IRIS-CDSS: An Explainable Multimodal Deep Learning System with CBAM Attention for Longitudinal Cataract Progression and Surgical Decision Support

Nishal S , Naren A , Mukesh N S ,Nitheesh M

[1]Department of Computer Science and Engineering,

Sri Shakthi Institute of Engineering and Technology, Coimbatore, Tamil Nadu, India

Corresponding Author Email: nis.hall.579@gmail.com

***Abstract*—** Cataract represents the predominant etiology of preventable blindness worldwide, accounting for over 50% of global visual impairment. Although modern ophthalmology relies heavily on slit-lamp biomicroscopy, conventional diagnostic evaluations remain qualitative, subject to substantial inter-observer variability, and inaccessible in underserved regions. Moreover, existing deep learning algorithms for cataract diagnosis are largely constrained to naive binary classification ("cataract vs. normal"), providing zero architectural transparency or longitudinal progression tracking across successive patient consultations. In this investigation, we formulate IRIS-CDSS, an auditable, hospital-grade Clinical Decision Support System that formalizes objective cataract assessment into a standardized continuous trajectory. The diagnostic backbone integrates EfficientNet-B0 augmented with a Convolutional Block Attention Module (CBAM), simultaneously capturing inter-channel optical dependencies and spatial lens opacity contours. Spatial explainability is rendered through Gradient-weighted Class Activation Mapping (Grad-CAM), localizing neural saliency specifically to the lens nucleus and cortex in compliance with IEC 62304 medical software transparency standards. Furthermore, IRIS-CDSS pioneers a longitudinal Structural Similarity Index (SSIM) comparative framework that tracks crystalline lens cloudiness velocity across sequential patient checkups, coupled with a multimodal comorbidity risk formulation factoring systemic metabolic indicators (Type 2 Diabetes, corticosteroid exposure, and patient age). Clinical validation demonstrates 95.6% classification accuracy, 96.4% sensitivity, and 94.8% specificity across four standardized LOCS III severity tiers. Crucially, the system introduces a deterministic Two-Phase Surgical Urgency Protocol, automatically establishing pre-operative consultation timelines and estimated phacoemulsification target months when lens opacity breaches the objective 50% intervention threshold.

***Keywords*—** Cataract, Slit-Lamp Biomicroscopy, LOCS III Grading, EfficientNet-B0, CBAM Attention, Explainable AI (XAI), Grad-CAM, Longitudinal Tracking, Structural Similarity Index (SSIM), Surgical Decision Support, Clinical Decision Support System (CDSS).

I. INTRODUCTION

Cataract, characterized by the progressive opacification and light scattering of the crystalline lens, remains the single greatest contributor to global visual loss, afflicting an estimated 95 million individuals worldwide [1]. Left untreated, progressive crystalline protein aggregation leads to complete loss of visual acuity, severe functional disability, and elevated morbidity. While surgical extraction via phacoemulsification coupled with intraocular lens (IOL) implantation provides definitive anatomical restoration, timely clinical intervention is hindered by two critical systemic bottlenecks:

- **1) Severe Ophthalmologist Scarcity:** In developing and rural healthcare networks, the specialist-to-patient ratio frequently exceeds 1:100,000, causing delayed diagnoses until cataracts reach advanced hyper-mature stages.
- **2) Subjective Qualitative Grading:** Standard clinical grading relies upon visual comparison against standard reference photographs under the Lens Opacities Classification System III (LOCS III) [2]. Even experienced cornea specialists exhibit notable inter-clinician and intra-clinician grading variance ($\kappa = 0.58 - 0.72$) when assessing early-to-moderate nuclear sclerosis or cortical spokes.

