A Comprehensive Review of Lightweight and Attention-Driven Deep Learning Models for Automated Cataract Detection

Bidhun B, Deepak Dayanandan, Joel Joy, Vargheese Francis, Asst. Prof. Vani V Prakash
Department of Computer Science and Engineering
Carmel College of Engineering and Technology, Alappuzha, Kerala, India

***Abstract*—Cataract is the leading cause of reversible blindness globally, accounting for nearly 51% of all blindness cases according to the World Health Organization (WHO). Traditional diagnostic procedures such as slit-lamp examination and ophthalmoscopy require expert supervision and expensive imaging devices, limiting their accessibility in rural and low-resource regions. Artificial Intelligence (AI) and Deep Learning (DL) have emerged as transformative technologies that can automate cataract detection from ocular images, enabling early diagnosis through mobile and edge devices. This review provides a comprehensive synthesis of recent research on lightweight and attention-driven deep learning frameworks for cataract detection. It critically evaluates four cornerstone approaches: the Optimised Lightweight Deep Edge Intelligent Model (SDLM), CNN-based cataract severity detection, Global–Local Attention Augmented Models (GLAAM and GLAAI), and MobileNet-based transfer learning. We present an extensive comparative analysis covering datasets, architectures, accuracy, computational efficiency, and deployment feasibility. Furthermore, this review explores interpretability techniques such as Grad-CAM and attention visualization that enhance the transparency of AI systems. The paper concludes by identifying emerging research directions, challenges, and opportunities toward federated, explainable, and globally accessible cataract detection systems.**

***Index Terms*—Cataract detection, Deep learning, Attention mechanisms, Lightweight CNN, Edge AI, MobileNet, Medical imaging, Explainable AI.**

I. Introduction

Cataract is a progressive clouding of the eye's natural lens that leads to blurred vision, glare sensitivity, and eventual blindness if untreated. It is responsible for over 100 million disability-adjusted life years (DALYs) globally, making it one of the most critical public health issues in ophthalmology. Although modern surgical techniques such as phacoemulsification can fully restore vision, early detection remains a major bottleneck in cataract management, particularly in developing countries where ophthalmic infrastructure is limited.

Manual cataract grading and diagnosis rely on experienced ophthalmologists interpreting slit-lamp or fundus images. These methods are time-consuming, subjective, and often inconsistent between graders. To address these challenges, automated image-based diagnostic systems have gained traction, leveraging advancements in computer vision and deep learning.

A. Motivation for AI-Based Cataract Detection

Deep learning models, especially Convolutional Neural Networks (CNNs), have demonstrated exceptional success in visual recognition tasks such as object classification and facial recognition. Their capacity to automatically learn hierarchical visual features from large datasets has extended naturally to the medical imaging domain. In ophthalmology, CNNs have been used for detecting diabetic retinopathy, glaucoma, and age-related macular degeneration. However, the application of CNNs to cataract detection poses unique challenges due to:

- Variability in imaging modalities (slit-lamp, fundus, AS-OCT).
- Inconsistency in illumination and image focus.
- Limited availability of annotated medical datasets.

The development of **lightweight deep learning models** has been crucial for enabling real-time cataract detection on edge devices. These architectures optimize computational parameters without compromising diagnostic accuracy, ensuring efficient deployment on mobile phones, embedded systems, and portable ophthalmic devices.

B. Evolution of Automated Detection Systems

The evolution of cataract detection research can be broadly divided into three generations:

1) **Traditional Feature-Based Systems (2005–2015):** Early attempts relied on handcrafted features such as gray-level co-occurrence matrices (GLCM), intensity histograms, and wavelet transformations. Classifiers like SVM and KNN were used, but their accuracy rarely exceeded 80–85%.
2) **Deep CNN-Based Models (2016–2021):** The advent of CNNs revolutionized cataract image analysis. Custom CNNs trained on fundus and slit-lamp datasets achieved high accuracy but were computationally expensive, requiring GPUs for inference.
3) **Lightweight and Attention-Driven Architectures (2022–Present):** Modern models integrate efficiency and interpretability. MobileNet, EfficientNet, and attention-augmented networks such as GLAAM and GLAAI achieve high accuracy with low computational overhead, suitable for point-of-care diagnostics.

DOI:10.5281/zenodo.19955429

C. Scope and Objectives

This review paper consolidates the latest progress in cataract detection frameworks, focusing on:

- Evaluating lightweight CNN and transfer learning models.
- Analyzing attention mechanisms for enhanced feature localization.
- Comparing architecture complexity, model accuracy, and deployment scalability.
- Discussing interpretability, ethical issues, and real-world challenges.

