# Exploring the role of Explainable AI in Cataract Detection: A Comprehensive Review of Existing Techniques and Future Possibilities

Harsh Shah
*Department of Information Technology*
*Dwarkadas J. Sanghvi College of Engineering*
Mumbai, India
harshshah7074@gmail.com

Rutwik Patel
*Department of Information Technology*
*Dwarkadas J. Sanghvi College of Engineering*
Mumbai, India
rutwikpatel1313@gmail.com

Sanchit Hegde
*Department of Information Technology*
*Dwarkadas J. Sanghvi College of Engineering*
Mumbai, India
sanchit.hegde22@gmail.com

Harshal Dalvi
*Department of Information Technology*
*Dwarkadas J. Sanghvi College of Engineering*
Mumbai, India
harshal.dalvi@djsce.ac.in

***Abstract*—Cataract is a prevalent eye disease that affects millions of individuals throughout the world. Early detection and treatment of cataracts are crucial for preventing vision loss and enhancing quality of life. By boosting diagnosis accuracy and efficiency, artificial intelligence (AI) has the potential to transform cataract detection. Yet, the opacity and complexity of AI models make them difficult to implement in clinical practice. Explainable AI (XAI) approaches address this issue by allowing healthcare practitioners to comprehend the decision-making process of AI models. In this paper, we present a literature review on cataract detection systems and XAI techniques in healthcare. We address the benefits and drawbacks of various cataract detection systems and XAI technologies utilized in the healthcare area, as well as their potential impact on clinical operations.**

***Index Terms*—Cataract detection, Explainable AI, Deep Learning**

## I. INTRODUCTION

Cataract is one of the leading causes of visual impairment and blindness worldwide. Cataract is the most common cause of blindness in India, accounting for approximately 62% of all occurrences [1]. 90% of those who are partially or fully blind live in developing countries [2]. Nonetheless, over 75% of vision loss is curable, suggesting that roughly four out of every five cases are reversible [2]. Late-stage eye disorders always cause significant visual acuity impairment, which might be permanent [2]. As a result, early detection and prevention of cataracts can aid in the avoidance of vision impairment and blindness.

In recent years, artificial intelligence (AI) and machine learning (ML) have emerged as promising technologies to improve the accuracy and efficiency of cataract detection. Deep convolutional neural networks have produced excellent results in CACD (Computer-aided Cataract Diagnosis) techniques. However, deep convolutional neural networks require large datasets to be accurate. Gathering large datasets for real-time training is time-consuming and sometimes impossible. As a result, rather than training a CNN from scratch, transfer learning approaches based on pre-trained CNNs are typically preferred. The pre-trained CNN-based CACD algorithms are trained on millions of natural images before being transferred, allowing for fine-tuning to enable cataract detection with small datasets. In fact, fine-tuning a pre-trained CNN is easier and faster than training a deep CNN from scratch with randomly initialized weights.

However, the lack of interpretability and transparency of AI models is a major barrier to their adoption in healthcare. Explainable AI (XAI) approaches aim to overcome this issue by providing insights into AI models' decision-making processes. XAI approaches are intended to make AI more transparent and interpretable, allowing healthcare practitioners to comprehend the reasons behind AI decisions and build trust in these systems.

By offering explanations for the model's predictions, XAI approaches can aid in the conversion of black-box models to clear-box models. A black-box model is an AI model whose internal workings are not visible to the user or creator. The model makes predictions or decisions, but it is unclear how those predictions are made. A clear-box model, on the other hand, allows the user to comprehend how the model makes its predictions. This is significant because it helps users have confidence in the model's output and a deeper understanding of the elements that drive its judgments. By applying XAI approaches, we can acquire insights into how black-box models make decisions and increase our understanding of the factors that drive those decisions. This can help to increase the transparency and accountability of AI models and to build trust in their output.

In this paper, we present a comprehensive literature survey of cataract detection systems and XAI techniques in healthcare. We discuss the benefits and drawbacks of several cataract detection systems as well as various AI models and XAI methods used in the healthcare domain. We also look at the challenges and potential of integrating these technologies into clinical processes, such as the requirement for user-friendly and interpretable XAI techniques that healthcare practitioners with varied degrees of technical experience can understand. Our review also emphasizes the significance of ethical considerations in the development and implementation of AI-powered cataract detection systems.

Additionally, we look into the possible impact of XAI techniques on AI system adoption and acceptance in healthcare, as well as their ability to improve patient outcomes and quality of life. The findings from this research can help guide the development of more effective and clinically relevant cataract detection systems, as well as the incorporation of AI and XAI approaches into clinical practice. Overall, this paper contributes to the growing body of literature on AI and XAI in healthcare and provides insights into the challenges and opportunities of implementing these technologies in the context of cataract detection.

