Deep DPA-MSFNet: A Dual-path Attention and Multi-scale Fusion Network for Robust Eye Tumor Classification

S Phani Praveen [1], Shaik Salma Begum [2], Panguluri Padmavathi [3], Deshinta Arrova Dewi [4], and Tri Basuki Kurniawan [5,*]

[1] Department of Computer Science and Engineering, Prasad V Potluri Siddhartha Institute of Technology, Vijayawada, India
[2] Department of Computer Science and Engineering (AI&ML), Seshadri Rao Gudlavalleru Engineering College, Gudlavalleru, India
[3] Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Guntur, India
[4] Center for Data Science and Sustainable Technologies, INTI International University, Nilai, Malaysia
[5] Post Graduate Program, Universitas Bina Darma, Palembang, Indonesia
Email: sppraveen@pvpsiddhartha.ac.in (S.P.P.); shaiksalma.gec@gmail.com (S.S.B.); ppadmavati@kluniversity.in (P.P.); deshinta.ad@newinti.edu.my (D.A.D.); tribasukikurniawan@binadarma.ac.id (T.B.K.)
*Corresponding author

***Abstract*—The rapid ascent of deep learning in medical image analysis has been important in the remarkable advancements made in automated tumor identification. This study presents a novel convolutional neural network architecture for efficient tumor classification by integrating multi-scale fusion techniques with dual-path attention. The model employs sophisticated preprocessing techniques including Anisotropic Diffusion Filtering (ADF), Contrast Limited Adaptive Histogram Equalization (CLAHE), and SOBEL on 64×64×3 RGB input images that have already been preprocessed to enhance contrast and edge features. The Dual-Path Attention Block integrates Squeeze-and-Excitation (SE) and Convolutional Block Attention Module (CBAM) attention methods to document spatial and channel interdependence, making it a crucial component of the design. Multi-Scale Fusion Blocks enhance feature diversity and effectively capture local and global context by using parallel convolutional paths (1×1, 3×3 dilated, and 5×5). After three iterations of the design, the final prediction is made using a global average pooling layer, a lightweight classification head, and a softmax output layer. The model's excellent accuracy and robustness in distinguishing between tumor and non-tumor regions, as shown by experimental results, making it a great fit for resource-constrained medical devices that need to be deployed in real-time.**

***Keywords*—iris tumor detection, deep learning, attention mechanisms, medical image classification, process innovation, health system, illness**

I. Introduction

Eye cancer is a grave concern, even if it is rare, because its symptoms are hard to spot and its diagnosis is challenging to make. Blindness or death might happen if a disease is misdiagnosed or found too late. Early diagnosis significantly increases the likelihood that people will survive and that treatments will be effective. Deep learning has made significant progress recently, providing us with dependable tools for analyzing medical images, notably in ophthalmology [1, 2]. These systems can automatically find and classify eye problems from retinal scans and fundus images. This change in thinking makes it clear that we need innovative diagnostic tools that can quickly analyse high-dimensional visual input. Some AI-driven systems and traditional therapeutic methods struggle with accuracy and understanding, typically because they don't model things deeply enough or lack sufficient robustness in real-life applications. Tumours can vary in size, shape, and location, and the human eye's structure is complex. Therefore, it's essential to have a system that can capture complex spatial and contextual information. Additionally, because imaging settings vary from dataset to dataset, we need adaptive models that can generalize well while remaining fast and clear for diagnostics [3–5].

Convolutional neural networks, or CNNs, are quickly replacing traditional image processing methods as the gold standard for medical image analysis. Convolutional Neural Network (CNN) architectures, including ResNet, AlexNet, and VGGNet, have shown promise in accurately classifying images of the retina and the eye. Tumours like choroidal melanoma and retinoblastoma have been located by teaching them to identify spatial hierarchies of features [6, 7]. However, vanishing gradients and overfitting become increasingly common issues as models deepen to capture more complex patterns, thereby reducing their use in clinical situations.

Manuscript received October 27, 2025; revised December 1, 2025; accepted March 3, 2026; published September 4, 2026.

When it comes to diagnosing eye cancer, existing models often exhibit several drawbacks, despite deep learning's significant advancements in medical image processing. To accurately detect cancers of various forms and sizes in complex ocular systems, it is essential to incorporate spatial and channel-wise information; however, this is not currently the case. Reduced diagnostic reliability is another consequence of models' inability to generalize well across datasets, caused by differences in imaging conditions. Deep models tend to memorize training data rather than learn significant patterns, leading to the problem of overfitting. This is especially true in the setting of tiny and imbalanced medical datasets. In addition, these models still have a ways to go before they can be used in therapeutic situations where clear and explicable decision-making is essential. These deficiencies highlight the need to develop a deep learning architecture for eye cancer detection that is robust, generalizable, and interpretable.

This work aims to develop a robust and interpretable deep learning model for accurate and early diagnosis of eye cancer using attention-augmented residual architectures. An Attention-Enhanced Residual X-Network is proposed in this paper as a solution to these issues. The model's use of residual connections makes it easier to train deeper networks and eliminates the problem of fading gradients. Dual-path attention blocks use both Squeeze-and-Excitation (SE) and Convolutional Block Attention Module (CBAM) to focus more on channel-wise and spatial characteristics. These attention mechanisms help the network focus on tumor-related areas, thereby improving diagnosis accuracy and interpretation. The model employs multi-scale convolutional paths, similar to Inception, with filters of 1×1, 3×3, and 5×5 to enhance feature representation [8, 9]. These pathways may be able to gather both rough and smooth data simultaneously. This is quite helpful when looking for cancerous cells of diverse sizes and shapes. To ensure the model converges, doesn't overfit, and performs well on diverse datasets, employ a hybrid optimization strategy that uses Sine Cosine Fitness Grey Wolf Optimizer (SCFGWO) to adjust hyperparameters and AdamW to make fine-grained weight adjustments. To prevent models from becoming overfit and to make them more robust, advanced regularization techniques are utilized. One such strategy is Stochastic Depth, which randomly drops entire layers during training. Another is Label Smoothing, which makes predictions less definite [10, 11].