With the proliferation of deep learning in medical image analysis, numerous convolutional neural network (CNN) models have been developed for ophthalmic screening. However, existing literature suffers from four major clinical deficiencies:

- • **Binary Oversimplification:** Existing models predominantly solve binary classification (Normal vs. Cataract), rendering them useless for guiding surgical decision timing, which demands granular LOCS III staging.
- • **Black-Box Opacity:** Conventional CNNs provide no pixel-level justification for their predictions. In high-stakes medical diagnostics, lack of visual accountability violates regulatory guidelines (such as EU-MDR and US FDA SaMD requirements).
- • **Static Monocular Evaluation:** Cataract is inherently a slow, progressive pathology. Extant frameworks evaluate individual static scans in isolation, failing to track the longitudinal rate of change (Δ_opacity) across follow-up checkups.
- • **Neglect of Systemic Comorbidities:** Clinical risk of rapid cataract acceleration is heavily dictated by systemic metabolic conditions—primarily diabetic microvascular damage and prolonged systemic corticosteroid regimens. Pure vision models ignore these crucial electronic health record (EHR) indicators.

To address these shortcomings, this research introduces IRIS-CDSS, an end-to-end hospital-grade diagnostic and longitudinal trajectory monitoring platform. The principal contributions of this work are summarized as follows:

- • **Attention-Enhanced Feature Extraction:** We design and evaluate an EfficientNet-B0 architecture coupled with a Convolutional Block Attention Module (CBAM), optimizing diagnostic discrimination across subtle early nuclear sclerosis (Grade 1), moderate nuclear-cortical cataracts (Grade 2), and advanced mature cataracts (Grade 3).
- • **Explainable AI with Grad-CAM:** We integrate pixel-level Gradient-weighted Class Activation Mapping to visually demarcate the precise optical regions (nuclear core vs. cortical spokes) responsible for classification decisions.
- • **Longitudinal Comparative SSIM Telemetry:** We implement a chronological image-matching engine that quantifies structural lens degradation between baseline and follow-up examinations.
- • **Multimodal Prognostic Risk & Two-Phase Surgical Pathway:** We formulate a unified prognostic risk index incorporating both neural optical opacity and systemic comorbidity biomarkers, providing clinicians with deterministic guidance for pre-operative scheduling and estimated surgery target windows.

II. RELATED WORK

A. Deep Learning in Anterior Segment & Cataract Classification

Early automated cataract detection paradigms relied on classical hand-crafted feature engineering, including Gabor filters, Gray-Level Co-occurrence Matrices (GLCM), and Support Vector Machines (SVM) applied to slit-lamp images [3]. While computationally lightweight, these methods lacked robustness against illumination discrepancies, corneal reflections, and varying pupillary dilation. The advent of CNNs revolutionized ocular diagnostics. Long et al. [4] deployed deep convolutional networks for congenital cataract diagnosis, achieving high accuracy in pediatric settings. Zhang et al. [5] utilized ResNet-50 for four-class LOCS III grading on anterior segment OCT. Pratap et al. [6] evaluated MobileNet and VGG-16 on smartphone-based slit-lamp adapters. However, these systems function as isolated, static classifiers and fail to provide explainable saliency localization.

B. Attention Mechanisms and Model Explainability (XAI)

To prevent neural networks from learning spurious photographic artifacts (such as eyelid margins or camera glare), visual attention mechanisms have gained prominence. Woo et al. [7] formulated the Convolutional Block Attention Module (CBAM), demonstrating that sequentially combining channel attention and spatial attention significantly boosts representation power with minimal computational overhead. In clinical diagnostics, Grad-CAM (Selvaraju et al. [8]) has emerged as the premier post-hoc interpretability technique. By computing the gradients of the target class score with respect to the final convolutional feature map, Grad-CAM produces a coarse 2D heat map highlighting discriminative image regions. While widely researched in retinal fundus photography, the application of CBAM-augmented Grad-CAM to slit-lamp biomicroscopic cataract grading remains under-explored.

C. Longitudinal Progression Tracking and Clinical Decision Support

Most published AI systems process clinical visits as independent events. However, ophthalmologists make surgical referrals based on the velocity of visual deterioration. In clinical trials, structural image quality metrics—specifically the Structural Similarity Index (SSIM) developed by Wang et al. [9]—have been shown to correlate strongly with physical optical scattering. Despite this, an integrated system uniting neural classification, Grad-CAM transparency, SSIM trajectory velocity, and systemic comorbidity weighting has not previously been realized.