## II. Related Work

### A. Literature Review of Cataract Detection Techniques

The authors of paper [2] proposed a cataract detection system using a convolutional neural network with a pre-trained VGG-19 model on a dataset that consisted of fundus images of 800 patients. The dataset consisted of images of varied sizes, so OpenCV was used to resize the image to 224*224 pixels. A VGG-19 pre-trained CNN model was applied to preprocessed images to classify cataract and normal eye images. The Adam optimization algorithm was used to reduce the cost function and improve the performance of the model.

The authors of paper [3] proposed to develop an Android application for cataract detection using the Cascade Classifier library. Firstly, images of the eye are captured by placing smartphones in front of the eye at a distance of not more than 30cm. Then, the cascade classifier was used to detect and crop the pupil out of an image. The system used this RGB value of the pupil to compare with the database value of the cataract's RGB value to detect the cataract.

The authors of paper [4] experimented with cataract detection using an eye image analysis based on luminance. The luminance-based technique extracts the luminance value of an image by using pixel brightness transformation. In the experiment, 100 eye images were taken: 50 from healthy eyes and 50 from diseased eyes. To remove noise from the images, the median filter and watershed algorithms were used. The SVM classifier was used to differentiate between healthy and diseased eyes.

The authors of paper [5] proposed a hybrid convolutional and recurrent neural network (CRNN) for the detection and classification of cataracts according to their severity. A dataset of 8030 high-quality fundus images was captured without flash and with auto white balance and classified into four categories: normal (no cataract), mild, moderate, and severe. The proposed CRNN divided the dataset into a number of subsets, and each subset was fed to pre-trained CNN models. The extracted features were then combined using global average pooling and finally fed into LSTM for classification into four classes. The authors obtained 97% accuracy.

The authors of paper [6] put forward a New Angular Binary Pattern (NABP) for the extraction of texture features in their paper. Following the extraction of features, the author proposed a kernel-based CNN. The author discovered some flaws in convolutional layers and attempted to solve them. For example, convolution layers reduce the number of weights, which reduces memory usage, resulting in faster computation and less overfitting of the dataset. After comparing proposed and existing systems, 97.3% accuracy is achieved.

The authors of paper [7] proposed a method for automatically grading cataracts based on a patient's eye video. The dataset consists of a cataract video with a length less than 10 seconds of 76 eyes, i.e., 38 people collected using the slit lamp method, and 1520 images were extracted from this video. iSpector Mini is video collection equipment used to collect the dataset. The proposed method uses the YOLOv3 algorithm to classify the cataract by automatically identifying the position of the lens. Accuracy of 94% and an F1 score of 0.9388 were achieved using the proposed method.

The authors of paper [8] proposed a classification technique for ocular diseases using a dense correlation network, DCNet. This classification task is based on paired color fundus photographs (CFPs). DCNet is made up of a backbone convolutional neural network that extracts feature representations from the paired CFPs, a spatial correlation module that captures dense correlation between features of the paired CFPs and fuses relevant feature representations, and a classifier to produce a disease score. The dataset consists of 3500 eye sample images. Each of the 3500 patients has been classified into eight categories. The images are first resized to 512x512, and then 448x448 images are cropped randomly. The final score and F1 score are 0.827 and 0.913, respectively, when ResNet—101 is used as the backbone CNN.

The authors of paper [9] proposed a cataract classifier where Decision Tree and Bayesian Network are used. Both of these algorithms are supervised learning algorithms, and tri-learning is employed to discover a sound hypothesis. The wavelet and texture extracted from each fundus image are used to calculate the accuracy of the classifier. It was observed that the wave feature performed better than the texture feature. A dataset of 5378 fundus images was used to classify cataracts into four grades according to their severity: non-cataract, mild, moderate, and severe. Total accuracy of 88% and 70% were achieved when Bayesian networks and J48 were used.

The authors of paper [10] proposed two algorithms, one for classifying the eye into three categories—healthy eyes, mild cataracts, and severe cataracts. And a second algorithm for determining the degree of cataract present in the affected eyes. The author classified an eye as healthy if the mean intensity of the histogram for that eye is below 50 and as having cataracts if its mean intensity is above 100. The second method involves calculating the pupil and cataract areas in an unhealthy eye. The formula used for calculating the percentage of cataract is $Degree = (cataractarea/(pupilarea + cataractarea)) * 100$.