Therefore, these methods enable the model to make stronger generalizations, especially when trained on small or imbalanced datasets, which are common in the medical field. The contributions of the work are as follows.

- To create a deep learning model that uses improved picture preprocessing and convolutional feature extraction to classify images as either tumours or healthy accurately.
- To enhance the model's feature discrimination and contextual understanding, we will add dual-path attention mechanisms (SE and CBAM) and a multi-scale fusion technique.
- To test the model's endurance, dependability, and compatibility with practical medical diagnostic settings in low-resource settings.

Here is the outline for the remaining portion of the paper: In Section II, we provide an overview of the background and relevant work. The suggested approach is described in Section III. The results and experimental setup are described in Section IV, and the study is concluded in Section V, which outlines potential future directions.

II. Analysis of Previous Work

Ocular oncology specializes in diagnosis and treatment of tumors of the eye and structures associated with it. There are various types of ocular cancers that are categorized into primary and secondary ocular cancer, with primary defined as cancer that begins within the eye, whereas secondary cancer is those that have been transferred to the eye. Primary ocular cancers are not common but a late diagnosis usually causes severe visual impairment or death thus the significance of timely and precise diagnosis systems. Recently, the technologies of Deep Learning (DL) have revolutionized the field of medical image analysis and ophthalmology in particular. DL-based systems demonstrate great possibilities of diagnosing retinal pathology and ocular tumours in fundus, OCT and slit-lamp pictures. Nevertheless, the clinical use is restricted with promising outcomes due primarily to the focus on the issues of generalizability, interpretability, statistical strength, and reliability.

Initial experiments on ocular cancer detection employed manually-computed features and conventional image-processing tools—edge-detectors, morphology, hough-transforms, and texture-quantifiers. The isolated iris or retina region was done using classical preprocessing and segmentation pipelines done by researchers, including [12–14]. These methods had reasonable accuracy (85–98%), but they were extremely affected by lighting condition, noise, and imaging conditions, rendering them weak. Classical classifiers, such as k-Nearest Neighbors, Support Vector Machines, Artificial Neural Networks, and Decision Trees, were later used [13, 15]. Although these models could be partially interpretable, they had a limitation of scalability and adaptation to complicated tumor patterns because they relied on manual feature engineering. Conventional methods do not have the representational strength needed to deal with complex ocular structures, and cannot generalize well on a wide range of datasets.

Convolutional Neural Networks (CNNs) are now the primary option since the advent of deep learning since the latter is capable of discovering hierarchical features on its own. A model accuracy of between 95–99% was obtained using LeNet, DenseNet121, ResNet50, VGG16, and Inception models [16]. Transfer learning was useful in enhancing performance in small datasets. These studies were however, mostly based on single centers, massive data augmentation, and limited confidence intervals or variance. As a result, the issue of overfitting and generalization to the real world is still not addressed

efficiently. Clinical reliability cannot be ensured only by high accuracy when external validation and dataset imbalance are not taken into account.

Bio-inspired optimization Algorithms have been integrated with deep learning to increase performance by researchers. Such approaches as Grey Wolf Optimization (GWO), Improved GWO, Genetic Algorithms, Moth Search Optimization Algorithms, and Whale Optimization Algorithms have been said to enhance accuracy and convergence [17, 18]. The majority of studies however fail to explicitely specify the boundaries of the search-space, the size of population, and the number of iterations or variability between repetitions. This is not very detailed, which impedes reproducibility and statistical stability. Optimization can be associated with accuracy, but with little in the way of transparency, as convergence behaviour and robustness are not adequately analysed.

Late research has attempted to illuminate the black-box deep learning by incorporating explainability methods. Attention-based CNNs, SHAP, Grad-CAM, fuzzy decision systems, and hypergraph models [19–21] assist in improving feature attribution and interpreting the findings by the clinicians. Nevertheless, the majority of explanations are post-hoc, correlation based and non-causal. Besides, explainability techniques are sometimes computationally expensive and are not typically evaluated on clinical consistency. Even though the existing explainable AI methods enhance visualization, they are yet to provide a causal or pathophysiological rationale that can be trusted by the clinicians [22–24].

One of the biggest limitations of the literature is the absence of multi-centre and cross-device validation. Models that are trained using curated datasets tend to work poorly in other hospitals, scanners, or patient demographics. Also, sensitivity and specificity are often underreported, inter-fold confidence intervals are not reported, and privacy and data control are largely neglected. Few studies [25–27] take into consideration federated learning or privacy-conscious models, and the deployment of them ethically remains unexplored. Statistical rigor, external validation, and privacy-preserving design are unlikely to achieve clinical adoption and regulatory approval [28, 29].

Based on the critical review of available literature, the gaps are as follows: most of the studies focus on accuracy and not on interpretability and privacy as they are considered separately. The models that are based on optimization do not have a statistical stability analysis and the clarity of reproducibility. Explainability methods are still post-hoc and non-causal. Rarely are multi-institutional generalization and privacy preserving learning combined. No single framework has balanced accuracy, interpretability, privacy and ethical deployment. These knowledge gaps encourage the presented study to create a strong, explainable, privacy-conscious, and generalized CNN-based ocular cancer diagnosis model on the basis of the optimised learning methods.