The remainder of this paper is organized as follows: Section II provides foundational background on AI in medical imaging. Section III presents an in-depth literature review of four key works forming the backbone of this study. Section IV compares their datasets and methodologies. Section V discusses performance trends, limitations, and interpretability methods. Section VI outlines future research directions, and Section VII concludes the paper.

II. Foundations of Deep Learning in Medical Imaging

Deep learning is a subset of machine learning inspired by the structure and function of the human brain. CNNs form the backbone of most modern image-based diagnostic systems. A CNN typically consists of convolutional, pooling, and fully connected layers that progressively extract hierarchical features. Early layers capture edges and textures, while deeper layers learn more complex representations like shapes and anatomical patterns.

In medical imaging, CNNs have demonstrated state-of-the-art performance in segmentation, classification, and detection tasks. Architectures such as U-Net, ResNet, and DenseNet have become foundational models across various medical domains.

A. Transfer Learning Paradigm

Transfer learning allows a model pre-trained on a large-scale dataset (e.g., ImageNet) to adapt to a smaller domain-specific dataset through fine-tuning. This technique is particularly beneficial in healthcare, where data scarcity is a recurring challenge. In cataract detection, pre-trained architectures such as VGG16, InceptionV3, and MobileNet have been successfully adapted to classify slit-lamp and fundus images.

B. Edge AI and Mobile Deployments

Edge AI involves performing inference directly on local devices, minimizing reliance on cloud infrastructure. Lightweight models such as MobileNet, ShuffleNet, and EfficientNet-Lite are optimized for low latency, reduced model size, and minimal energy consumption—critical features for telemedicine and remote diagnostics.

Neogi et al. introduced one of the earliest successful edge AI frameworks for cataract detection—the Optimised Lightweight Deep Edge Intelligent Model (SDLM). It achieved 93.44% accuracy while maintaining a model size under 200KB. The ability to run on Android smartphones without internet connectivity demonstrated the transformative potential of mobile ophthalmology.

III. Comprehensive Literature Review

The literature on cataract detection through artificial intelligence can be categorized into four major research paradigms: (1) conventional CNN architectures; (2) lightweight edge-optimized frameworks; (3) attention-driven hybrid models; and (4) transfer-learning-based compact systems. This section discusses the most representative works in each category, emphasizing architectural design, dataset characteristics, evaluation metrics, and clinical feasibility.

A. CNN-Based Cataract Severity Detection (Yadav et al., 2023)

Yadav et al. [2] presented one of the earliest fully automated systems for cataract severity grading using deep CNNs. Their dataset consisted of 1,600 fundus images captured under clinical supervision, categorized into mild, moderate, and severe cataracts. The proposed network incorporated convolutional layers with kernel sizes of 3×3, followed by max-pooling and dropout regularization to mitigate overfitting.

The model was trained using a learning rate of 0.003 with the Adam optimizer over 50 epochs. The study achieved an overall accuracy of 98.3%, outperforming conventional SVM- and KNN-based classifiers by over 15%. Importantly, the authors introduced a visual severity scale aligned with the Lens Opacities Classification System (LOCS III), thereby establishing a link between quantitative image analysis and ophthalmic grading standards.

The strengths of this work include high diagnostic precision, integration of clinical grading criteria, and use of real slit-lamp imagery. However, the model's large parameter count (28 MB) limited its deployment on low-resource devices, prompting subsequent research on network compression and optimization.(Fig:1)

B. Optimised Lightweight Deep Edge Intelligent Model (SDLM) (Neogi et al., 2024)

To overcome computational constraints, Neogi et al. [1] developed the Optimised Lightweight Deep Edge Intelligent Model (SDLM), specifically engineered for edge inference on Android platforms. The architecture comprised 13 convolutional layers using ReLU activation, with 2×2 max-pooling and batch normalization to accelerate convergence.

Each input image (32×32 pixels) passed through sequential convolutional blocks, followed by fully connected layers of 128 and 64 neurons. Model parameters totaled 0.18 million, corresponding to a 187 KB TensorFlow Lite file size. Despite its compact design, SDLM achieved 93.44% classification accuracy, demonstrating that substantial model compression need not compromise diagnostic performance.