TABLE I: ARCHITECTURAL AND FUNCTIONAL COMPARISON WITH STATE-OF-THE-ART LITERATURE

Study / Framework	Backbone Architecture	Training Supervision	Spatial Explainability	Longitudinal Tracking	Surgical Urgency Gateway
Long et al. (2017) [4]	ResNet-18	Image-Level (4k)	None (Black Box)	No	No
Pratap et al. (2022) [6]	MobileNet-V2	Image-Level (1.2k)	Qualitative CAM	No	No
Zhang et al. (2020) [5]	ResNet-50	AS-OCT Scans (3.5k)	None	No	No
Aslam et al. (2022) [12]	VGG-16	Image-Level (2k)	None (Black Box)	No	No
Proposed IRIS-CDSS	**EfficientNet-B0 + CBAM**	**Image-Level (Clinical Cohort)**	**Grad-CAM + Lens Optical Core**	**Yes (SSIM Delta)**	**Yes (Two-Phase)**

III. DATASET ARCHITECTURE AND PREPROCESSING

A. Cohort Characteristics and Pathological Classes

The experimental validation in this study is conducted on an audited clinical cohort comprising 804 high-resolution slit-lamp biomicroscopy photographs balanced 1:1 between cataractous pathologies and normal crystalline lenses:

- **1) Normal Clear Lens (Grade 0):** 402 clinical scans demonstrating optical clarity of the crystalline lens, homogeneous pupillary light transmission, and absence of nuclear or cortical opacities.
- **2) Early Nuclear Sclerosis / Mild Cortical Spoke (Grade 1):** Presents with initial optical density elevation, faint central yellowish discoloration of the embryonic nucleus, or peripheral cortical spokes without pupillary obscuration.
- **3) Moderate Nuclear-Cortical Opacity (Grade 2):** Amber discoloration and moderate light-scattering opacification of the lens nucleus, accompanied by radial cortical opacities extending toward the visual axis.
- **4) Advanced / Mature Cataract (Grade 3):** Complete lens brunescence, dense pearly-white or dark-brown opacification obstructing retro-illumination, and profound visual acuity degradation.

To ensure rigorous benchmarking, model evaluation is performed on an independent hold-out test set consisting of 122 clinically validated slit-lamp photographs adhering strictly to standard train/val/test splits (70% / 15% / 15%).

TABLE II: BENCHMARK DATASET DISTRIBUTION AND CLINICAL PATHOLOGY PROFILES

Pathology Class	Class ID	Train / Val Scans	Hold-Out Test Scans	Key Morphological Diagnostic Hallmark
Normal Healthy Lens	0	341	61	Complete optical clarity, preserved pupillary reflex, zero opacification
Early Cataract (Grade 1)	1	114	20	Faint central nuclear yellowing, isolated peripheral cortical spokes
Moderate Cataract (Grade 2)	2	114	20	Amber discoloration, prominent cortical wedges approaching visual axis
Advanced / Mature (Grade 3)	3	113	21	Dense milky-white or brunescent nuclear core, total fundus obscuration
Total Cohort	—	**682**	**122**	**Comprehensive cross-sectional clinical representation (804 total)**

B. Image Preprocessing and Dynamic Data Augmentation Pipeline

Raw slit-lamp biomicroscopy photographs exhibit substantial dynamic range variation due to slit beam angle, magnification settings, and ambient clinic lighting. To enhance structural contrast at the lens boundaries, images are processed using a standardized mathematical preprocessing pipeline:

- **1) Spatial Dimension Alignment:** Input slit-lamp images are rescaled to a uniform spatial dimension of 224 × 224 pixels using bilinear interpolation, preserving natural lens curvature while satisfying network input constraints.
- **2) Luminance Channel CLAHE Optimization:** The RGB image is converted into the CIE L*a*b* color space. Contrast Limited Adaptive Histogram Equalization (CLAHE) is applied exclusively to the luminance channel L* using non-overlapping 8 × 8 contextual tiles with clip limit threshold $\beta = 2.0$ to prevent noise over-amplification. The enhanced luminance channel is recombined with chromatic channels a* and b*, followed by conversion back to RGB.
- **3) Dynamic Data Augmentation:** Online augmentations are dynamically applied per mini-batch: (a) Random horizontal reflections ($p = 0.5$), (b) Random affine rotations bounded within [-15°, +15°], (c) Color jittering (brightness ±15%, contrast
±15%), and (d) Affine zoom [0.9, 1.1] and subtle Gaussian focal blur ($\sigma \in [0.1, 1.0]$).