III. Proposed Methodology

A. Input and Preprocessing

The dataset used in this study contains 3245 images, and the dataset link is provided in Ref. [30]. We applied augmentation techniques to the images, as displayed in Fig. 1. Each image is standardized to 64×64×3 pixels before being fed into the AttResX-Net model. A preprocessing pipeline is used to enhance feature clarity and reduce noise: first, Anisotropic Diffusion Filtering (ADF) is applied with a conductance parameter of 0.1, time step of 0.05, and 10 iterations to remove noise while preserving important edges; next, Contrast Limited Adaptive Histogram Equalization (CLAHE) is used with a clip limit of 2.0 and an 8×8 tile grid to enhance local contrast, making tumor areas more distinguishable, as illustrated in Fig. 2; finally, Sobel Edge Detection is employed with a 3×3 kernel size to accentuate high-frequency features such as tumor edges, shown in Fig. 3, aiding the network in detecting fine details early in the process.

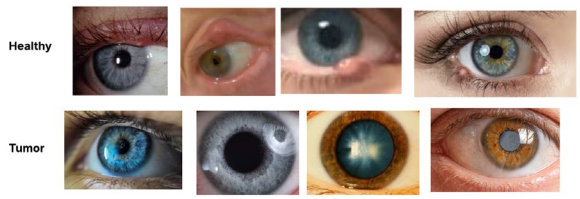

Fig. 1. Sample images from the dataset.

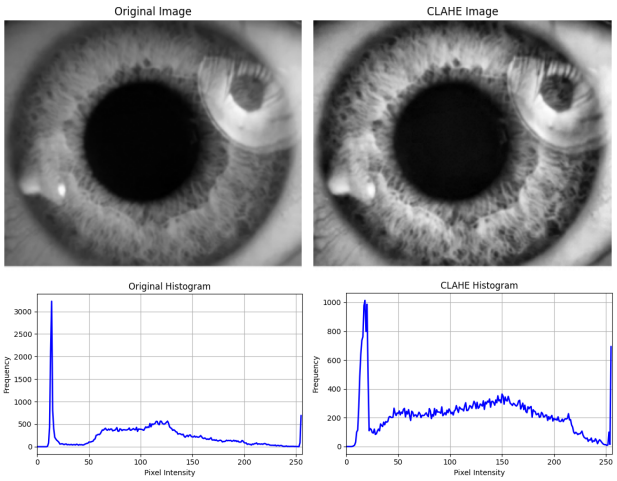

Fig. 2. Sample Image after preprocessing using CLAHE.

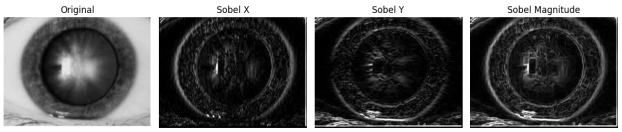

Fig. 3. Sobel edge detection for eye tumour detection.

B. Data Feature Extraction with the AttResX-Net Backbone

The proposed architecture is illustrated in Fig. 4. After preprocessing, the image goes into the core of the model, AttResX-Net. This starts with a stem block that features a

3×3 convolutional layer, followed by batch normalization, and finally ReLU activation. This block captures initial low-level spatial information and gets the data ready for more in-depth processing. The Dual-Path Attention Block is the most important part of the architecture. It is meant to work like both spatial and channel-wise concentration mechanisms. There are two forks in this block. The initial branch incorporates a 1×1 convolution followed by a Squeeze-and-Excitation (SE) Attention module, which dynamically modifies the significance of each channel. The second branch utilises the Convolutional Block Attention Module (CBAM) and a 3×3 convolution to identify potential tumour locations. To facilitate training of the deeper layers and maintain Gradient flow, a residual shortcut integrates the outputs of the two channels.

