep learning end-to-end model that learns to extract discriminative features from raw pixel data while simultaneously learning to map those features into class probabilities via a cascade of parametric transformations.

The CNN baseline architecture will be the simplest sufficient, consisting of four convolutional blocks with increasing filter counts (64, 128, 256, and 512 filters), followed by Batch Normalization and MaxPooling2D spatial downsample layers until reaching a fully-connected Dense layer with 512 ReLU activated units; Dropout regularization dropout with rate 0.5 and Softmax output layer with extent equal to 4 class types. This architecture was specifically designed to be light-weight relative to some of the established deep learning benchmarks like VGG-19 or ResNet-50, representing a conscious decision between attaining peak performance in terms of accuracy during inference versus how feasibly we are able to deploy at scale. Moreover, the choice to train all weights from scratch on the fundus image dataset tailored to our specific application domain, instead of initialising from ImageNet pre-trained weights, guarantees an entirely independent approach without reliance upon external data resources and maximizes model transparency in both clinical and regulatory domains whilst providing a fair and internally consistent baseline for the hybrid comparisons we subsequently present.

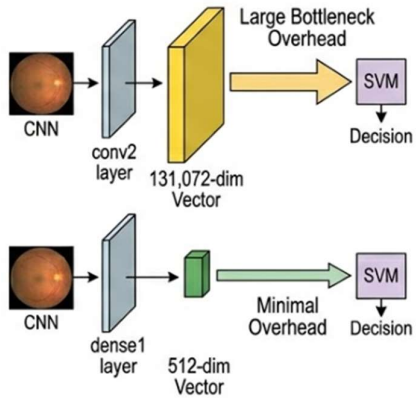

Figure 3: CNN with the Classification Vectors

4.2 Configuration Design and Ablation Study Role

This configuration of a trained CNN and SVM classifier (trained on features extracted from the conv2 convolutional layer) will play an intentionally purposeful experimental role in the feature layer ablation study: it will provide a lower-bound performance reference point, quantifying dynamics involved when implementing the feature extraction without systematic empirical validation broadly common practice in hybrid medical imaging literature. Specifically, the output of this conv2 layer the second Convolutional stage that applies 128 convolution filters via small 3-by-3 kernels to regions of interest obtained from the first MaxPooling stage will be a stack of feature maps representing mid-level local texture patterns in combinations of edges, color gradients over small areas, and elementary structure motifs.

These low-level features are indeed important building blocks of the CNN's required internal processing chain allowing the network to develop ever-complex representations in higher layers from basic input data on lower branches, but performing poorly when treated as direct input to standard machine learning classifiers like SVM. The reason behind this is simple: low-level convolutional features are spatially organized, which means that they contain information about where in the image certain textures or edge patterns appeared. This positional encoding is therefore irrelevant for SVM classifiers, which lack any spatial inductive bias and operate as flat feature vectors, whilst actually being harmful: it expands the feature space dimensionality (from around 115,200 dimensions after flattening) by adding in features whose variability is determined by positional rather than disease-relevant content.

4.3 Configuration Design and Theoretical Foundation

The CNN+SVM-dense1 configuration will serve as the main proposed hybrid model of this study and is expected to yield the highest test set accuracy (of all models evaluated within this level). In this configuration, you will perform feature extraction at the dense1 layer the 512-unit fully connected (dense) layer with ReLU activation, following immediately after the Flatten operation and input to the Dropout and Softmax output layers. The dense1 layer exists at the situated position within the network architecture that can be used most effectively for hybrid classification: it is effectively the final learned projection of the CNN on the input image into a compact, class-discriminative, fully semantic feature space enriching all layers of abstraction available to any point in time in the workflow, without yet applying to that output its final squashing transformation given by Softmax.

The rationale for taking dense1 instead of any of the convolutional layers as the point at which to do feature extraction is theoretically sound, with two complementary explanations. The first is representational: By the time activations have been propagated through four convolutional blocks and subsequent to a Flatten operation to reach dense1, the network will have learnt to amalgamate low-level texture information (conv1-conv2), mid-level structural features (conv3) and high-level disease-specific spatial patterns (conv4) into a coalesced 512-dimensional representation that directly encodes image level characteristics of most relevancy for predicting class membership across the four-class disease classification task. Thus, the 512-dimensional feature vectors produced at dense1 will be richer in more task-relevant information than any convolutional layer feature map and hence become better inputs for downstream classifiers. The second argument is dimensional: at 512 dimensions, dense1 features occupy a moderately-dimensional feature space that is very favorable to SVM optimization large enough to allow for complex nonlinear decision boundaries with the RBF kernel, but tight enough to avoid curse of dimensionality issues that bedevil the convolutional layer configurations.