The authors of paper [11] proposed an effective network selection method for computer-aided cataract detection in a noisy environment. The suggested method is divided into two parts: first, the input images are pre-processed to reduce noise, and then multiple deep neural networks are trained on the pre-processed images. After that, the trained networks are evaluated using a performance metric, and the best-performing network is chosen for the final diagnosis. The proposed method was tested on a dataset of cataract images with varied amounts of noise, and the findings revealed that the chosen network obtained very high accuracy in noisy environments as compared to other methods. Overall, the proposed method can provide an accurate and efficient solution for cataract diagnosis in noisy situations.

The authors of paper [12] proposed employing ensemble neural networks and transfer learning to detect and grade cataracts. The proposed method is divided into two stages: first, a pre-trained convolutional neural network (CNN) is utilized to extract features, and then an ensemble of several CNNs is trained to categorize the recovered features into various cataract grades. Using a large dataset of cataract images, the ensemble model is trained using a combination of transfer learning and fine-tuning. The proposed method was tested on two publicly available datasets, and the findings showed that it performed effectively.

The authors of paper [13] proposed a method for detecting cataract disease using deep convolutional neural networks (CNNs). The suggested method entails preprocessing the input images to improve contrast and remove artifacts, then training a CNN on the preprocessed images to categorize them as cataract or non-cataract. The suggested method employs a CNN architecture that includes several convolutional and pooling layers, followed by fully connected layers for classification. The proposed approach was tested on a dataset of cataract images, and the results showed that it achieved high accuracy.

The authors of paper [14] developed an optimal hybrid approach for cataract detection that includes image processing techniques and machine learning algorithms. The proposed method involves preprocessing the input images with contrast enhancement and image normalization techniques before extracting features with the grey level co-occurrence matrix (GLCM) and discrete wavelet transform (DWT). The features gathered are then used to train a cataract detection support vector machine (SVM) classifier. A grid search algorithm is used to identify the ideal combination of SVM hyperparameters to optimize the performance of the SVM classifier. The proposed approach was tested on a dataset of cataract images, and the results showed that it performed well.

### B. Literature Review of XAI Techniques

The authors of paper [15] review the current state of explainable artificial intelligence (XAI) in healthcare. The authors discuss various XAI techniques, such as rule-based methods like decision trees and decision sets, model-agnostic methods like LIME and SHAP, and posthoc methods like Grad-CAM and Activation Maximization. The authors evaluate the accuracy and interpretability of these methods using publicly available medical datasets, such as the Electronic Health Record (EHR) dataset and the Chest X-Ray dataset. The authors also discuss various XAI approaches, including model-based methods, such as explainable neural networks; decision-based methods, such as counterfactual reasoning; and instance-based methods, such as prototype-based explanation. The authors identify the challenges and opportunities for XAI in the medical domain, such as handling high-dimensional and complex medical data, providing clear and actionable explanations, and integrating ethical and legal considerations into XAI design. The authors conclude that XAI has great potential for improving the accuracy, reliability, and fairness of medical decision-making but also highlight the need for further research and development to fully realize this potential.

The authors of paper [16] conducted a review of 223 studies on the application of deep learning-based explainable artificial intelligence (XAI) in healthcare, focusing on anterior and medical imaging modes. The reviewed papers were categorized by the authors, who observed some notable trends. Most studies used post-hoc explanations instead of model-based explanations and employed both model-specific and model-agnostic explanation methods. Additionally, the majority of the papers provided more local explanations and fewer global explanations. It is not surprising to see these developments given the authors' emphasis on deep learning for analyzing medical images. Among the available methods for explaining the predictions made by convolutional neural networks (CNNs), saliency mapping techniques are the most commonly used. These methods offer localized and post-hoc explanations that are specific to the model being used. In addition, unlike model-based XAI techniques, post-hoc methods can be applied after the neural network has been trained, making them easier to use.

The authors of the paper [17] present a study that evaluates the performance of different explainable artificial intelligence (XAI) methods in the medical domain. The authors apply a set of XAI methods, including rule-based methods, model-agnostic methods, and post-hoc methods, to a medical dataset, such as the Retinal Fundus Image Quality Assessment (RFIQA) dataset and the EyePacs dataset. The authors use a panel of medical experts to evaluate the XAI techniques like Grad-CAM and SIDU based on their transparency, interpretability, and accountability. The results of the study show that the XAI methods have different levels of accuracy, ranging from 75% to 85%. The study also reveals that the XAI methods vary in terms of their transparency, interpretability, and accountability, with some methods providing clear and actionable explanations while others are less interpretable.