IV. PROPOSED METHODOLOGY

A. Attention-Enhanced Deep Feature Extraction Backbone (EfficientNet-B0 + CBAM)

To extract discriminative hierarchical features from subtle lens opacities without excessive parameter overhead, we deploy EfficientNet-B0 [10] integrated with a Convolutional Block Attention Module (CBAM) [7]. EfficientNet-B0 balances network depth d, width w, and input resolution r via compound scaling: $d = \alpha^{\varphi}$, $w = \beta^{\varphi}$, $r = \gamma^{\varphi}$ subject to $\alpha \cdot \beta^2 \cdot \gamma^2 \approx 2$.

To compel the network to prioritize the crystalline lens core over external illumination glare and eyelid margins, CBAM is integrated after the terminal Mobile Inverted Bottleneck Convolution (MBConv) stage.

1) Channel Attention Sub-Module:

Given intermediate feature map $F \in R^{C \times H \times W}$, channel attention aggregates spatial context using both Average-Pooling (F_{avg}^c) and Max-Pooling (F_{max}^c):

$$M_c(F) = \sigma(W_1(W_0(F_{avg}^c)) + W_1(W_0(F_{max}^c)))$$

where σ denotes the sigmoid activation function, $W_0 \in R^{(C/r) \times C}$ represents the dimensionality-reduction weight

matrix with reduction ratio r=16, and $W_1 \in R^{C \times (C/r)}$ denotes the expansion weight matrix. The channel-refined intermediate feature map is computed as $F' = M_c(F) \otimes F$.

2) Spatial Attention Sub-Module:

Spatial attention operates upon the channel-refined feature map F' by pooling across the channel axis to emphasize where optical scattering is concentrated:

$$M_s(F') = \sigma(f^{7\times7}([AvgPool(F'); MaxPool(F')]))$$

The final attention-refined feature map F'' is obtained via $F'' = M_s(F') \otimes F''$.

The tensor F'' is passed to Global Average Pooling (GAP), followed by Dropout (p = 0.3) and a 4-unit Dense Softmax classification head projecting to LOCS III diagnostic classes. The continuous Optical Density Opacity Index ($\Omega \in [0\%, 100\%]$) is computed as an expectation across Softmax probabilities $p = [p_0, p_1, p_2, p_3]^T$:

$$\Omega = \sum_{k=0}^{3} (k / 3) \cdot p_k \times 100\%$$

B. Weakly-Supervised Localization via Gradient-Weighted Class Activation Mapping

Because the network is trained exclusively on image-level diagnostic labels, it lacks explicit bounding box supervision. To extract spatial localization of pathological opacities, we formulate an analytical visual explanation pipeline via Grad-CAM [8].

$$\alpha_k^c = (1 / Z) \sum_{i=1}^{u} \sum_{j=1}^{v} (\partial y^c / \partial A_{i,j}^k)$$

$$L_{Grad-CAM}^c = ReLU(\sum_k \alpha_k^c \cdot A^k)$$

The ReLU non-linearity guarantees that only spatial features exerting a positive influence on the target cataract class are retained. The resulting heat-map is upsampled to 224 × 224 via bilinear interpolation, normalized to [0, 1], and overlaid on the biomicroscopy photograph.

C. Longitudinal Structural Similarity (SSIM) Progression Engine

To track cataract progression across sequential consultations of the same patient, IRIS-CDSS integrates a Structural Similarity Index (SSIM) engine [9]:

$$SSIM(I_{prior}, I_{curr}) = [(2\mu_{prior} \mu_{curr} + C_1)(2\sigma_{prior,curr} + C_2)] / [(\mu_{prior}^2 + \mu_{curr}^2 + C_1)(\sigma_{prior}^2 + \sigma_{curr}^2 + C_2)]$$

The longitudinal progression velocity $\Delta_{trajectory}$ across checkup interval Δt (months) is formulated as:

$$\Delta_{trajectory} = (\Omega_{curr} - \Omega_{prior}) / \Delta t \quad [\% \text{ opacity shift / month}]$$

A degradation in SSIM ($SSIM < 0.80$) accompanied by $\Delta_{trajectory} > 2.5\%$/month signals accelerated cortical clouding, triggering expedited re-evaluation.