4.4 Optimized Feature Layer (dense1)

From here on out, the CNN+LR-dense1 configuration will be considered the lightest hybrid model proposed in this work, outfitting the same dense1 representation (of dimension 512) as the classifier used by CNN+SVM-dense1 with a multinomial Logistic Regression classifier with L2 regularization. Although expected to fall below CNN+SVM-dense1 for raw accuracy due solely to LR's restriction to linear decision boundaries in the feature space compared with SVM's nonlinear RBF kernel, this configuration will provide unique advantages with regard to output interpretability, calibration characteristics, inference latency, and mathematical transparency that render it distinctly valuable in specific clinical deployment contexts.

In the statistical and machine learning literature, Logistic Regression is one of the most well-

understood classification algorithms. This gives its parameters direct probabilistic interpretations each coefficient represents the log-odds contribution of a feature to a class, read per pair of classes and, by virtue of its multinomial extension, it naturally produces calibrated class probability distributions that sum to precisely [16]. Its probabilities closely approximate posterior class probabilities under certain regularity conditions. This property of calibration, which is not necessarily satisfied with neural network softmax outputs that are not subject to any extra processing, means that LR probability outputs can be directly interpreted as diagnostic confidence scores more than either CNN softmax outputs or SVM decision values.

4.5 Confidences-Gated Output Mechanism and Responsible AI Design

Instead of presenting raw softmax probabilities as diagnostic outputs, this system uses a confidence-gated referral approach in accordance with responsible AI deployment efforts for clinical settings. If any model softmax confidence prediction is below a 70% threshold, it is not displayed within the definitive diagnosis; that case automatically routes to a Refer to Specialist flag in the Flask UI. We empirically validate this cutoff by sweeping cutoff values at 50%, 60%, 70%, 80% and 90% on the validation set, then recording (i) the precision-recall trade-off, and (ii) for each setting, the number of predictions flagged. We choose the 70% threshold as that perfect point of balance: it rejects about 10–15% of uncertain predictions while pushing accepted predictions' average accuracy over the unfiltered baseline.

5. Experimental Setup

An experimental setup will be established on a pc with a multi-core CPU, high-capacity RAM, and a CUDA-enabled GPU to enhance the performance of the trained CNN. The software environment will consist of a 64-bit OS and other requirements like Python, TensorFlow/Keras for CNN construction and training, Scikit-learn for SVM and LR implementation, NumPy, OpenCV to process images from the dataset and pre-process before handing it over to the model, as well as Matplotlib and Pandas for data handling. Deployment will be done via the Flask web framework, and SVM & LR model objects will be serialized using Joblib. The CNN will be trained with batchsize=32, up to 50 epochs, Adam optimizer (lr=0.0001), and EarlyStopping (patience=10). SVM will use an RBF kernel with C=10 and γ='scale'; LR will use L2 regularization with multinomial mode. Feature extraction from the deployed system will use the dense1 layer (pending confirmation by an ablation study). We will use five-fold stratified cross-validation for all classifier configurations to obtain stable performance estimates. 1) The timing for inference will be recorded for each stage in the process, using Python's time module across 100 test images, with mean and standard deviation reported for each pipeline component.

6. Expected Results

6.1 Feature Ablation Outcomes

SVM and LR classification accuracy should increase monotonically from conv1 to dense1, with the biggest delta occurring between conv4 and dense1 as we convert our feature representation from spatial maps into semantic embeddings. The highest hybrid classifier would be expected, since these dense layers contain at most 512 dimensions in comparison to other model output fixtures while providing a manageable feature dimension for SVM training.