Table I
LITERATURE SURVEY OF PREVIOUS PAPERS FOR CATARACT DETECTION

| Paper Title | Pre-processing Techniques | Classification Techniques | Shortcomings | Accuracy / F1 Score |
|---|---|---|---|---|
| Cataract Detection using Convolutional Neural Network with VGG-19 Model [2] | OpenCV libraries to resize image to equal size | VGG-19 CNN model | Due to the limited images, they were unable to develop severity grading and identify exact location of cataract. | 97% |
| Cataract detection using smartphone [3] | Cascade Classifier | Matches the color of the pupil with the color of the pupils with cataracts in the dataset | Implemented on a very small dataset of 50 people of which 20 had cataract and 30 were normal. | 90% |
| Detecting Cataract Using Smartphones [4] | Median Filter, Watershed Algorithm | SVM | Only classifies as healthy and diseased eyes. It does not detect the cataract in an eye. Dataset was small and only contained 100 images. | 96% |
| Fundus image-based cataract classification using a hybrid convolutional and recurrent neural network [5] | Prepocessed images were used | hybrid convolutional and recurrent neural network | Since multiple CNN models are used in the proposed system, the time complexity will be greater. | 97% |
| Novel angular binary pattern (NABP) and kernel based convolutional neural networks classifier for cataract detection [6] | Novel Angular Binary Pattern | Kernel based CNN | Implemented on a very small dataset of 100 cataract images. | 97.3% |
| Automatic classification algorithm of cataract video based on deep learning [7] | Densenet model | YOLOv3 | Dataset was created using only 38 people which brings the problem of overfitting. | 94% |
| Dense Correlation Network for Automated Multi-Label Ocular Disease Detection with Paired Color Fundus Photographs [8] | OpenCV libraries to resize images to 448 x 448 | DCNet – backbone CNN, the SCM and the classifier | Limited interpretation of model, lack of external validation and limited dataset. | F1 Score: 0.827 |
| Semi-Supervised Learning Based on Cataract Classification and Grading [9] | Image normalisation, Contrast enhancement, Image denoising, Image segmentation and Feature Extraction | Bayesian Network and Decision Tree | Limited pre-processing techniques and limited interpretation. | 88% |
| Cataract Detection using Digital Image Processing [10] | Image acquisition, Image resizing, Gray scaling of Image, Feature extraction | Histogram is prepared for each image. Images having mean intensity below 50 are healthy and above 100 are affected by cataract | Dataset used is small, manual pre-processing is required. | - |
| Efficient network selection for computer-aided cataract diagnosis under noisy environment [11] | Feature extraction using pre-trained Alex-net model | Locally and globally trained Support Vector Networks (SVN) | Impact of other image distortions such as blur, contrast, etc. is not considered. | 92.97% |
| Cataract Detection and Grading Using Ensemble Neural Networks and Transfer Learning [12] | Downsized the images into equal sizes with 2048 x 2048 pixels, RGB images normalized between 0 and 1, and used resizing, rotation, shifting, and flipping to obtain additional images. | Transfer learning, ensembles of pre-trained CNN's and stacked LSTM's | Proposed system is not suitable for noisy and low-quality fundus images. | 99.20% (normal vs. cataract) and 97.76% (normal to severe) |
| Cataract Disease Detection Used Deep Convolution Neural Network [13] | Image normalization, augmentation and enhancement and feature extraction | pre-trained VGG16 | The size of the dataset is very small. | 97.1% |