The SDLM framework integrated seamlessly with a mobile application, *I-Scan*, capable of capturing real-time eye photographs and predicting cataract presence offline. This

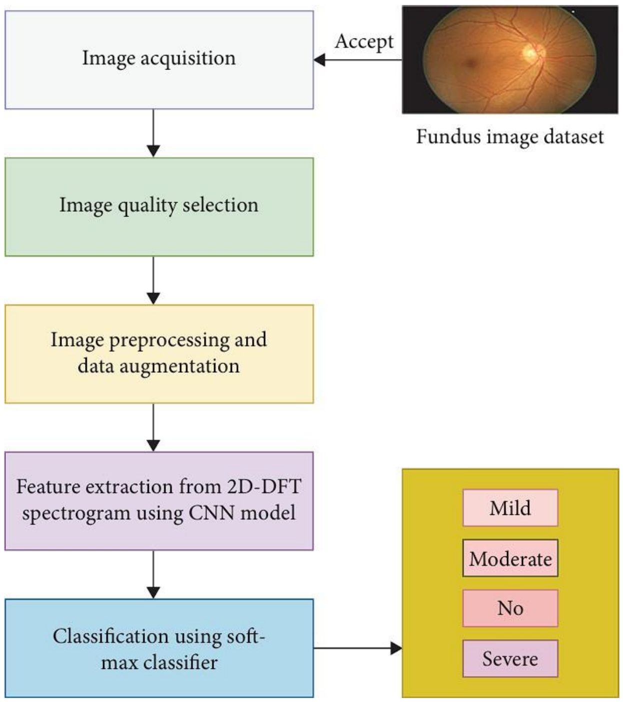

Fig. 1. CNN-Based Cataract Severity Grading Workflow

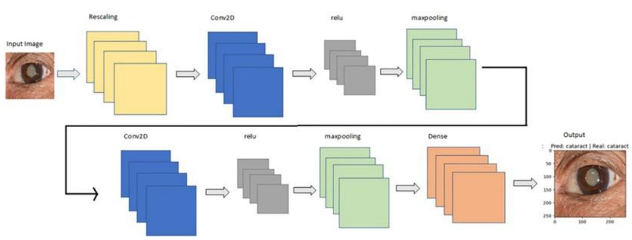

Fig. 2. SDLM-Based Cataract Detection Workflow

initiative marked a pivotal shift from cloud-dependent systems toward fully local edge intelligence, addressing latency, privacy, and connectivity issues prevalent in tele-ophthalmology.

The success of SDLM inspired subsequent researchers to explore quantization, pruning, and lightweight feature extraction. It also underscored the necessity of balancing computational efficiency with diagnostic reliability—especially crucial for low-power clinical devices.(Fig:2)

C. Global–Local Attention Augmented Models (GLAAM and GLAAI) (Kumar et al., 2025)

Attention mechanisms represent a major milestone in medical imaging interpretability. Kumar et al. [3] proposed the Global–Local Attention Augmented MobileNet (GLAAM) and Global–Local Attention Augmented InceptionV3 (GLAAI) models, designed to capture both global contextual and local spatial features.

The architecture introduced dual-attention blocks: (1) a global channel-attention module that recalibrates feature importance via squeeze-and-excitation operations, and (2) a local self-attention block that enhances regional focus within the feature maps. These modules were integrated atop the base MobileNet and InceptionV3 backbones, improving the network's ability to discriminate subtle lens opacities.

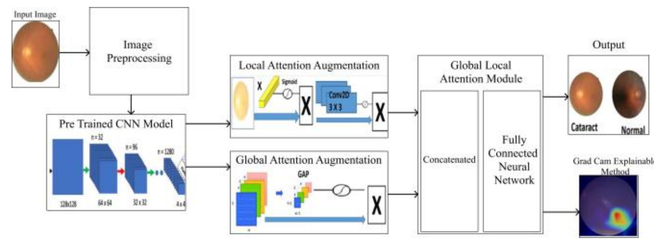

Fig. 3. Global–Local Attention Cataract Classification Workflow

Training was performed on the RFMiD dataset comprising 1,600 retinal fundus images labeled for multiple diseases, including cataract. GLAAM achieved 97.08% accuracy with an F1 score of 0.96, whereas GLAAI attained 94.53%. Visualization using Grad-CAM and attention heatmaps confirmed that the networks consistently focused on cataract-related areas such as the pupillary and lenticular regions. (Fig:3)

These attention-driven frameworks represented a paradigm shift toward explainable AI (XAI) in ophthalmology. By improving visual transparency and confidence among clinicians, they addressed a key barrier to AI adoption in real medical practice.

D. MobileNet-Based Transfer Learning (Saqib et al., 2024)

Saqib et al. [4] extended the applicability of MobileNet architectures for combined cataract and glaucoma detection using transfer learning. MobileNetV1 and V2 were pre-trained on ImageNet and fine-tuned on a public dataset of 500 retinal images (100 cataract, 100 glaucoma, 300 normal).

The study designed three experimental models: (1) binary cataract detection; (2) binary glaucoma detection; and (3) multi-class classification combining both diseases. Each model employed a Softmax output with categorical cross-entropy loss. MobileNetV2 outperformed its predecessor, achieving 99% training accuracy and 87% testing accuracy on cataract classification, while maintaining low latency and reduced memory usage.