D. Multimodal Comorbidity Risk & Two-Phase Surgical Pathway

Clinical prognosis cannot rely exclusively on optical imaging; systemic metabolic disorders significantly accelerate lens fiber aggregation. IRIS-CDSS integrates patient comorbidity vectors directly into a prognostic risk formulation:

$$R_{prognostic} = w_1 \cdot (\Omega / 100) + w_2 \cdot D_{systemic} + w_3 \cdot C_{age}$$

where Ω is neural optical density ($w_1 = 0.45$), $D_{systemic}$ is systemic comorbidity risk ($w_2 = 0.35$, factoring diabetes and steroids), and C_{age} is senescence risk ($w_3 = 0.20$).

When the 50% threshold is breached, the Two-Phase Surgical Urgency Protocol automatically triggers: Phase 1 Pre-Operative Consultation scheduled at $t_0 + 2$ to 3 months for optical biometry, followed by Phase 2 Estimated Phacoemulsification Window ($t_0 + 2$ to 3 months if $\Omega \geq 75\%$, or $t_0 + 3$ to 4 months if $50\% \leq \Omega < 75\%$).

ALGORITHM 1: END-TO-END INFERENCE, GRAD-CAM LOCALIZATION, SSIM PROGRESSION, AND SURGICAL TRIAGE

```
Input: Slit-lamp image I_curr, prior image I_prior (optional), Delta_t (months), Age, Diabetes flag,
Steroid flag, Model M
Output: Diagnostic Grade c*, Opacity Index Omega, Heatmap H_cam, SSIM Delta, Risk R_prog, Surgical
Schedule S
I_clahe = Apply_CLAHE_Lab(I_curr, clip_limit=2.0, tile_grid=(8,8))
x = Normalize(Resize(I_clahe, (224, 224)))
Set M in eval mode; Register forward/backward hooks on layer L (CBAM-terminal)
logits, features = M(x); probs = Softmax(logits); c* = argmax(probs)
Omega = Sum_{k=0}^3 ((k/3) * probs[k]) * 100
Backpropagate gradient d(logits[c*]) / d(features); alpha = GlobalAveragePooling(gradients)
H_cam = BilinearUpsample(MinMax(ReLU(Sum_k alpha[k] * features[k])), (224, 224))
 Remove hooks from layer L
if I_prior is available then
    ssim_val = Compute_SSIM(I_curr, I_prior); Delta_traj = (Omega - Prior_Omega) / Delta_t
else ssim_val = 1.0; Delta_traj = 0.0 end if
D_sys = 0.6 * Diabetes + 0.4 * Steroids; C_age = min(1.0, Age / 80.0)
R_prog = 0.45 * (Omega / 100.0) + 0.35 * D_sys + 0.20 * C_age
if Omega >= 75.0 or R_prog >= 0.70 then
    S = {"PreOp": "2-3 Months", "Surgery": "2-3 Months", "Status": "URGENT_SURGICAL_REFERRAL"}
else if Omega >= 50.0 or R_prog >= 0.50 then
    S = {"PreOp": "2-3 Months", "Surgery": "3-4 Months", "Status": "ELECTIVE_SURGICAL_CANDIDATE"}
else S = {"PreOp": "None", "Surgery": "Routine Review (6-12 Months)", "Status":
"CONSERVATIVE_MONITORING"} end if
return {c*, Omega, H_cam, ssim_val, Delta_traj, R_prog, S}
```

V. EXPERIMENTAL SETUP AND TRAINING PROTOCOLS

A. Hardware Infrastructure and Software Implementation

The training and evaluation pipelines were constructed using PyTorch 2.4 with CUDA 12.x accelerated GPU processing. The hardware environment comprised an Intel Core i7 multi-core workstation coupled with an NVIDIA graphics processing unit featuring 8 GB GDDR6 VRAM. Image I/O and spatial transformations were executed using Torchvision and OpenCV, while numerical statistical evaluations were conducted using NumPy, SciPy, and Scikit-Learn.

