XAI meets Ophthalmology: An Explainable Approach to Cataract Detection using VGG-19 and Grad-CAM

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 the most common cause of blindness in the world, and early detection is critical for successful treatment. Deep learning algorithms have recently demonstrated promising results in automated cataract detection from fundus images. Yet, the inability of these black box models to be interpreted raises concerns regarding their clinical application. In this paper, we present an explainable approach for detecting cataracts using the VGG-19 convolutional neural network (CNN) and the Gradient-weighted Class Activation Mapping (Grad-CAM) visualization technique. We trained and tested our model using 2,112 high-resolution fundus images from a publicly available dataset. According to our findings, the proposed method obtains a very high accuracy of 97%. Furthermore, the Grad-CAM visuals show the image regions that contribute to the model's decision, providing insights into the diagnosis process and increasing the model's trustworthiness. The findings of this study could assist in improving patient outcomes and lowering healthcare costs by allowing for the early detection and diagnosis of cataracts. Overall, the combination of VGG-19 with Grad-CAM offers a viable option in the medical domain for identifying cataracts and understanding CNN-based decisions.**

Index Terms*—*Cataract Detection, Explainable AI, Interpretability, Deep Learning, Decision Support System

I. Introduction

According to the World Health Organization (WHO), at least 2.2 billion people worldwide have a near or distant visual impairment, and at least 1 billion of these cases might have been avoided or are still unaddressed [1]. Cataracts are a major cause of moderate or severe distant vision impairment or blindness among these 1 billion people, accounting for around 94 million or 9.4% [1]. Almost 90% of the individuals who are partially or completely blind are from developing countries, yet around 75% of eyesight loss is treatable [2]. Late stages of ocular diseases always result in substantial visual acuity impairment, which may be permanent; however, 80% of visual impairment is treatable or recoverable if treated early [2].

Cataract is a common eye disease that causes the lens to fog, resulting in poor vision and, if untreated, blindness. Cataracts are a widespread eye condition that affects people of all ages and is the biggest cause of blindness worldwide. Cataract detection at an early stage is critical for timely treatment and improved clinical outcomes. Nonetheless, reliable cataract diagnosis continues to be a significant challenge in ophthalmology.

Cataract detection is an important area of ophthalmology since early detection is key to effective treatment. Cataracts have traditionally been diagnosed through eye examination and medical history. But computer-aided diagnosis (CAD) has grown in importance as a tool for detecting cataracts. Current advances in artificial intelligence and machine learning show promise in terms of enhancing the accuracy and efficiency of cataract detection. Early detection is critical for effective treatment, and computer-aided diagnosis can help immensely. However, one of the major issues with employing these techniques is that they frequently work as black boxes. A black-box model is a type of machine learning model in which the underlying workings or algorithm that generates the result are hidden from the user. In other words, the model is regarded as a "black box" that accepts inputs and generates outputs while concealing the intermediary steps or decision-making processes. As a result, the model's lack of transparency and interpretability might make it difficult to grasp how it makes its predictions or conclusions, limiting its clinical relevance because it demands explanations or accountability for the model's outputs.

In recent years, Explainable Artificial Intelligence (XAI) has emerged as a promising solution to this problem, assisting us in developing more transparent and explainable AI systems that provide insights into the decision-making process of machine learning models. XAI refers to developing AI systems that can explain how they make decisions in a manner understandable to humans. This is significant because, as AI grows more pervasive in our lives, it is critical that we can trust it and have confidence in its decision-making processes. Yet, if AI simply makes decisions without explanation or transparency, it can be difficult for people to comprehend why certain decisions are being made or to have faith in the correctness of those decisions. The purpose of XAI is to make AI systems more explainable in order to promote openness and accountability. This allows us to better understand and manage the risks associated with AI, as well as discover and reduce biases or faults in AI systems. Ultimately, XAI can lead to more accurate, fair, and trustworthy AI systems on which humans may rely.

As a result, in order to overcome all of the disadvantages of traditional machine learning and deep learning models, we propose an approach that uses XAI to increase the interpretability and trustworthiness of deep learning models in medical image analysis. In our paper, we explore the use of XAI in cataract detection by training the VGG-19 model on a large dataset of cataract and non-cataract images. We then employed Grad-CAM, a visualization technique that generates heatmaps to highlight the parts of the image that contributed the most to the model's decision, to provide more insights into the decision-making process of our model. By doing so, we intend to build a more visible and interpretable decision-making process that doctors can use as a decision support system for rapid and accurate cataract diagnosis.

Our approach detected cataracts with high accuracy and gave useful insights into image areas with cataract-related characteristics. This has the potential to increase the accuracy and efficiency of cataract diagnosis as well as the clinical decision-making process in ophthalmology. Doctors can have increased confidence in the diagnosis and treatment of cataracts by gaining insights into the model's decision-making process.