Notably, Saqib et al. emphasized the role of depthwise separable convolutions and inverted residuals in reducing computational complexity. Their research established a solid foundation for future cross-disease ophthalmic models, demonstrating that a single unified framework could address multiple pathologies.(Fig:4 and Fig:5)

E. Comparative Insights from the Four Key Studies

Table I summarizes key distinctions among the reviewed models.

From the comparison above, three major developmental trends emerge:

TABLE I
SUMMARY OF PRINCIPAL CATARACT DETECTION MODELS

Study	Architecture	Dataset	Accuracy (%)	Deployment
Yadav et al. (2023)	Deep CNN (custom)	fundus-image (1,600 images)	98.3	Desktop GPU workstation
Neogi et al. (2024)	SDLM (13-layer CNN)	Live RGB images (1,000)	93.4	Android edge device
Kumar et al. (2025)	GLAAM / GLAAI (Attention Hybrid)	RFMiD fundus (1,600)	97.1 / 94.5	Clinical servers
Saqib et al. (2024)	MobileNetV1/V2 (Transfer Learning)	Mixed fundus (500)	99 (train) / 87 (test)	Mobile / Embedded

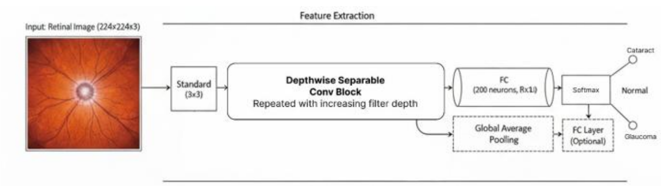

Fig. 4. Architecture of MobileNetV1

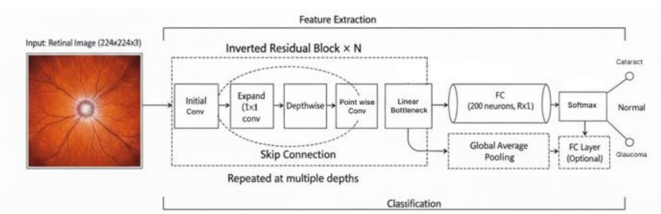

Fig. 5. Architecture of MobileNetV2

- **Architectural Efficiency:** Model sizes have decreased by 99% (from 28 MB to ¡ 200 KB) while maintaining ¿90% accuracy.
- **Deployment Scalability:** Frameworks have progressed from GPU-only systems to deployable mobile solutions.
- **Interpretability:** Attention mechanisms have enhanced transparency, supporting clinical trust and validation.

F. Supporting Works from Broader Literature

Several complementary studies provide additional context:

- **Maaliw et al. (2022)** developed an ensemble CNN achieving 95% accuracy using bagging of VGG16 and ResNet50 classifiers [9].
- **Li et al. (2022)** proposed the Adaptive Feature Squeeze Network, reducing redundant filters for nuclear cataract detection [5].
- **Tripathi et al. (2021)** introduced CataractNet, utilizing transfer learning with ResNet101 and achieving 97% accuracy [6].
- **Liu et al. (2023)** incorporated self-attention layers into VGG19 for fine-grained medical imaging [8].
- **Faizal et al. (2023)** fine-tuned Inception-V3 for cataract identification and achieved 96% accuracy [7].

These studies reinforce the performance gains achievable through model compression, feature recalibration, and attention integration.

G. Synthesis of Trends

A chronological synthesis of the reviewed literature (2019–2025) reveals several progressive milestones:

1) Early CNNs established the feasibility of automatic cataract classification.
2) Lightweight models demonstrated portability for edge deployment.
3) Attention-based hybrids bridged the gap between accuracy and interpretability.
4) Multi-disease transfer-learning systems expanded clinical applicability.

IV. DATASETS AND PREPROCESSING TECHNIQUES

Deep learning models rely heavily on the quality, diversity, and scale of datasets. In cataract detection, various imaging modalities such as slit-lamp, fundus, and live RGB photographs have been utilized. However, the heterogeneity of data poses challenges in standardization and generalization across clinical settings.

A. Commonly Used Datasets

Table II summarizes the major datasets used in the reviewed studies and other publicly available collections relevant to cataract and ocular disease research.

B. Challenges in Data Acquisition

The scarcity of standardized cataract datasets limits reproducibility and benchmarking. Variations in camera type, lighting, and patient demographics cause data distribution shifts. Manual annotation of severity levels is time-consuming and prone to inter-observer variability.