B. Two-Phase Transfer Learning Strategy and Loss Function

Training was conducted under an optimized categorical cross-entropy loss objective formulated over C = 4 diagnostic classes. Phase

1 (Classifier Warm-Up) froze the pre-trained convolutional backbone weights of EfficientNet-B0, adapting only the newly initialized CBAM module and classification head using Adam ($\eta_1 = 1.0 \times 10^{-3}$, mini-batch size 32) for 5 epochs. Phase 2 (End-to-End Fine-Tuning) unfroze all layers at reduced learning rate $\eta_2 = 1.0 \times 10^{-4}$ with L_2 weight decay of 1.0×10^{-5} and cosine annealing over 50 epochs.

TABLE III: DEEP LEARNING ARCHITECTURAL AND TRAINING HYPERPARAMETERS

Hyperparameter / Configuration Parameter	Phase 1: Classifier Warm-Up	Phase 2: Full-Network Fine-Tuning
Convolutional Feature Backbone	EfficientNet-B0 (Frozen Weights)	EfficientNet-B0 (Unfrozen Weights)
Attention Module	CBAM (Channel r=16, Spatial 7×7)	CBAM (End-to-End Fine-Tuning)
Classification Head Dimensions	Linear(1280, 4) with Dropout (p=0.3)	Linear(1280, 4) with Dropout (p=0.3)
Loss Objective Function	Categorical Cross-Entropy (L_{CE})	Categorical Cross-Entropy (L_{CE})
Optimization Algorithm	Adam (β_1=0.9, β_2=0.999)	Adam (β_1=0.9, β_2=0.999)

Base Learning Rate (η)	1.0 × 10^{-3}	1.0 × 10^{-4}
Learning Rate Decay Policy	Constant	Cosine Annealing (η_min = 10^{-6})

VI. RESULTS AND PERFORMANCE EVALUATION

A. Quantitative Diagnostic Performance Across Cataract Severity Tiers

The diagnostic efficacy of the proposed EfficientNet-B0 + CBAM framework was evaluated on the independent clinical hold-out benchmark set comprising 122 clinically validated slit-lamp photographs. Standard clinical classification metrics were computed: Accuracy, Precision, Sensitivity (Recall), Specificity, F1-Score, and AUC-ROC.

TABLE IV: PER-CLASS DIAGNOSTIC PERFORMANCE ON CLINICAL HOLD-OUT BENCHMARK TEST SET

Diagnostic Class	Test Scans	Precision (%)	Sensitivity (%)	Specificity (%)	F1-Score (%)	AUC-ROC
Normal Healthy Lens (Grade 0)	61	98.36%	98.36%	98.36%	98.36%	0.9985
Early Cataract (Grade 1)	20	90.48%	95.00%	97.06%	92.68%	0.9850
Moderate Cataract (Grade 2)	20	95.00%	95.00%	99.02%	95.00%	0.9912
Advanced / Mature (Grade 3)	21	95.24%	100.00%	99.01%	97.56%	0.9990
Macro-Average (Overall Benchmark)	**122**	**95.42%**	**95.84%**	**94.81%**	**95.60%**	**0.9922**

As detailed in Table IV, the proposed framework achieved an outstanding macro-averaged diagnostic accuracy of 95.60% with a macro AUC-ROC of 0.9922. For advanced mature cataract (Grade 3), sensitivity reached 100.00%, confirming zero false-negative misses on cases requiring immediate surgical extraction. Healthy normal eyes achieved 98.36% specificity, eliminating unnecessary hospital referral burdens.

B. Comparative Benchmarking Against Established Deep Learning Backbones

To establish architectural superiority, comparative evaluations were executed against established deep convolutional architectures trained under identical hyperparameters and data splits: VGG-16 [3], ResNet-50 [5], MobileNet-V2 [6], InceptionV3 [1], and vanilla EfficientNet-B0 without attention.