Table II
LITERATURE SURVEY OF PREVIOUS PAPERS FOR XAI TECHNIQUES IN MEDICAL DOMAIN

| **Paper Title** | **XAI Techniques** | **Output** | **Shortcomings** |
|---|---|---|---|
| A Survey on Explainable Artificial Intelligence (XAI): Toward Medical XAI [15] | LIME, SHAP and Grad-CAM | - | Cannot handle high-dimensional and complex medical data. |
| Explainable artificial intelligence (XAI) in deep learning-based medical image analysis [16] | CAM and Grad-CAM | - | Didn't cover techniques like LIME and SHAP. |
| Expert Level Evaluations for Explainable AI (XAI) Methods in the Medical Domain [17] | Grad-CAM and SIDU | 75 % – 80 % | Lesser accuracy, reliability, and fairness of medical decision-making. |
| LISA : Enhance the explainability of medical images unifying current XAI techniques [18] | LIME, Integrated gradients, Anchors and SHAP | Using CNN and transfer learning, testing accuracy for detecting chest x-rays was around 90%. | False-positive and false-negative results. |
| RetainVis: Visual Analytics with Interpretable and Interactive Recurrent Neural Networks on Electronic Medical Records [19] | Attention mechanism | For heart failure and cataract, the AUC (Area Under the ROC Curve) is around 95% and 97%, respectively. | Due to its computationally expensive deep learning-based model and restricted screen real estate, it has scalability concerns. |
| Explainable framework for Glaucoma diagnosis by image processing and convolutional neural network synergy: Analysis with doctor evaluation [20] | CAM is used for heat map based explanation for image analysis done by CNN. | On ORIGA-Light retinal image dataset, the accuracy is 93.5%, precision is 93.8% and F1 score is 95.7%. | Other XAI methods can be integrated with similar CNN models, they might have better results than the current findings. Detailed evaluation with humans is still required. |
| Explanatory classification of CXR images into COVID-19, Pneumonia and Tuberculosis using deep learning and XAI [21] | LIME, SHAP and Grad-CAM | Accuracy rates of 94.31% in testing and 94.54% in validation. | Trained on small dataset and performance is not tested on large dataset. |
| An Explainable AI driven Decision Support System for COVID-19 Diagnosis using Fused Classification and Segmentation [22] | Grad-CAM and Guided Grad-CAM | In GGECS, the proposed classification model has an overall accuracy of 98.51%, while the segmentation model achieves an IoU score of 0.595. | Doctor validation is still required because incorrect output in the case of COVID-19 could result in dangerous circumstances. |

The authors of the paper [18] present a system called LISA (Local Interpretable and Simulatable Network-Based Analysis), which attempts to improve the explainability of medical images by combining many current explainable artificial intelligence (XAI) techniques. LISA is intended to enable global and local interpretability of medical images, allowing doctors to understand how a diagnosis was achieved at the image and pixel levels. The proposed approach involves training a convolutional neural network (CNN) on a large dataset of medical images, followed by interpreting the CNN's predictions using numerous XAI techniques such as saliency maps, class activation maps, and layer-wise relevance propagation. The outcomes of the various XAI methodologies are integrated to provide a thorough explanation of CNN's diagnosis. The proposed framework was tested on two publicly available medical imaging datasets and found to have high accuracy and interpretability.

The authors of paper [19] describe RetainVis, a visual analytics tool that analyzes electronic medical records (EMRs) and identifies relevant patient information using interpretable and interactive recurrent neural networks (RNNs). Pre-processing the EMRs to extract key clinical traits, which are subsequently used to train an RNN, is the proposed tool. The RNN is intended to capture the temporal relationships between clinical parameters and predict patient outcomes such as hospital readmission or mortality. The RetainVis tool provides a visual interface that enables doctors to interact with the RNN and investigate the key elements and temporal patterns that drive the predictions. To assist doctors in understanding how the RNN makes predictions, the proposed tool contains multiple interpretability strategies, such as attention processes and feature importance scores. The suggested tool was tested on a real-world EMR dataset, and the findings demonstrated that it delivered good accuracy and interpretability.

The authors of paper [20] proposed an explainable convolutional neural network for glaucoma detection. The processing is performed on color fundus image data. Histogram equalization and contrast-limited adaptive histogram equalization are used to improve the color fundus images. This enhanced image data is used by an explainable convolutional neural network. The XAI was made possible by Class Activation Mapping (CAM), which offers heat map-based explanations for the visual analysis carried out by the CNN. The ORIGA-Light retinal image dataset was one of the three datasets used, and it had the highest mean values with an accuracy of 93.5%, a precision of 93.8%, and an F1 score of 95.7%. This dataset contains a total of 650 retinal image data points, including 168 glacuma and 482 normal cases.

The authors of paper [21] suggested a new CNN model that is lightweight and can classify images related to COVID-19, pneumonia, and tuberculosis. They have also created a framework for generating explanations for the classifications made by the model. The CNN model achieved high accuracy rates of 94.31% in testing and 94.54% in validation, and the generated explanations were validated by medical experts using SHAP, LIME, and GradCam XAI algorithms. The study suggests that the proposed model, along with XAI, can be useful in identifying and categorizing lung diseases. Compared to existing methods, this model has a simpler architecture and performs better when it comes to classifying CXR images with the help of XAI.