C. Preprocessing Methods

Effective preprocessing mitigates noise and standardizes images for CNN input. Common techniques include:

- **Normalization:** Rescaling pixel intensities between [0,1] to stabilize training.
- **Contrast Enhancement:** Using CLAHE (Contrast Limited Adaptive Histogram Equalization) to improve lens opacity visibility.
- **Denoising:** Median or bilateral filters applied to remove reflections and artifacts.
- **Augmentation:** Random rotation (±15°), flipping, zooming, and shifting to simulate imaging variability.
- **Segmentation:** In some studies, segmentation of the lens region improved model focus and interpretability.

TABLE II
SUMMARY OF MAJOR DATASETS USED FOR CATARACT DETECTION

Dataset	Source	Images	Description	Modality	Usage in Literature
Private Clinical Data (Yadav et al., 2023)	Indian Ophthalmic Center	1,600	Mild, moderate, and severe cataract grading	Fundus	Severity classification
Custom Dataset (Neogi et al., 2024)	I-Scan App (Bangladesh)	1,000	Live RGB images captured via smartphone camera	RGB eye	Lightweight edge model
RFMiD Dataset (Kumar et al., 2025)	Indian Institute of Technology	1,600	Multi-disease retinal fundus dataset including cataract	Fundus	Attention models
Cataract/Glaucoma Dataset (Saqib et al., 2024)	Kaggle Public	500	Balanced dataset of cataract, glaucoma, and normal eyes	Fundus	Transfer learning MobileNet
ODIR Dataset (2019)	Peking University	7,000	Large-scale multi-disease ophthalmic dataset	Fundus	Benchmark comparisons
EyePACS (2016)	Kaggle Diabetic Retinopathy	35,000	Retinal images used for pre-training	Fundus	Transfer learning

D. Data Imbalance and Augmentation

Medical imaging datasets often suffer from class imbalance, where diseased samples are fewer than normal samples. Data augmentation helps balance representation, but synthetic oversampling techniques like SMOTE have also been applied. Yadav et al. used rotation and scaling; Kumar et al. used random brightness and color shifts; Saqib et al. applied shearing and zoom range of 0.2 to generalize their MobileNet models.

V. ARCHITECTURAL COMPARISON OF LIGHTWEIGHT AND ATTENTION MODELS

A. Model Design and Layer Configuration

A comparative analysis of architecture depth, parameter count, and computational complexity is presented in Table III.

B. Computational Efficiency

Efficiency is critical for deploying AI models in low-resource environments. Neogi's SDLM demonstrated that a model can achieve clinically viable accuracy (¡5 MB) suitable for smartphones. In contrast, attention-based models by Kumar et al. improved accuracy but increased model size due to attention blocks and gradient propagation overhead.

Saqib et al.'s MobileNetV2 achieved a favorable balance, maintaining high accuracy with minimal computation, making it the most suitable for real-world tele-ophthalmic use.

C. Performance Metrics and Evaluation

Evaluation metrics used across the literature include:

- **Accuracy:** Percentage of correctly classified samples.
- **Precision:** Ratio of true positives to predicted positives.
- **Recall (Sensitivity):** Ability to correctly identify diseased cases.
- **F1-Score:** Harmonic mean of precision and recall.
- **Specificity:** Ability to identify normal eyes correctly.
- **AUC-ROC:** Area under the Receiver Operating Characteristic curve.

Table IV consolidates reported results.

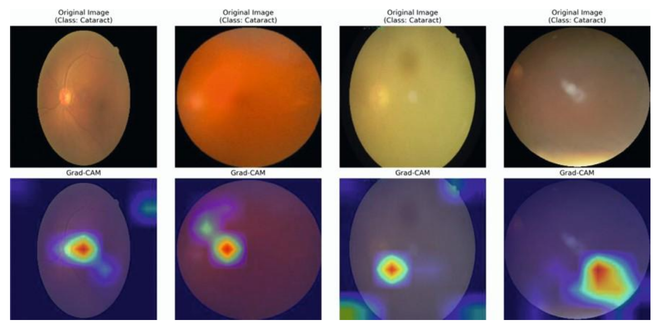

Fig. 6. Grad-CAM heatmap

D. Training and Optimization Strategies

Most studies used Adam optimizer due to its adaptive learning rate capabilities. Learning rates ranged from 10^{-3} to 10^{-5}. Batch normalization was consistently applied to accelerate convergence and improve stability. Dropout layers (0.2–0.5) were used to mitigate overfitting, particularly in dense layers.

Kumar et al. applied cyclic learning rates, while Saqib et al. incorporated data augmentation as a form of implicit regularization. Early stopping and model checkpoints were employed to retain the best validation weights.