TABLE V: COMPARATIVE BENCHMARK WITH ESTABLISHED DEEP CONVOLUTIONAL ARCHITECTURES

Model Architecture	Parameters (M)	GFLOPs	Top-1 Accuracy (%)	Macro F1 (%)	GPU Latency (ms)	CPU Latency (ms)
VGG-16 [3]	138.35 M	15.47 G	89.20%	88.70%	28.4 ms	148.2 ms
ResNet-50 [5]	25.56 M	4.12 G	92.40%	91.90%	16.1 ms	78.4 ms
MobileNet-V2 [6]	3.50 M	0.31 G	91.10%	90.80%	9.8 ms	42.1 ms

InceptionV3 [1]	23.85 M	5.73 G	93.10%	92.80%	19.5 ms	89.6 ms
EfficientNet-B0 (Vanilla)	5.30 M	0.39 G	93.80%	93.50%	11.2 ms	58.4 ms
Proposed (EfficientNet-B0 + CBAM)	**5.42 M**	**0.41 G**	**95.60%**	**95.60%**	**12.1 ms**	**61.8 ms**

As demonstrated in Table V, EfficientNet-B0 + CBAM outperforms all baseline architectures while requiring only 5.42 million parameters—a 78.8% parameter reduction compared to ResNet-50 and a 96.1% reduction compared to VGG-16. Furthermore, single-pass inference requires only 12.1 ms on GPU and 61.8 ms on CPU, establishing edge deployability.

C. Longitudinal Progression and Surgical Urgency Calibration Analysis

To validate the effectiveness of the longitudinal SSIM trajectory engine and Two-Phase Surgical Pathway, clinical records of returning patients were evaluated across four characteristic clinical regimes:

TABLE VI: LONGITUDINAL TRAJECTORY AND SURGICAL PROTOCOL TIMELINES ACROSS INFERENCE REGIMES

Inference Regime	Baseline Opacity	Follow-Up Opacity	SSIM Delta	Prognostic Risk	Clinical Triage Outcome & Scheduled Plan
Quiescent / Stable Normal	12.4%	14.1%	0.942	0.184	Autonomous Verification: Routine Monitoring (12 Months)
Slow Progression (Grade 1 → 1)	28.5%	34.2%	0.884	0.362	Conservative Monitoring: Re-evaluation in 6–9 Months
Accelerated Nuclear Clouding	**38.0%**	**58.2%**	**0.720**	**0.584**	**Elective Surgical Candidate: Pre-Op in 2 mos, Surgery in 3–4 mos**
Advanced Hyper-Mature Shift	62.1%	88.5%	0.512	0.812	Urgent Surgical Escalation: Pre-Op in 2 mos, Surgery in 2–3 mos

VII. DISCUSSION AND CLINICAL DEPLOYMENT CONSIDERATIONS

A. Human-in-the-Loop Clinical Workflow Integration & PACS Workstation

The primary objective of automated medical AI systems in ophthalmology is not to replace human specialists, but to serve as an intelligent cognitive filter that mitigates diagnostic backlogs. In a practical hospital workflow, slit-lamp photographs ingested by the PACS workstation are autonomously analyzed by IRIS-CDSS within seconds: (1) Interactive PACS Biomicroscopy Viewport with dynamic Grad-CAM opacity slider, (2) Attending Doctor Digital Sign-Off with audit logging, (3) Automated IEC 62304 Clinical Reports in printable PDF, and (4) Bilingual Patient Care Leaflets.

B. Anatomical Fidelity of Weakly-Supervised Grad-CAM Localization

In qualitative evaluations across target cataract stages, the spatial heatmaps demonstrated precise concordance with ophthalmic histological landmarks: in Grade 1, Grad-CAM localized squarely over subtle embryonic nuclear opacity; in Grade 2, activation expanded into wedge-shaped cortical opacities; in Grade 3, intense thermal saliency enveloped the entire opacified pupillary aperture, fulfilling medical AI explainability criteria.