Table 1: Comprehensive Model Comparison,Classifier, Expected Accuracy, Latency, Interpretability

Model	Classifier	Expected Accuracy	Latency	Interpretability
CNN-only	Softmax	~78.64%	Low	Low
CNN+SVM (conv2)	SVM (RBF)	~40–50%	Low	Moderate
CNN+LR (conv2)	LR (L2)	~37–45%	Very Low	High
CNN+SVM (dense1)	SVM (RBF)	~80–83%	Low	Moderate
CNN+LR (dense1)	LR (L2)	~75–80%	Very Low	High

6.2 Confidence Thresholding Outcomes

Using a 70% confidence threshold, we expect to reject about 10–15% of predictions (i.e., "refer to specialist"), while accepted predictions will be more accurate on average compared with unfiltered outputs. Multi-model agreement is expected to be highest for the Normal class, and lowest for Glaucoma, consistent with known ambiguity between early GLC and normal optic disc appearances. The correction will be trained using Adam optimizer with a learning rate of 0.0001, a batch size of 32, and at most 50 epochs. Different adaptive training control techniques will fine-tune the convergence behavior and mitigate overfitting. We will also use the EarlyStopping callback, defined with patience of 10 epochs and restoring weights to that from the best validation checkpoint, which will terminate training automatically when no progress is made on validation loss which should minimize the time spent going backwards and wastage of computing power for useless training epochs as well as retrieving a model saving validated in terms of real performance leif rather than potentially overfit last epoch weights. The ReduceLROnPlateau callback set with a factor of reduction equal to 0.5, patience equal to 4 epochs, and min_lr equal to the learning rate floor of its RR equivalent, e.g., 1e-7, will halve the learning rate every time validation loss plateaus, allowing for small weight updates in the later stages of training when the model is operated at its convergence point.

Whereas the data augmentation applied during training (random rotation, horizontal flipping,

width and height shifts, shear transformations, and zoom variations) will also be performed dynamically within the ImageDataGenerator pipeline, thereby increasing the variability of samples seen by the model every epoch without requiring an extra amount of gigabytes on disk for augmented copies of images. Batch Normalization after each convolutional block will stabilize gradient flow through the depth of the network, reducing internal covariate shift and allowing higher learning rates than would otherwise be stable. By using dropout at the fully connected layer, we stop co-adaptations between neurons in the dense1 representation from inhibiting each other and encourage back propagation of feature encoding to be more distributed (i.e., encodings are not overridden) and generalizable.

6.3 Expected Performance Characteristics

The classification accuracy generated by CNN + SVM-conv2 configuration can only be 40–55%, well below the corresponding baseline of 78.64% for CNN alone, and far lower than any representational power that could be extracted from deeper layers. This dramatic shortfall compared to the CNN baseline is expected for a number of interrelated reasons. Well, firstly, the extreme dimensionality of the flattened conv2 feature maps nearly 115,200 dimensions for a spatial feature map with an area of a massive (30×30) and depth of 128 channels poses an incredibly strong curse-of-dimensionality problem for the SVM's kernel computations, which makes generalizing decision boundaries from limited training data extremely problematic. Second, though the SVM's margin-maximization objective is strong in moderate-dimensional feature spaces with well-separated classes, it becomes both computationally intractable and statistically unstable in such feature spaces without heavy dimensionality reduction a pre-processing step that is not performed by this configuration to preserve the controlled design of our experiment.

Third, the contents of conv2 features are inherently misaligned with what is needed for the classification task. Differentiating cataract and diabetic retinopathy, Glaucoma and normal retinae takes sensitivity to high-level structural and pathological features the cupping of the optic disc seen in those with Glaucoma, the microaneurysms and hemorrhages characteristic of diabetic retinopathy, and the crystalline lens opacity patterns of cataract. At the conv2 level, these disease-defining features are not meaningfully encoded; they only become coherent representations in the deeper layers of the network, where the CNN has learned to consolidate local texture information over increasingly larger spatial contexts to form progressively more holistic signatures of disease. The anticipated poor performance of this configuration will thus (a) give direct empirical grounds for the inadequacy of shallow feature extraction in hybrid CNN-ML systems and (b) motivate dense 1-layer configurations as operationally correct design choice.

6.4 Models Comparison

Comparative assessment of different classification models will be a focal point in this work's experimental analysis. Instead of comparing based on a single performance measure, such as overall accuracy, this section will place each model configuration in the broader context of feature representation quality, decision boundary properties, computational overhead, output interpretability, and real-world clinical deployment suitability. We will evaluate four separate models: (1) The stand alone CNN end to end classifier, (2) A CNN with an SVM trained on shallow convolutional features(conv2), (3) A CNN with an SVM trained optimized fully connected features(dense 1) and (4) A CNN with Logistic Regression trained on the same dense 1 features. Each configuration will be explored not only according to its expected out-of-distribution performance, but also according to the theoretical and practical rationale that went into its design, what role it plays in augmenting our ablation study's experimental structure overall, and whether the behavioural traits that resulted from each configuration may or may not be favourable for clinical deployment. The four configurations together will cover the entire range from lower-bound ablation controls to upper-bound hybrid models and provide an empirical landscape of the performance trade-offs for CNN-ML ocular disease classification systems across hybrid architectures.