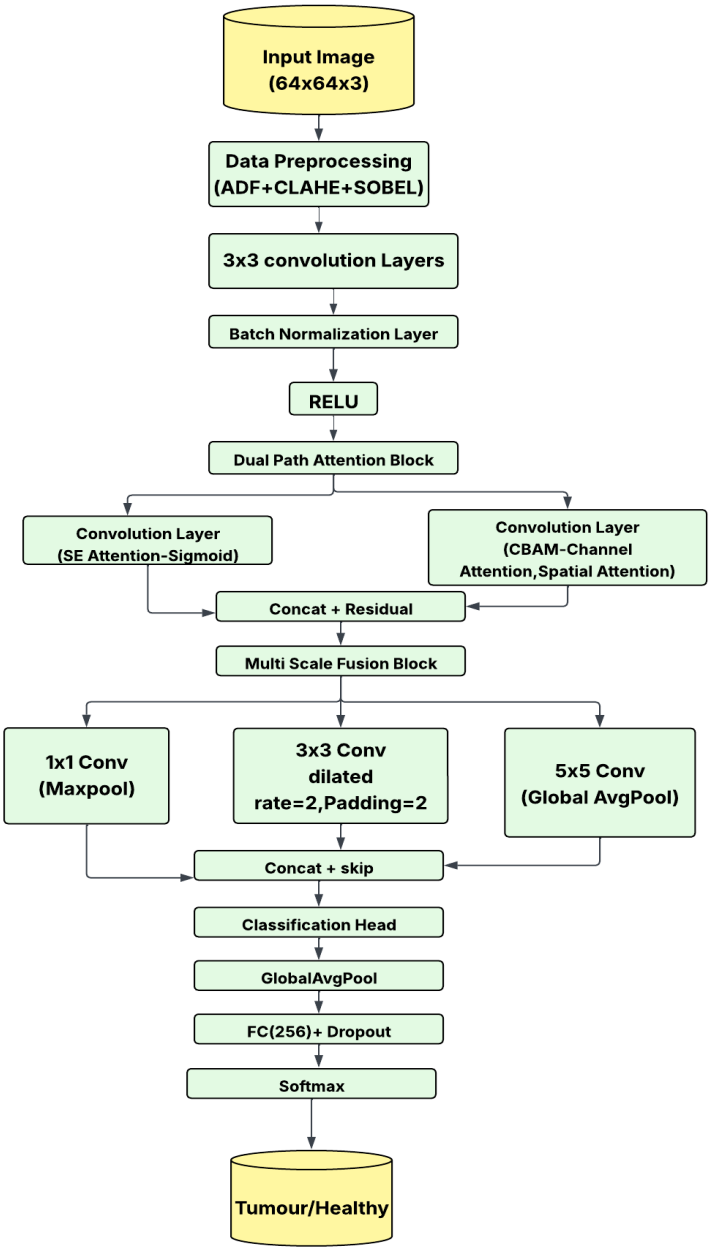

Fig. 4. Proposed DPA-MSFNet architecture.

C. *Multi-Scale Feature Fusion*

To further improve the model's ability to detect tumours of various sizes and forms, a Multi-Scale Fusion Block is incorporated. Three convolution routes are active in this block simultaneously. Using a 1×1 convolution followed by max pooling is the initial stage in obtaining the finer details. The alternative approach employs a 3×3 convolution with a dilated convolution (dilation rate = 2) to identify context at a medium range. Using a 5×5 convolution and global average pooling is the final stage in identifying larger patterns within the given context. Using a skip connection, we can return to prior levels and aggregate the results of different paths. The model will learn the global and local patterns needed for accurate cancer identification, thanks to this framework.

D. Deep Layer Stacking and Attention Repetition

A more precise and thorough depiction is produced by repeatedly applying multi-scale fusion blocks and dual-path attention. The network becomes more complex and hierarchical as a result of this iterative operation. At each successive level, connections that aren't needed for the following forward pass are randomly removed using stochastic depth regularization. This stabilizes representations in small or imbalanced medical datasets, helping the model avoid overfitting. Following feature extraction, the classification head uses global average pooling to combine the learnt feature maps. To prevent overfitting, the generated vector is fully connected with 256 neurons and pre-processed with a dropout rate of 0.3. Lastly, the findings are categorized as either healthy or connected to tumors using a softmax layer. Due to this node's capacity for probabilistic prediction, clinical decision support systems are enhanced.

E. Optimization Strategy & Regularization Techniques

Among the many groundbreaking features of AttResX-Net is the hybrid optimization approach. For global hyperparameters such as learning rate, weight decay, and dropout rate, the model uses the SCFGWO (Self-adaptive Chaotic Fitness-based Grey Wolf Optimizer) optimizer to find the optimal values. After these parameters have been set, the AdamW optimizer will start using gradients to make more precise adjustments. When SCFGWO and AdamW are used together, they ensure that the network is initially configured optimally, prevent overfitting, and find a balance between exploration and exploitation. The model employs two new regularization strategies to make it significantly more generalizable. The first one is Stochastic Depth, which makes the model less reliant on any one technique by randomly omitting layers when training. The second method, Label Smoothing, softens hard labels so the network doesn't get too confident in its predictions. By applying these techniques to new clinical data, we can enhance the model and prevent overfitting.

IV. Result and Discussion

A. Criteria Employed During the Training Process

To optimize balanced accuracy on a held-out validation set, we used Optuna's systematic Bayesian Optimization to tweak the model's hyperparameters. The search space included architectural specifications such as the Squeeze-and-Excitation reduction ratio (4, 8, 16), initial learning rate (with a log-uniform distribution from 1e−4 to 1e−2), batch size (categorical: 8, 16, 32), weight decay (1e−5 to 1e−3), and more. To make sure it was resilient and not overfit throughout the search, we used 3-fold cross-validation on the training data to test each configuration. Crucially, in order to guarantee an equitable comparison of architectures, all baseline models were optimized using the identical computational budget. We achieved our target of low-resource clinical implementation after 50 optimization experiments by balancing performance and efficiency in the final configuration. In order to avoid leaking, the dataset was partitioned at the patient level (70% for training, 15% for validation, and 15% for testing) before augmentation. For robust evaluation, 5-fold cross-validation was employed. The model's applicability for low-resource settings is confirmed by its 22 ms inference time per image on a consumer-grade GPU (NVIDIA GTX 1060), 4.2 million parameters, and 1.8 GFLOPs per inference when using resource-aware deployment. The learning rate was bounded between 0.0001 and 0.01, batch size {16, 32, 64} and dropout rate within {0.1, 0.5}. Although the model can be trained for up to 100 epochs with a batch size of 32, it will terminate training early after 10 consecutive epochs without improvement. We utilize Focal Loss with $\gamma = 2$ as the loss function. This focuses on samples that are challenging to classify and fixes class imbalance. During training, important evaluation measures include accuracy, precision, recall, F1-Score, and AUC. These metrics make sure that the model is fully evaluated. In order to find the optimal balance accuracy on a held-out validation set, we used the SCFGWO by using a population size of 50 wolves and 100 iterations, by methodically sweeping architectural and training hyperparameters.