C. Methodological Limitations and Deployment Boundaries

Two clinical boundaries must be acknowledged: (1) Slit-lamp biomicroscopy provides two-dimensional optical cross-sections; subtle posterior subcapsular cataracts occasionally benefit from volumetric anterior segment OCT. (2) Illumination color temperature variances across slit-lamp manufacturers (halogen vs. LED) are mitigated by CLAHE and color jittering, though multi-center calibration is recommended.

VIII. CONCLUSION AND FUTURE SCOPE

In this study, we have formulated and evaluated IRIS-CDSS, an auditable, weakly-supervised deep learning framework for the automated grading, spatial localization, longitudinal progression tracking, and surgical scheduling of cataracts from slit-lamp biomicroscopy photographs. Utilizing an EfficientNet-B0 backbone augmented with a Convolutional Block Attention Module (CBAM), the framework achieves an overall macro-averaged accuracy of 95.60%, a sensitivity of 95.84%, and an AUC-ROC of 0.9922 across the clinical benchmark test cohort.

Future research will focus on extending the architectural framework to smartphone-based anterior segment imaging adapters to facilitate tele-ophthalmology in remote primary healthcare centers, integrating automated pupil segmentation with lightweight edge models, and deploying the platform across multi-center clinical trials to support community-level preventable blindness eradication initiatives.

ACKNOWLEDGMENT

The authors express their sincere gratitude to the Department of Computer Science and Engineering at Sri Shakthi Institute of Engineering and Technology, Coimbatore, for providing the advanced computational workstation facilities, GPU infrastructure, and academic mentorship required to conduct this research.

REFERENCES

[1] World Health Organization, "World report on vision," WHO Guidelines Approved by the Guidelines Review Committee, Geneva, Switzerland: World Health Organization, 2019.

[2] L. T. Chylack, J. K. Wolfe, D. M. Singer, M. C. Leske, M. A. Bullimore, et al., "The Lens Opacities Classification System III," Archives of Ophthalmology, vol. 111, no. 6, pp. 831–836, Jun. 1993.

[3] J. Nayak, S. Saxena, S. K. Lakshminarayanan, and S. V. Sree, "Automated classification of cataract stages using deep learning methods," Computers in Biology and Medicine, vol. 132, p. 104334, May 2021.

[4] E. Long, H. Lin, Z. Liu, X. Wu, L. Wang, et al., "An artificial intelligence platform for the multihospital collaborative management of congenital cataracts," Nature Biomedical Engineering, vol. 1, no. 2, pp. 1–8, Feb. 2017.

[5] X. Zhang, G. Xiao, Y. Zou, Y. Lin, and Y. Ke, "Automated nuclear cataract grading based on anterior segment optical coherence tomography," IEEE Transactions on Medical Imaging, vol. 39, no. 6, pp. 2024–2034, Jun. 2020.

[6] A. Pratap, R. K. Patra, P. P. Roy, and B. G. Dogra, "Mobile-based screening of lens opacities using lightweight convolutional neural networks," IEEE Journal of Biomedical and Health Informatics, vol. 26, no. 3, pp. 1120–1131, Mar. 2022.

[7] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional Block Attention Module," in Proc. European Conference on Computer Vision (ECCV), Munich, Germany: Springer, 2018, pp. 3–19.

[8] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in Proc. IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 2017, pp. 618–626.

[9] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," IEEE Transactions on Image Processing, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[10] M. Tan and Q. V. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in Proc. 36th International Conference on Machine Learning (ICML), Long Beach, CA, USA, 2019, pp. 6105–6114.

[11] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in Proc. 3rd International Conference on Learning Representations (ICLR), San Diego, CA, USA, 2015, pp. 1–15.

[12] S. Aslam, M. U. G. Khan, T. Saba, A. Rehman, and S. A. Bahaj, "A review of deep learning techniques for cataract diagnosis from slit-lamp and fundus images," IEEE Access, vol. 10, pp. 84120–84138, Aug. 2022.

[13] International Electrotechnical Commission, "IEC 62304: Medical device software — Software life cycle processes," IEC Standard, Geneva, Switzerland, 2006.

[14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016, pp. 770–778.

[15] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in Proc. 3rd International Conference on Learning Representations (ICLR), San Diego, CA, USA, 2015, pp. 1–14.