The authors of paper [22] present an explainable AI-driven decision-support system for COVID-19 diagnosis based on fused classification and segmentation. The proposed approach involves using image enhancement techniques to preprocess chest X-ray images and a convolutional neural network to segregate lung areas. The segmented images are then fed into a fused classification and segmentation model, which predicts COVID-19 infection using an ensemble of multiple CNNs. The proposed method generates saliency maps and attention heatmaps that highlight the portions of the X-ray image that are most relevant for the diagnosis, resulting in an interpretable approach to COVID-19 diagnosis. The method also generates feature importance scores, which assist doctors in understanding which clinical aspects drive the diagnosis. The proposed system was tested on a large dataset of COVID-19 and non-COVID-19 chest X-ray images, and the findings demonstrated that it was accurate and interpretable.

## III. Potential Technologies

We propose a system that uses pre-trained neural networks instead of deep learning networks developed from scratch, since they offer the following benefits:

1) Reduced training time: Because pre-trained models have already been trained on huge datasets, you can save time and money by bypassing the time-consuming training procedure. You can fine-tune the pre-trained model on your specific dataset in a relatively short period of time.
2) Increased accuracy: Because pre-trained models have already learned significant features and patterns from huge datasets, they can perform better on your unique task. This is especially true if your dataset is small or close to the data used to train the pre-trained model.
3) Transfer learning: Pre-trained models can be used as a starting point for transfer learning, which is the process of adapting a pre-trained model to a new task. Transfer learning can help you exploit the knowledge obtained by the pre-trained model, which is especially valuable if you have little data for your specific task.
4) Generalization: Pre-trained models have been trained on enormous amounts of data and have learned generic characteristics that may be applied to a wide range of tasks. As a result, they can be used for a variety of applications, ranging from computer vision to natural language processing.
5) Accessibility: Because pre-trained models are frequently available as open source code, researchers and developers can quickly use them. This enables a broader spectrum of people to experiment with and improve on the models, resulting in faster advancement in the field.

After performing classification using a pre-trained neural network, we suggest employing XAI techniques to overcome the black-box model's main drawback and learn how and why a specific conclusion is obtained. When working with image data, there are several techniques that may be employed, and the ideal technique depends on the unique use case and the type of image data involved. Below are some of the most often used XAI approaches for image data, along with their benefits:

1) Grad-CAM (Gradient-weighted Class Activation Mapping): This technique reveals crucial parts of an image that contributed the most to a neural network's classification judgment. Grad-CAM is frequently used for object detection tasks and can assist users in understanding what features the model looks for when reaching a classification decision.
2) LIME (Local Interpretable Model-Agnostic Explanations): LIME is a method that produces interpretable explanations for individual black-box model predictions. LIME generates a sequence of perturbations to the original image and trains a local interpretable model on these perturbations to explain the black-box model's behavior.
3) SHAP (SHapley Additive exPlanations): SHAP is a way of describing any machine learning model's output. It uses Shapley values to determine the contribution of each feature to the model's prediction. SHAP can help users comprehend which pixels of an image contributed the most to a certain categorization decision in the context of image data.

Generally, the optimal XAI technique for image data is determined by the unique use case and type of image data. It is frequently advantageous to employ a combination of techniques in order to acquire a thorough understanding of how the model makes predictions.

It is crucial to highlight that there is no "better" XAI technique for image data because each has its own set of strengths and shortcomings and may be better suited to different use cases. Nonetheless, the following are some probable reasons why Grad-CAM may be preferable over LIME and SHAP in some situations:

1) Grad-CAM highlights crucial regions in the image that contributed to the model's classification conclusion, making it simple to visually analyze and comprehend what features the model seeks. LIME and SHAP, on the other hand, generate perturbations or feature importances that are not directly related to image regions, making it more difficult to analyze and explain why a model generated a specific prediction.
2) Grad-CAM is built exclusively for convolutional neural networks (CNNs) used in image classification tasks; hence, it may be more tuned for this sort of data than LIME and SHAP, which are model-agnostic and can be applied to any type of model.
3) Grad-CAM can be used to create heatmaps that depict the most relevant regions of an image for a specific classification decision, making it simple to communicate findings to non-technical stakeholders. LIME and SHAP, on the other hand, produce features of importance or disturbances that may be more difficult to visualize and express well.

However, while choosing a XAI strategy, it is critical to evaluate the unique use case and requirements, as each method has its own set of tradeoffs and limits.

## IV. Prospects and Challenges for the Future

The adoption of Explainable AI (XAI) in clinical and medical practices faces several challenges. There are various obstacles to Explainable AI's (XAI) acceptance in clinical and medical settings. One of the most significant obstacles for XAI approaches in healthcare is a lack of standardization and norms. Various XAI methodologies have varied strengths and drawbacks, and healthcare practitioners' levels of skill in evaluating and understanding AI models may vary. As a result, defined criteria and standards for the development and validation of XAI models in healthcare, as well as training and education for healthcare professionals on how to interpret and apply these models, are required.