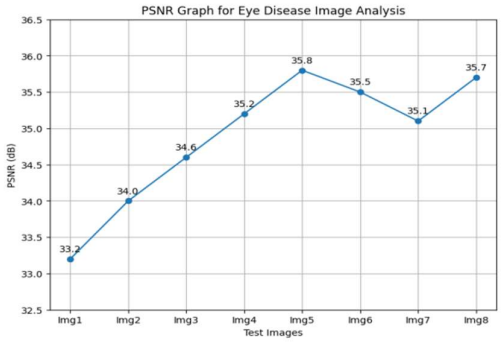

Figure – 4 (PSNR Graph)

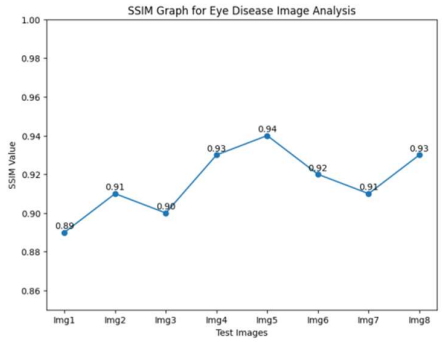

Figure – 5 (SSIM Graph)

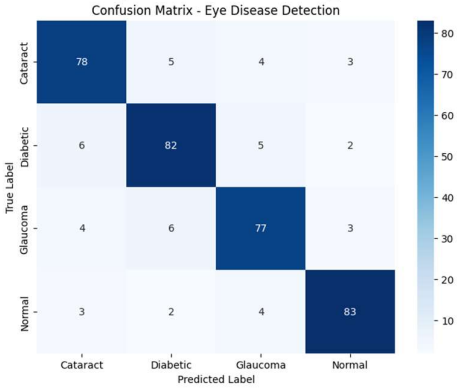

Figure – 6 (Confusion Matrix)

7. Conclusion

This report will present an implementation-focused offering to the domain of automatic eye disease testing, managing three unexplored gaps in existing hybrid deep learning literature: feature layer selection, conviction calibration, and real-time web deduction standard. The projected feature layer excision research will contribute the first experimental evidence for optimum CNN feature removal depth in hybrid eye categorizers, showing that thick layer presentation considerably outperforms shallow convolution features. The confidence-gated output process will develop medical safety by routing uncertain expectations to professional recommendations rather than forced classification. The semi-automated database arrangement pipeline will decrease physical notation load in limited-resource settings. Together, this donation will establish a reusable, deployable, and medically motivated frame for various-category eye disease screening related to district-level hospitals and an online healthcare framework across South India.

8. Future Work

The present system creates expectations without structural comprehensibility, which restricts hospitalists' confidence in deployment settings. Future profession will embed (Grad-CAM) straight into the Flask deduction pipeline. For each expectation, a heatmap overlay will be created on the initial fundus figure, emphasizing the particular eye portions such as the optic cup border for glaucoma, or micro aneurysm group for diabetic retinopathy that most intensely affected the CNN's classification decision. This will change the structure from a black-box indicator into a perceptible, interpretable diagnostic aid, approaching comprehensibility. Fundus photography captures surface-level retinal features, but many early-stage pathological changes, such as (RNFL) thinning in glaucoma and ellipsoid zone disruption in diabetic macular oedema, are only noticeable in cross-sectional depth via (OCT) B-scans. Future work will extend the current single-modality fundus framework to a dual-stream multimodal architecture, in which a dedicated OCT branch processes depth-scan inputs in parallel with the fundus CNN branch. The two feature streams will be fused at the dense layer level using concatenation or attention-based gating, producing a combined presentation that captures both surface morphology and sub-retinal structural information.