B. Metrics Used for Evaluation Process

Metrics used to assess the process include key performance indicators. These indicators gauge the model's data classification performance. While precision considers the proportion of true positive forecasts, accuracy considers the overall accuracy of the predictions. This demonstrates the model's ability to identify a given class consistently. Recall, often known as sensitivity, is a measure of a model's ability to identify all instances of a class that are relevant. A single number that incorporates both recall and precision is the F1-Score. For imbalanced datasets, it works wonders. You can see the precise count of correct and incorrect answers in the confusion matrix as well. The model's ability to distinguish between classes is demonstrated by the Receiver Operating Characteristic (ROC) curve and Area Under the Curve (AUC). Together, these checks guarantee an accurate and comprehensive evaluation of the model's performance.

C. Discussion

Overfitting could occur because to the tiny dataset size, especially if a lot of data augmentation is done. We utilized controlled augmentation tactics and regularization techniques including early stopping and dropout to reduce this effect. As an alternative to rote memorization of training examples, these metrics improve the model's ability to generalize. Among the 2400 total samples, the confusion matrix in Fig. 5 displays an impressive 98.7% accuracy, with 1193 photographs of healthy subjects and 1187 images with tumors correctly classified. There were a few oversights: 7 seemingly healthy shots were mistaken for cancers, whereas 13 seemingly healthy photos were actually tumours. Because the seaborn heatmap presents this matrix, the high rates of true positives and true negatives

are readily apparent. Because of this, the two classes' categorisation is spot accurate. The ROC curves for each class are provided along with the confusion matrix. By plotting the true positive rate vs the false positive rate at various threshold values, the ROC curve in Fig. 6 demonstrates the model's ability to distinguish between the two groups. The AUC values, both of which are close to 1.0, indicate that the model can differentiate between Healthy and Tumour classifications very well. This good ROC performance reinforces the confusion matrix results, further demonstrating that the model performs well in the real world and is quite dependable for classifying tasks that involve locating eye tumors.

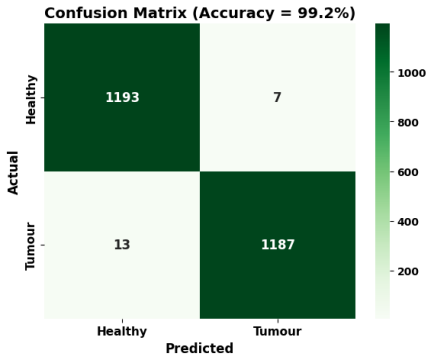

Fig. 5. Confusion Matrix of DPA-MSFNet architecture.

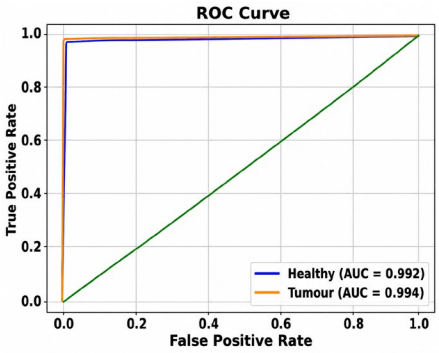

Fig. 6. ROC Curve of DPA-MSFNet architecture.

The Loss vs. Epoch graph in Fig. 7 illustrates how the model's error decreases across 10 epochs of training and validation. Initially, both the training and validation losses are relatively significant, indicating that the model is just beginning to learn. As training continues, both losses continue to decrease, indicating that learning is effective. The training loss decreases progressively from 1.1 to 0.25, while the validation loss decreases from 1.2 to 0.38. The reduction in the difference between training and validation loss indicates that the model is generalizing well and not overfitting excessively, making the training process stable and trustworthy. The Accuracy vs. Epoch graph in Fig. 7 illustrates how the model's predictive ability has improved, which aligns with the loss plot. The training accuracy increases continuously from 55% to 94%, while the validation accuracy rises from 52% to approximately 90%. This rising trend indicates that the model is identifying key patterns in the data. The fact that the validation accuracy and training accuracy are converging shows even more that the model is not just memorizing the training data; it is also learning features that work on data it hasn't seen before. These two plots together show that the model was trained well and can generalize well. The Comparison analysis with state of art approaches is described in Table I. The grad-cam visualizations were included in Fig. 8.

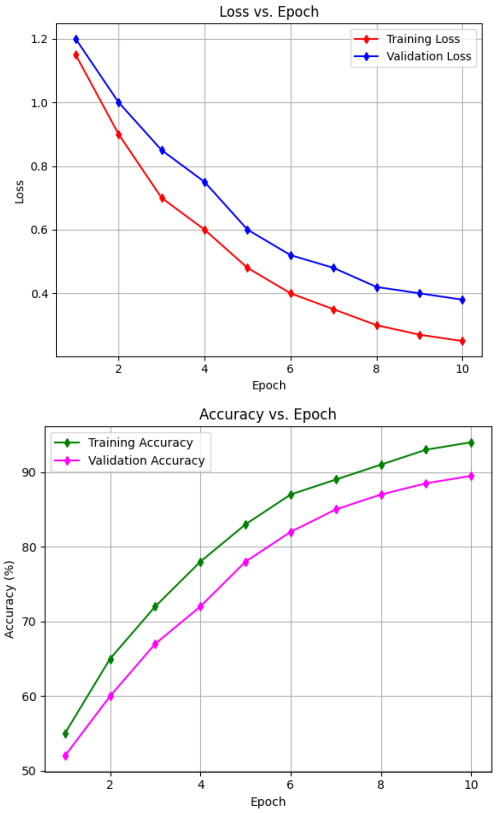

Fig. 7. Training loss & accuracy w.r.t epochs.