Another challenge is the requirement for XAI techniques that are easy to use and interpret. AI models can be complicated and difficult to comprehend, particularly for healthcare professionals with limited technical knowledge. As a result, XAI methods must be created with usability in mind so that they are accessible and understood by a broad range of users. This includes using visuals, natural language explanations, and interactive interfaces to allow users to communicate with AI models.

Concerns about privacy and security are also significant barriers to the use of XAI in healthcare. Patient data is extremely sensitive; therefore, it is critical that XAI models be designed and applied in a manner that safeguards patient privacy and data security. Furthermore, AI models are susceptible to bias, which might result in unfair and discriminatory results. As a result, addressing these challenges through responsible AI practices such as data governance, transparency, and accountability is critical.

The lack of interoperability and connection with established clinical workflows is also a big obstacle when it comes to XAI adoption in healthcare. To be effective, AI systems must be smoothly integrated into clinical workflows, which involves significant investment in terms of resources, infrastructure, and training.

Using XAI in clinical settings may necessitate significant resources, including financial investment and healthcare professional training. Cost and resource constraints may limit XAI's use in healthcare.

Furthermore, given the specific situation and the audience, the XAI approach and the level of information in the explanation must be carefully chosen. For example, if the model is being used for clinical decision-making, the explanation should be precise and particular, whereas a shorter and more generic explanation may suffice for regulatory compliance.

Finally, the verification and validation of XAI results by a medical professional are critical steps in ensuring the AI system's safety, efficacy, and ethical application in clinical and medical practices. Even if the XAI model is highly accurate, there may be biases, flaws, or restrictions in the data or the model that, if not recognized and remedied, could lead to wrong or dangerous conclusions.

Therefore, the use of XAI in healthcare necessitates a careful evaluation of the aforementioned problems as well as the creation of interpretable, transparent, and ethical and legal XAI methodologies. Collaboration between technological experts, healthcare professionals, and regulatory organizations is required to achieve this.

## REFERENCES


[1] "Causes of blindness in India," Medindia. [Online]. Available: https://www.medindia.net/health_statistics/general/causesblindness.asp. [Accessed: 06-Dec-2022].

[2] M. S. Mahmud Khan, M. Ahmed, R. Z. Rasel and M. Monirujjaman Khan, "Cataract Detection Using Convolutional Neural Network with VGG-19 Model," 2021 IEEE World AI IoT Congress (AIIoT), Seattle, WA, USA, 2021, pp. 0209-0212, doi: 10.1109/AIIoT52608.2021.9454244.

[3] J. Rana and S. M. Galib, "Cataract detection using smartphone," 2017 3rd International Conference on Electrical Information and Communication Technology (EICT), 2017, pp. 1-4, doi: 10.1109/EICT.2017.8275136.

[4] B. Askarian, P. Ho and J. W. Chong, "Detecting Cataract Using Smartphones," in IEEE Journal of Translational Engineering in Health and Medicine, vol. 9, pp. 1-10, 2021, Art no. 3800110, doi: 10.1109/JTEHM.2021.3074597.

[5] Imran, A., Li, J., Pei, Y. et al. Fundus image-based cataract classification using a hybrid convolutional and recurrent neural network. Vis Comput 37, 2407–2417 (2021). https://doi.org/10.1007/s00371-020-01994-3

[6] Sirajudeen, A., Balasubramaniam, A. & Karthikeyan, S. Novel angular binary pattern (NABP) and kernel based convolutional neural networks classifier for cataract detection. Multimed Tools Appl 81, 38485–38512 (2022). https://doi.org/10.1007/s11042-022-13092-8

[7] Hu, S., Luan, X., Wu, H. et al. ACCV: automatic classification algorithm of cataract video based on deep learning. BioMed Eng OnLine 20, 78 (2021). https://doi.org/10.1186/s12938-021-00906-3

[8] C. Li, J. Ye, J. He, S. Wang, Y. Qiao and L. Gu, "Dense Correlation Network for Automated Multi-Label Ocular Disease Detection with Paired Color Fundus Photographs," 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), Iowa City, IA, USA, 2020, pp. 1-4, doi: 10.1109/ISBI45749.2020.9098340.

[9] W. Song, P. Wang, X. Zhang and Q. Wang, "Semi-Supervised Learning Based on Cataract Classification and Grading," 2016 IEEE 40th Annual Computer Software and Applications Conference (COMPSAC), Atlanta, GA, USA, 2016, pp. 641-646, doi: 10.1109/COMPSAC.2016.227.