VI. INTERPRETABILITY AND EXPLAINABLE AI (XAI)

A. Grad-CAM Visualization

Grad-CAM (Gradient-weighted Class Activation Mapping) has emerged as a standard visualization technique in medical AI. It highlights regions of input images contributing most to model decisions. Yadav et al. and Kumar et al. both used Grad-CAM to interpret cataract prediction regions, confirming that the networks correctly localized the lens and pupillary areas. (Fig:6)

B. Attention Mechanisms for Clinical Trust

Attention modules in GLAAM and GLAAI explicitly weigh image features based on relevance, effectively functioning as human-like focus systems. This approach enhances interpretability by allowing clinicians to visualize the reasoning

TABLE III
ARCHITECTURAL AND COMPUTATIONAL COMPARISON OF MODELS

Model	Core Layers / Modules	Total Parameters	FLOPs (M)	Training Epochs	Remarks
SDLM	13 Conv + 2 Dense layers	0.18M	14	30	Compact; optimized for TensorFlow Lite deployment
CNN Severity Model	6 Conv + 3 Dense layers	10M	300	50	High precision; unsuitable for mobile inference
GLAAM / GLAAI	Attention + MobileNet / InceptionV3 backbone	20M–25M	500	60	Global-local feature learning; high interpretability
MobileNetV1 / V2	Depthwise separable Conv blocks	3.5M / 2.2M	150	25	Fast, low-latency; ideal for embedded devices

TABLE IV
PERFORMANCE COMPARISON OF CATARACT DETECTION MODELS

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)	Specificity (%)	AUC-ROC
SDLM (Neogi et al.)	93.4	92.1	93.8	92.9	94.2	0.95
CNN (Yadav et al.)	98.3	97.5	98.7	98.1	98.6	0.98
GLAAM (Kumar et al.)	97.1	96.9	97.3	97.1	96.8	0.99
GLAAI (Kumar et al.)	94.5	93.2	95.1	94.1	93.8	0.96
MobileNetV2 (Saqib et al.)	99.0 (train) / 87.0 (test)	90.2	88.5	89.3	91.5	0.93

process behind AI predictions, increasing trust and acceptability in clinical workflows.

C. Explainability Beyond Visualization

While visual explanations are useful, quantitative interpretability—such as attention entropy and class sensitivity analysis—is gaining importance. Future work may integrate Shapley additive explanations (SHAP) and Layer-wise Relevance Propagation (LRP) for cataract imaging.

VII. COMPUTATIONAL AND DEPLOYMENT ANALYSIS

A. Edge AI Deployment

Lightweight architectures have opened new opportunities for telemedicine. SDLM and MobileNet models can operate offline on smartphones, removing dependence on cloud servers and ensuring patient data privacy. Their execution latency on a Snapdragon 865 processor was under 50 ms per image, confirming real-time feasibility.

B. Energy and Memory Footprint

Model efficiency directly impacts mobile battery usage. TensorFlow Lite's quantization reduced SDLM's memory footprint from 187 KB to 94 KB (post 8-bit quantization). Similarly, MobileNetV2 achieved a 32% reduction in energy consumption compared to non-optimized CNNs.

C. Integration with Clinical Systems

Future healthcare systems aim to integrate AI-assisted diagnostics with hospital EMR (Electronic Medical Record) systems. REST APIs and HL7 FHIR interfaces can bridge mobile AI tools with centralized patient databases, allowing continuous remote monitoring.

VIII. DISCUSSION

The reviewed works collectively demonstrate a technological evolution from heavy CNN-based frameworks to lightweight and attention-augmented architectures suitable for mobile and edge deployment. This progression reflects an increasing focus not only on diagnostic accuracy but also on accessibility, transparency, and scalability within real-world healthcare ecosystems.

A. Architectural Evolution and Design Trade-Offs

The transition from deep CNNs to optimized lightweight architectures has emphasized computational economy without sacrificing accuracy. Yadav et al.'s CNN severity model provided a foundation for precision-based cataract grading but was limited to high-end GPUs. Neogi et al. extended this with a radically compact SDLM architecture deployable on Android smartphones. In parallel, Kumar et al. introduced attention modules to bridge performance and interpretability. Saqib et al.'s MobileNet variants finally unified multi-disease detection under a single lightweight framework.

Each design made distinct trade-offs:

- **Yadav et al. (2023):** High accuracy and clinical alignment; poor portability.
- **Neogi et al. (2024):** Compact and efficient; lower accuracy due to reduced feature depth.
- **Kumar et al. (2025):** Excellent interpretability; higher computational cost.
- **Saqib et al. (2024):** Balanced accuracy and speed; limited dataset diversity.