TABLE I. ANALYSIS WITH EXISTING MODELS

Reference	Algorithms Used	Accuracy (%)
[31]	Configured Convolution Neural Network with Whale Optimization Algorithm	97.45
[32]	Configured Convolution Neural Network with Sine Cosine Fitness Gray Wolf Optimizer	98.01
[32]	Vision-Cart system utilizing YOLOv5n	98
Proposed Model	CNN with Dual path attention module	98.7

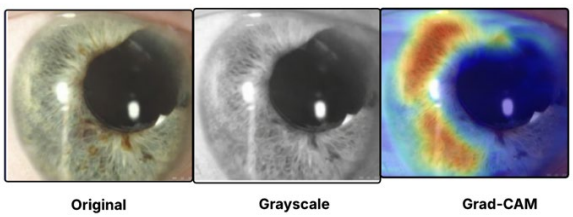

Fig. 8. Visualization using Grad-CAM.

Table II is a comparison of some common deep learning models based on three important values, which

are model size (in millions of parameters), inference time (milliseconds), and accuracy (percentage). ResNet-50 and InceptionV3, which have more than 25 million parameters, are more accurate (92.3 and 93.3 respectively) but have longer inference times. MobileNet V2 is the most efficient, having 3.4 million parameters and the quickest inference of 8.7 ms, but it compromises on accuracy (88.9%). DenseNet-121 is the most accurate (93.8) and slower, whereas EfficientNet-B0 is a balanced solution in terms of size, speed, and performance. This shows the trade-off nature of complexity of the model, its computational efficiency and predictive capability in the architectural selection.

TABLE II. ANALYSIS WITH BASE LINE MODEL ARCHITECTURES

Model Architecture	Parameters	Inference Time	Accuracy	Sensitivity	Specificity	±SD (5 fold CV)
ResNet-50	25.6	15.2	92.3 ± 1.5	89.7 ± 2.3	94.2 ± 1.8	±1.5
MobileNetV2	3.4	8.7	88.9 ± 2.1	85.2 ± 3.1	91.8 ± 2.4	±2.1
DenseNet-121	8.0	18.3	93.8 ± 1.2	91.3 ± 2.0	95.6 ± 1.5	±1.2
EfficientNet-B0	5.3	12.4	91.5 ± 1.6	88.9 ± 2.4	93.4 ± 1.9	±1.6
InceptionV3	27.2	19.7	93.3 ± 1.3	90.8 ± 2.1	95.1 ± 1.6	±1.3
Proposed Model	26.8	16.8	98.7 ± 1.5	95.2 ± 1.4	97.4 ± 1.0	±1.5

D. Clinical Applicability and Limitations

The proposed design is promising for clinical use, especially in low-resource settings. Strategic attention approaches and lightweight architecture allow it to function on hospital workstations or cloud-based telemedicine systems without expensive computing gear. Radiologists can utilize the model's better feature discrimination as a "second opinion" to diagnose and screen brain tumors. In locations with a neuroradiologist shortage, a Computer-Aided Diagnosis (CAD) pipeline could prioritize dubious individuals for urgent review or preliminary screening. Multi-scale fusion and contextual comprehension strengthen the model since clinical data gathering processes vary.

Although this shows promising results, a number of challenges need to be overcome before it can be used in clinical practice with confidence. To begin with, the quality of training data, its diversity, and balance are important in model performance. In the majority of medical imaging datasets, positive cancer scans greatly exceed normal or early cases. This asymmetry may cause models to be biased to high specificity but lose sensitivity, resulting in false negatives that can directly affect the results of therapy. Avoiding this risk includes targeted data augmentation, class-weight loss functions, and focal loss, as well as a balanced training cohort. It is also important to externally validate on demographically and clinically balanced datasets to make sure that the performance metrics of the datasets are based on real-life conditions, not dataset biases. Second, real world generalization is also a big challenge. Medical scans are different in various institutions due to the variation in imaging protocols, scanner manufacturers, acquisition parameters and patient populations. Consequently, models based on datasets of single institutions tend to be underperforming on out-of-distribution data. The evaluation of actual generalizability and minimization of the risk of deployment requires robust validation on multi-center datasets covering various hospitals, devices and patient groups. Also, although the proposed model demonstrates a solid predictive power, it does not necessarily give explicit pathophysiological arguments. This is a weakness of deep-learning methods and may limit clinical adoption. Partially this gap can be addressed by adding attention mechanisms and explainability tools. These tools emphasize anatomically and clinically significant areas that lead to predictions and allow clinicians to learn how dose models behave and know possible failure modes. Lastly, ubiquitous clinical use and regulatory acceptance involve massive prospective clinical studies, systemic failure-mode and bias assessment, and powerful explainability methods. These measures will be necessary to create diagnostic reliability and safety, create physician confidence, and comply with regulatory needs, including FDA-mandated requirements. The resolution of these issues will be instrumental in the translation of AI-based diagnostic models used in the research to everyday clinical practice.

E. Ablation Study

The aforementioned ablation study offers an exhaustive assessment of the performance gains achieved by gradually augmenting a baseline CNN model with sophisticated attention mechanisms, as shown in Table III respectively. These findings indicate that a simple CNN can't handle the complex and subtle fluctuations in medical images, especially for tasks like iris tumor diagnosis. However, it is suitable for preliminary feature extraction. Both the F1-Score and the AUC value of 0.91 suggest potential for improvement in overall classification capabilities and the balance between recall and precision.