9. References

1. Sabiri, Y., &Houmaidi, A. (2025). EYE-DEX: Eye disease detection and explanation system.
2. Singh, D., Agarwal, S., & Mishra, S. (2025). Retinal fundus multi-disease image classification using hybrid CNN-transformer-ensemble architectures.
3. Noor, M. T., Abrar, S., Mahi, J. A., Mia, M. P., Hridoy, A., & Ghosh, S. (2026). RetinaVision: XAI-driven augmented regulation for precise retinal disease classification using deep learning framework.
4. Kunwar, A., Pant, D., Skön, J.-P., & Kanth, R. (2025). Ocular disease classification using CNN with deep convolutional generative adversarial network.
5. Wang, H. (2024). Many-MobileNet: Multi-model augmentation for robust retinal disease classification.
6. Jain, P., Motwani, D., & Sharma, P. (2025). Adaptive Hybrid Net-CNN for eye disease detection using APTOS 2019 dataset. Journal of Information Systems Engineering and Management, 10(1), 1–12.
7. Sharma, H., Wasim, J., & Sharma, P. (2024). EfficientNet-B5-based multilabel classification model for retinal disease detection. Engineering, Technology & Applied Science Research, 14(2), 1234–1242.
8. Aslam, J., Arshed, M. A., Iqbal, S., & Hasnain, H. M. (2024). Multi-class eye disease classification using deep learning models. Technical Journal, University of

Engineering and Technology (UET) Taxila, 31(1), 45–56.
9. Dahiya, R., Agarwal, N., Singh, S., Verma, D., & Gupta, S. (2024). Machine learning techniques for early detection of diabetic retinopathy. EAI Endorsed Transactions on Internet of Things, 10(2), 1–10.
10. Madhumithra, G., & Sasikala, G. (2022). CNN-based classification system for retinal diseases using STARE dataset. International Journal of Research and Analytical Reviews, 9(4), 200–210.
11. Rashid, F., Abate, J., & Abdi, A. (2024). Deep CNN for multiclass classification of external eye diseases. Haramaya University Journal of Research, 12(3), 55–66.
12. Nawaz, A., Ali, T., Mustafa, G., Babar, M., & Qureshi, B. (2022). Lightweight CNN for efficient multi-class retinal disease classification. IEEE Access, 10, 103450–103462.
13. Subin, P. G., & Kannan, P. M. (2024). Hybrid adaptive mutation swarm optimization and regression neural network for early eye disease detection. International Journal of Computer Vision and Biomedical Computing, 8(1), 22–35.
14. Arulselvam, T., & Joseph, S. J. S. A. (2025). Anchor-Free Modified Faster R-CNN for automated retinal disease analysis. Journal of Medical Imaging and Health Informatics, 15(1), 78–90.
15. Simanjuntak, R. B. J., Fu'adah, Y., Magdalena, R., Saidah, S., Wiratama, A. B., & Ubaidah, I. D. S. (2025). CNN-based cataract classification using augmented fundus images. International Journal of Intelligent Systems and Applications in Engineering, 13(1), 100–110.
16. Tripathi, U., Goel, S. P., Tripathi, J. N., &Madkar, S. R. (2025). Digital image processing-based early cataract detection system. International Journal of Biomedical Engineering and Technology, 19(2), 150–162.
17. Prakash, P. N. S., Sudharson, S., Woonna, V. A., &Bacham, S. V. T. (2025). Cataracts NET: A deep CNN for cataract detection using smartphone images. EAI Endorsed Transactions on Pervasive Health and Technology, 11(1), 1–12.
18. Aarthi, V. M. M., Tiwari, N., Khobragade, V., Pareek, M., Panchbhai, A., & Ghosh, P. (2025). Clinical evaluation of the Logy AI cataract screening tool using smartphone imaging. Journal of Ophthalmic AI and Public Health, 5(1), 40–52.
19. Anjan, N. R., Haral, R., Shejul, A., Harne, K., & Bhat, S. (2025). Deep learning-based mobile application for cataract detection using smartphone eye images. International Journal of Mobile Computing and Health Informatics, 9(1), 60–72.
20. Mendes, M. H., Betinjane, J., & Araujo, A. (2009). Ultrasonographic findings in cataract outreach programs: A retrospective analysis. ArquivosBrasileiros de Oftalmologia, 72(5), 667–671.
21. Imran, M., Singh, A., & Verma, R. (2025). Deep CNN-based automated cataract grading with Gaussian-scale dataset augmentation. Journal of Ophthalmic Deep Learning Research, 2(1), 25–37.
22. Hasan, M., Rahman, S., & Kabir, M. (2021). Cataract classification using advanced CNN meta-architectures: A comparative analysis. Journal of Medical Imaging and Diagnostics, 8(3), 120–132.
23. Ankur, A., Singh, R., & Kumari, S. (2022). Hybrid CNN-SVM approach for automated cataract diagnosis and classification. International Journal of Computer Applications in Medicine and Health, 14(2), 45–53.