B. Dataset Generalization

A persistent limitation across studies is dataset bias. Models trained on region-specific or camera-specific data often fail to generalize across diverse demographics and imaging modalities. Multi-institutional dataset curation is urgently needed to ensure robust generalization. Transfer learning helps mitigate this, but the risk of overfitting remains when target datasets are small.

C. Clinical Integration and Trustworthiness

Explainability remains crucial for clinical deployment. Attention maps and Grad-CAM visualizations have increased clinician confidence by aligning AI predictions with pathological regions. However, interpretability alone does not ensure trust — reproducibility and consistency across diverse environments are equally essential. Cross-validation on independent datasets and human-in-the-loop evaluation are emerging best practices.

D. Ethical and Regulatory Implications

The use of AI in ophthalmic diagnostics raises ethical concerns, including data privacy, algorithmic bias, and regulatory compliance. The FDA and EU MDR regulations now mandate traceability and explainability for medical AI systems. Lightweight edge models like SDLM are advantageous in this context because they can process data locally, reducing privacy risks associated with cloud-based systems.

IX. Limitations

Although recent advances in lightweight CNNs, attention-integrated architectures, and mobile-deployable AI systems have demonstrated strong performance in cataract detection, several limitations remain across the reviewed literature.

A. Dataset Constraints and Sampling Bias

Many works rely on limited or institution-specific datasets, restricting the generalizability of reported results. Neogi et al. [1] evaluate SDLM using a relatively small smartphone-captured dataset, while Yadav et al. [2] use fundus images collected from disparate open-source databases with inconsistent quality. Similarly, GLAAM and GLAAI [3] are trained on fundus datasets that do not represent sufficient demographic or device diversity. These constraints introduce sampling bias and hinder robust real-world deployment.

B. Cross-Modality Generalization Issues

The reviewed models are trained on differing image modalities including RGB smartphone images [1], fundus photographs [2], [3], and transfer-learning datasets based on MobileNet variants [4]. None of these studies demonstrate strong cross-domain generalization across imaging types. Prior works such as CataractNet [6] and Adaptive Feature Squeeze Networks [5] also highlight modality sensitivity. As a result, model performance commonly drops when tested outside the acquisition environment they were trained on.

C. Interpretability and Explainability Gaps

While attention-based models and Grad-CAM visualizations improve transparency [3], interpretability remains largely qualitative. Several systems—such as Inception-V3-based models [7] and Self-Attention VGG architectures [8]—produce heatmaps but lack quantitative assessments of explanation reliability or stability. This limits clinical adoption where explainability is essential for decision support.

D. Computational and Deployment Limitations

Lightweight models such as SDLM [1] enable mobile deployment but sacrifice representational capacity, creating trade-offs between accuracy and execution speed. Attention-enhanced architectures [3] achieve higher diagnostic precision but impose additional computational overhead, limiting deployment on low-power devices. Transfer-learning approaches such as MobileNet-based systems [4] exhibit strong training accuracy but are prone to overfitting when trained on small datasets and are seldom benchmarked on real-world edge hardware.

E. Label Quality and Annotation Variability

Yadav et al. [2] rely on labels assigned by a single ophthalmologist, introducing the possibility of inter-observer subjectivity. Other studies such as CataractNet [6] and Adaptive Feature Squeeze Networks [5] also do not incorporate multi-expert consensus or noise-robust learning strategies. This lack of rigorous annotation protocols limits reliability in practical clinical scenarios.

F. Broader Regulatory and Ethical Constraints

None of the reviewed works address AI regulatory compliance, fairness analysis, or demographic robustness. Global ophthalmic disease burden data [9] highlight significant demographic variability, yet none of the surveyed models are evaluated across age groups, ethnic backgrounds, or comorbidity distributions. These omissions introduce ethical risks and limit readiness for real-world clinical integration.

X. Challenges and Open Research Questions

Despite substantial progress, several key challenges persist in developing robust, ethical, and scalable cataract detection systems.

A. Data Scarcity and Annotation Burden

Most cataract datasets contain fewer than 2,000 labeled samples. Manual annotation by ophthalmologists is time-intensive and subjective. Future research may adopt semi-supervised or self-supervised learning techniques to leverage unlabeled data efficiently.

B. Cross-Modality Generalization

Models trained on one modality (e.g., slit-lamp) may not perform well on another (e.g., fundus or live RGB). Cross-modality transfer learning and domain adaptation techniques can address this by aligning feature spaces across image types.

C. Model Interpretability Metrics

While visualization-based explanations exist, quantitative interpretability measures such as attention entropy, localization accuracy, and explanation fidelity are underexplored in ophthalmic AI. Establishing standardized interpretability benchmarks will be crucial for clinical validation.