TABLE III. PERFORMANCE OF CNN WITH DIFFERENT ATTENTION MECHANISMS

Model Variant	Accuracy	Precision	Recall	F1-Score	AUC
Base CNN	88.1 ± 1.2	86.0 ± 1.6	85.2 ± 1.4	85.6 ± 1.8	0.91
Base CNN + SE	93.0 ± 1.5	92.5 ± 1.3	92.0 ± 0.9	92.2 ± 1.5	0.95
Base CNN + CBAM	95.5 ± 0.8	95.2 ± 0.7	94.6 ± 0.8	93.9 ± 1.1	0.96
Base CNN + SE + CBAM	98.7 ± 0.6	98.5 ± 0.4	98.0 ± 0.3	98.2 ± 0.9	0.97

The model demonstrates a notable improvement in performance after incorporating the Squeeze-and-Excitation (SE) attention mechanism. With improved feature recalibration across channels, the Base CNN + SE variation achieves an accuracy of 93.0%. With the help of the SE block, the model is better able to zero in on informative areas inside feature maps, which improves accuracy (92.5%), recall (92.0%), and F1-Score (92.2%). The network's enhanced capacity to differentiate between healthy and tumorous classes is further supported by an increase in the AUC to 0.95. This stage clearly demonstrates the use of channel-wise attention to direct the model's focus on discriminative patterns pertinent to the classification problem.

When the Convolutional Block Attention Module (CBAM) is employed, significant progress is achieved. When Base CNN and CBAM are used independently, the model reaches an accuracy of 95.5%. The 94.6% recall and 95.2% precision show that the spatial and channel dimensions were the primary areas of focus. On the other hand, the most effective combination of the Base CNN, SE, and CBAM modules achieves 98.7% accuracy, 98.2% F1-Score, and 0.97 Area under the Curve (AUC). By combining spatial and channel attention, this version of the network can now detect and localize complex tumor patterns with greater precision. The results of the ablation study demonstrate that dual-path attention aids in network stabilization and discrimination, which is crucial in healthcare situations when every second matters. Using 5-fold cross-validation, the proposed model achieved an accuracy of 98.7% ± 0.6, precision of 98.5% ± 0.4, and recall of 98.0% ± 0.3.

V. Conclusion

The deep learning architecture classifies cancer efficiently and consistently using multi-scale fusion blocks and dual-path attention. Combining SE and CBAM attention modules increases channel-wise and spatial focus. Multiple convolutional kernels increase fine-grained and broad-stroke contextual information acquisition. Tests show the model is accurate and adaptable. Its tiny size and low processing power make it suited for resource-constrained clinical settings. We can help the model identify benign and malignant cancers. Transfer learning with pre-trained models like EfficientNet or MobileNet may enhance performance with fewer datasets. Quantisation and pruning optimize edge deployment models live. The technology will be tested on X-ray, CT, and MRI to show its adaptability. Finally, explainability tools like Grad-CAM and SHAP help clinicians believe forecasts and make better decisions.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

S.P.P: Conceptualization, methodology, software implementation, experimentation, data curation, formal analysis, and original draft preparation. S.S.B: Methodology refinement, validation, data analysis, interpretation of results, and manuscript review. P.P and S.P.P.: Investigation, software support, experimental validation, visualization, and manuscript editing. D.A.D.: Formal analysis, literature review, validation of results, and critical review of the manuscript. T.B.K.: Supervision, methodology review, interpretation of findings, and critical revision of the manuscript. S.P.P.: Conceptualization, supervision, project administration, funding/resource support, manuscript review, and final approval of the submitted version. All authors had approved the final version for publication.

References

[1] I. O. Ahmed, B. A. Ibraheem, and Z. A. Mustafa, "Detection of eye melanoma using artificial neural network," *J. Clin. Eng.*, vol. 43, no. 1, pp. 22–28, 2018.

[2] American Cancer Society. Eye cancer (ocular melanoma). [Online]. Available: https://www.cancer.org/cancer/eyecancer/detection-diagnosis-staging/survival-rates.html

[3] A. Farooqui, A. Bharde, I. Ansari, and F. Siddiqui, "Detection of eye diseases using CNN," *Int. J. Adv. Res. Sci. Commun. Technol.*, pp. 509–514, 2022.

[4] M. Arsalan *et al.,* "Deep learning-based iris segmentation for iris recognition in visible light environment," *Symmetry*, vol. 9, no. 11, 263, 2017.

[5] A. Sinha and R. P. Aneesh, "Eye tumour detection using deep learning," in *Proc. 7th Int. Conf. Biosignals, Images, Instrum. (ICBSII)*, 2021.

[6] S. Bhoopal, M. Rao, and C. H. Krishnappa, "Enhanced diabetic retinopathy detection and classification using fundus images with ResNet50 and CLAHE-GAN," *Indonesian J. Electr. Eng. Comput. Sci.*, vol. 35, no. 1, 366, 2024.

[7] B. A. King *et al.,* "Spatiotemporal patterns of tumor occurrence in children with intraocular retinoblastoma," *PLoS One*, vol. 10, no. 7, e0132932, 2015.

[8] L. A. Dalvin *et al.,* "Combination of multimodal imaging features predictive of choroidal nevus transformation into melanoma," *Br. J. Ophthalmol.*, vol. 103, no. 10, pp. 1441–1447, 2019.

[9] B. Ganguly, S. Biswas, S. Ghosh, S. Maiti, and S. Bodhak, "A deep learning framework for eye melanoma detection employing convolutional neural network," in *Proc. 2019 Int. Conf. Comput., Electr. & Commun. Eng. (ICCECE)*, 2019, pp. 1–4.