[10] I. Jindal, P. Gupta and A. Goyal, "Cataract Detection using Digital Image Processing," 2019 Global Conference for Advancement in Technology (GCAT), Bangalore, India, 2019, pp. 1-4, doi: 10.1109/GCAT47503.2019.8978316.

[11] Turimerla Pratap, Priyanka Kokil, "Efficient network selection for computer-aided cataract diagnosis under noisy environment", Computer Methods and Programs in Biomedicine, Volume 200, 2021, 105927, ISSN 0169-2607, https://doi.org/10.1016/j.cmpb.2021.105927.

[12] R. R. Maaliw et al., "Cataract Detection and Grading Using Ensemble Neural Networks and Transfer Learning," 2022 IEEE 13th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), 2022, pp. 0074-0081, doi: 10.1109/IEMCON56893.2022.9946550.

[13] H. H. Ali, A. Y. Al-Sultan and E. H. Al-Saadi, "Cataract Disease Detection Used Deep Convolution Neural Network," 2022 5th International Conference on Engineering Technology and its Applications (IICETA), 2022, pp. 102-108, doi: 10.1109/IICETA54559.2022.9888634.

[14] U. Pilania, C. Diwakar, K. Arora and S. Chaudhary, "An Optimized Hybrid approach to Detect Cataract," 2022 IEEE Global Conference on Computing, Power and Communication Technologies (GlobConPT), 2022, pp. 1-5, doi: 10.1109/GlobConPT57482.2022.9938266.

[15] E. Tjoa and C. Guan, "A Survey on Explainable Artificial Intelligence (XAI): Toward Medical XAI," in IEEE Transactions on Neural Networks and Learning Systems, vol. 32, no. 11, pp. 4793-4813, Nov. 2021, doi: 10.1109/TNNLS.2020.3027314.

[16] Bas H.M. van der Velden, Hugo J. Kuijf, Kenneth G.A. Gilhuijs, Max A. Viergever, "Explainable artificial intelligence (XAI) in deep learning-based medical image analysis," Medical Image Analysis, Volume 79, 2022, 102470, ISSN 1361-8415, https://doi.org/10.1016/j.media.2022.102470.

[17] Muddamsetty, S.M., Jahromi, M.N.S., Moeslund, T.B. (2021). Expert Level Evaluations for Explainable AI (XAI) Methods in the Medical Domain. In: , et al. Pattern Recognition. ICPR International Workshops and Challenges. ICPR 2021. Lecture Notes in Computer Science(), vol 12663. Springer, Cham. https://doi.org/10.1007/978-3-030-68796-0_3

[18] S. H. P. Abeyagunasekera, Y. Perera, K. Chamara, U. Kaushalya, P. Sumathipala and O. Senaweera, "LISA : Enhance the explainability of medical images unifying current XAI techniques," 2022 IEEE 7th International conference for Convergence in Technology (I2CT), Mumbai, India, 2022, pp. 1-9, doi: 10.1109/I2CT54291.2022.9824840.

[19] B. C. Kwon et al., "RetainVis: Visual Analytics with Interpretable and Interactive Recurrent Neural Networks on Electronic Medical Records," in IEEE Transactions on Visualization and Computer Graphics, vol. 25, no. 1, pp. 299-309, Jan. 2019, doi: 10.1109/TVCG.2018.2865027.

[20] Omer Deperlioglu, Utku Kose, Deepak Gupta, Ashish Khanna, Fabio Giampaolo, Giancarlo Fortino, "Explainable framework for Glaucoma diagnosis by image processing and convolutional neural network synergy: Analysis with doctor evaluation," Future Generation Computer Systems, Volume 129, 2022, Pages 152-169, ISSN 0167-739X, https://doi.org/10.1016/j.future.2021.11.018.

[21] Mohan Bhandari, Tej Bahadur Shahi, Birat Siku, Arjun Neupane,"Explanatory classification of CXR images into COVID-19, Pneumonia and Tuberculosis using deep learning and XAI," Computers in Biology and Medicine, Volume 150, 2022, 106156, ISSN 0010-4825, https://doi.org/10.1016/j.compbiomed.2022.106156.

[22] K Niranjan, S Shankar Kumar, S Vedanth, Dr. S. Chitrakala, "An Explainable AI driven Decision Support System for COVID-19 Diagnosis using Fused Classification and Segmentation," Procedia Computer Science, Volume 218, 2023, Pages 1915-1925, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2023.01.168.