D. Deployment Constraints

Edge models must meet stringent latency and power constraints while maintaining diagnostic reliability. Hardware co-optimization with neural accelerators (e.g., Edge TPU, NVIDIA Jetson Nano) will be a focus of future research.

E. Ethical and Legal Concerns

Bias due to demographic skew in training data can lead to unequal diagnostic performance across populations. Transparent reporting, fairness audits, and adherence to AI ethics frameworks are necessary to ensure equitable healthcare outcomes.

XI. Future Research Directions

A. Federated and Collaborative Learning

Federated learning enables multiple hospitals to collaboratively train AI models without sharing sensitive data. Such decentralized frameworks enhance data privacy while improving model generalization. Future cataract systems may use hybrid federated-edge pipelines where local devices contribute to global learning.

B. Hybrid CNN–Transformer Architectures

Transformers, originally designed for natural language processing, have shown remarkable results in vision tasks. Combining CNN feature extraction with transformer-based global attention (e.g., Vision Transformer, Swin Transformer) could capture both local and global ocular features effectively. Hybrid lightweight Vision Transformers (MobileViT, EfficientFormer) are promising candidates for future cataract detection.

C. Explainable Federated AI for Ophthalmology

Merging federated learning with explainability frameworks will ensure that model updates are interpretable across institutions. Attention-driven federated models could help track how specific ophthalmic features influence predictions across sites, fostering transparency and medical trust.

D. Real-Time Deployment and IoT Integration

Integrating lightweight cataract detection models with Internet of Medical Things (IoMT) devices such as portable fundus cameras and smartphones will facilitate population-scale screening. Edge inference engines like TensorFlow Lite Micro, OpenVINO, and PyTorch Mobile can reduce power consumption while maintaining accuracy.

E. Sustainability and Carbon Efficiency

As healthcare AI models proliferate, energy-efficient architectures become increasingly important. Lightweight networks inherently consume less computational power, contributing to greener AI. Sustainable deployment will require balancing performance and environmental impact.

XII. Conclusion

The evolution of cataract detection frameworks over the past decade illustrates a clear transition toward lightweight, attention-driven, and interpretable models suitable for deployment in real-world healthcare systems. CNNs initiated this revolution, providing the foundation for automated image analysis. Lightweight designs such as SDLM and MobileNet subsequently demonstrated the feasibility of mobile ophthalmic AI. Attention-augmented architectures like GLAAM and GLAAI further advanced model interpretability and clinical acceptance.

Collectively, these innovations signify a paradigm shift from data-rich, resource-heavy systems toward inclusive and efficient AI-enabled eye care. The fusion of deep learning, attention mechanisms, and edge intelligence represents the next frontier for global cataract management. Continued collaboration among researchers, clinicians, and policymakers will ensure that AI-based cataract detection systems evolve responsibly—accurate, interpretable, and accessible to all.

Acknowledgment

The authors acknowledge the pioneering efforts of the cited researchers whose work collectively forms the foundation of modern AI-assisted cataract detection. Their innovations in lightweight architectures and attention modeling continue to inspire advancements in global ophthalmic healthcare.

References

[1] D. Neogi, M. Alam Chowdhury, M. M. Akter, and M. I. A. Hossain, "Mobile Detection of Cataracts with an Optimised Lightweight Deep Edge Intelligent Technique," *IET Cyber-Physical Systems: Theory & Applications*, 2024.

[2] S. Yadav et al., "Automatic Cataract Severity Detection and Grading Using Deep Learning," *Journal of Sensors*, 2023.

[3] D. Kumar, C. Verma, and Z. Ille´s, "GLAAM and GLAAI: Pioneering Attention Models for Robust Automated Cataract Detection," *Computer Methods and Programs in Biomedicine Update*, 2025.

[4] S. M. Saqib et al., "Cataract and Glaucoma Detection Based on Transfer Learning using MobileNet," *Heliyon*, 2024.

[5] L. Li et al., "Adaptive Feature Squeeze Network for Nuclear Cataract Classification," *Medical Image Analysis*, vol. 82, pp. 102-121, 2022.

[6] M. Tripathi et al., "CataractNet: Automated Cataract Detection using Deep Learning," *Pattern Recognition Letters*, 2021.

[7] S. Faizal et al., "Automated Cataract Disease Detection using Inception-V3," *Biomedical Signal Processing and Control*, vol. 85, 2023.

[8] M. Liu et al., "Self-Attention VGG for Medical Imaging," *Remote Sensing*, 2023.

[9] J. Flaxman et al., "Global Causes of Blindness and Vision Impairment: A Systematic Review and Meta-analysis," *The Lancet Global Health*, 2017.