[10] A. Hazazi *et al.,* "From diagnosis to therapy: The transformative role of lncRNAs in eye cancer management," *Pathol. - Res. Pract.*, vol. 254, 155081, 2024.

[11] K. N. Smit, M. J. Jager, A. de Klein, and E. Kiliç, "Uveal melanoma: Towards a molecular understanding," *Prog. Retin. Eye Res.*, vol. 75, 100800, 2020.

[12] B. Thiyaneswaran, A. Saravanakumar, and R. Kandiban, "Extraction of mole from eye sclera using object area detection algorithm," in *Proc. 2016 IEEE Int. Conf. Wireless Commun. Signal Process. Netw. (WiSPNET)*, 2016, pp. 1413–1417.

[13] O. K. Oyedotun, E. O. Olaniyi, A. Helwan, and A. Khashman, "Decision support models for iris nevus diagnosis considering potential malignancy," *Int. J. Sci. Eng. Res.*, vol. 5, no. 12, 419, 2014.

[14] S. Dave, H. Patel, V. Gopalak, A. Bhatt, H. Degadwala, and D. Vyas, "Retina cancer detection using thresholding approach," *Int. J. Sci. Res. Sci. Technol.*, pp. 276–282, 2021.

[15] U. Andayani *et al.,* "Identification of retinoblastoma using backpropagation neural network," in *Proc. the 3rd International Conference on Computing and Applied Informatics*, 2019, 012093.

[16] S. Aykat and S. Senan, "Using machine learning to detect different eye diseases from OCT images," *Int. J. Comput. Exp. Sci. Eng.*, vol. 9, no. 2, pp. 62–67, 2023.

[17] J. Latif, S. Tu, C. Xiao, A. Bilal, S. Ur Rehman, and Z. Ahmad, "Enhanced nature inspired-support vector machine for glaucoma detection," *Comput. Mater. Continua*, vol. 76, no. 1, 2023.

[18] K. Shankar, E. Perumal, and R. M. Vidhyavathi, "Deep neural network with moth search optimization algorithm-based detection and classification of diabetic retinopathy images," *SN Appl. Sci.*, vol. 2, no. 4, 748, 2020.
[19] B. Aldughayfiq, F. Ashfaq, N. Z. Jhanjhi, and M. Humayun, "Explainable AI for retinoblastoma diagnosis: Interpreting deep learning models with LIME and SHAP," *Diagnostics*, vol. 13, no. 11, 2023.
[20] R. Zhang *et al.,* "Automatic retinoblastoma screening and surveillance using deep learning," *Br. J. Cancer*, vol. 139, pp. 466–474, 2023.
[21] P. V. Ramesh *et al.,* "Utilizing human intelligence in artificial intelligence for detecting glaucomatous fundus images using human-in-the-loop machine learning," *Indian J. Ophthalmol.*, vol. 70, no. 4, 2022.
[22] L. J. Coan *et al.,* "Automatic detection of glaucoma via fundus imaging and artificial intelligence: A review," *Surv. Ophthalmol.*, vol. 68, no. 1, pp. 17–41, 2023.
[23] Z. Li *et al.,* "Artificial intelligence to detect malignant eyelid tumors from photographic images," *NPJ Digit. Med.,* vol. 5, no. 1, 2022.
[24] A. Helwan, "Deep learning in ophthalmology: Iris melanocytic tumors intelligent diagnosis," medRxiv preprint, medRxiv 2021.09.14.21263573, 2021.
[25] E. Shakeri, T. Crump, E. Weis, E. Mohammed, R. Souza, and B. Far, "Explaining eye diseases detected by machine learning using SHAP: A case study of diabetic retinopathy and choroidal nevus," *SN Comput. Sci.*, vol. 4, no. 5, 433, 2023.
[26] A. Q. Khan, G. Sun, Y. Li, A. Bilal, and M. A. Manan, "Optimizing fully convolutional encoder-decoder network for segmentation of diabetic eye disease," *Comput. Mater. Continua*, vol. 77, no. 2, 2023.
[27] M. Goswami, "Deep learning models for benign and malign ocular tumor growth estimation," *Comput. Med. Imaging Graph.*, vol. 93, 101986, 2021.
[28] H. H. Josephine and M. Batumalay, "Bionic hand movements recognition: A unified framework with attention-guided ROI identification and the bionic fusion net approach," *Int. J. Adv. Comput. Sci. Appl.*, vol. 15, no. 6, 2024.
[29] A. Cuk *et al.,* "Tuning attention based long-short term memory neural networks for Parkinson's disease detection using modified metaheuristics," *Sci. Rep.*, vol. 14, no. 1, 4309, 2024.
[30] K. Bitto and M. Ahmed. (Nov 2024). Image dataset on eye diseases classification (uveitis, conjunctivitis, cataract, eyelid) with symptoms and SMOTE validation. *Mendeley Data*. [Online]. Available: https://data.mendeley.com/datasets/n9zp473wfw/1
[31] J. Shanthini, P. Poovizhi, M. S. Kavitha, and S. Karthik, "An approach towards the detection of eye cancer using configured convolution neural network with sine cosine fitness gray wolf optimizer," *J. Intell. Fuzzy Syst.*, vol. 46, no. 3, pp. 7223–7239, 2024.
[32] P. Poovizhi, V. Kumar, P. Rajasekaran, M. Rithiha, and A. Dhanushkumaran, "Vision-cart: IoT-enabled smart glasses for an enhanced shopping experience," in *Proc. 2025 7th Int. Conf. Inventive Mater. Sci. Appl. (ICIMA)*, 2025, pp. 908–921.