II. Related Work

A. Literature Review of Cataract Detection Techniques

The authors of the paper [3] proposed a cataract diagnosis system employing a convolutional neural network with a pre-trained VGG-19 model on 800 patient fundus images. OpenCV resized the dataset to 224 x 224 pixels. A VGG-19 pre-trained CNN model classified cataract and normal eye images from preprocessed images. Adam optimisation reduced the cost function and improved model performance. The authors of the paper [4] suggested a hybrid convolutional-recurrent neural network (CRNN) for cataract detection and severity classification. 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 fed each dataset subset to pre-trained CNN models. LSTM classified the collected features into four classes after global average pooling. The authors obtained 97% accuracy. The authors of the paper [5] presented a cataract classification using a decision tree and a Bayesian network. Using tri-learning, these supervised learning algorithms find a good hypothesis. Each fundus image's wavelet and texture determine classifier accuracy. It was observed that the wave feature performed better than the texture feature. Cataracts were classified as non-cataract, mild, moderate, or severe using 5378 fundus images. Bayesian networks and J48 obtained 88% and 70% accuracy, respectively. The authors of the paper [6] proposed two algorithms: one for classifying eyes into three categories: healthy, mild, and severe cataracts, and the second method for determining cataract severity. The authors 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 approach calculates an unhealthy eye's pupil and cataract regions. The formula used for calculating the percentage of cataract is:
$Degree = (cataractarea/(pupilarea + cataractarea)) * 100$.

The authors of the paper [7] provided an effective network selection approach to computer-aided cataract detection in noisy environments. First, input images are pre-processed to remove noise, and then numerous deep neural networks are trained on them. After that, a performance metric evaluates the trained networks and selects the best one for the final diagnosis. The proposed method was evaluated on a dataset of cataract images with various noise levels, and the results showed that the chosen network had very good accuracy in noisy environments compared to other methods. Overall, the proposed method can provide an accurate and efficient solution for cataract diagnosis in noisy situations. The authors of the paper [8] proposed ensemble neural networks and transfer learning to detect and grade cataracts. A pre-trained convolutional neural network (CNN) extracts features, and then an ensemble of CNNs categorizes the retrieved data into cataract grades. Transfer learning and fine-tuning train the ensemble model using a large cataract image dataset. The proposed method was tested on two publicly available datasets, and the findings showed that it performed effectively. The authors of the paper [9] proposed a method for detecting cataract disease using deep convolutional neural networks (CNNs). Preprocessing input images to increase contrast and remove artefacts is followed by training a CNN to classify them as cataracts or non-cataracts. The proposed method uses a CNN architecture with numerous convolutional and pooling layers and fully connected classification layers. The proposed approach was tested on a dataset of cataract images, and the results showed that it achieved high accuracy. The authors of the paper [10] developed an optimal cataract detection hybrid using image processing and machine learning algorithms. Before extracting features with GLCM and DWT, the input images are contrast-enhanced and normalized. A cataract detection SVM classifier is trained using these features. A grid search algorithm finds the best SVM hyperparameters to optimize SVM classifier performance. 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 used in Medical Domain

The authors of the paper [11] analyze the current state of XAI in healthcare. The authors examine XAI approaches such as decision trees, decision sets, LIME, SHAP, Grad-CAM, and Activation Maximisation. These approaches are tested using publicly available medical datasets, including the EHR and Chest X-Ray datasets. The authors also explore model-based, decision-based, and instance-based XAI methodologies, such as explainable neural networks, counterfactual reasoning, and prototype-based explanation. The authors identify medical XAI difficulties and potential, such as processing high-dimensional and complicated medical data, delivering clear and actionable explanations, and integrating ethical and legal considerations into XAI design. The authors conclude that XAI offers significant promise to improve medical decision-making accuracy, dependability, and fairness, but additional research is needed to realise this potential. Using anterior and medical imaging modes, the authors of the paper [12] reviewed 223 studies on deep learning-based XAI in healthcare. The majority of studies used model-specific and model-agnostic post-hoc explanations. Local explanations outnumbered global explanations in most publications. The authors' focus on deep learning for medical image analysis explains these findings. Saliency mapping is the most prevalent way for explaining CNN predictions. These methods give model-specific post-hoc explanations. Post-hoc approaches can be used after the neural network is trained, making them easier to employ than model-based XAI methods. The authors of the paper [13] compare medical XAI approaches. Rule-based, model-agnostic, and post-hoc XAI approaches are used for medical datasets such as the Retinal Fundus Image Quality Assessment (RFIQA) and EyePacs datasets. A panel of medical specialists evaluates Grad-CAM and SIDU XAI procedures for transparency, interpretability, and accountability. The study found that XAI approaches have 75%–85% accuracy. The study also shows that XAI systems vary in transparency, interpretability, and accountability, with some providing clear and actionable explanations and others being less interpretable.

The authors of the paper [14] introduce LISA (Local Interpretable and Simulatable Network-Based Analysis), which combines numerous XAI techniques to increase medical image explainability. LISA allows doctors to comprehend how a diagnosis was made at the image and pixel levels through global and local image interpretability. Saliency maps, class activation maps, and layer-wise relevance propagation are used to evaluate the CNN's predictions after training it on a huge dataset of medical images. CNN's diagnosis is explained by integrating XAI methods. The proposed framework was tested on two publicly available medical imaging datasets and found to have high accuracy and interpretability. The authors of the paper [15] describe RetainVis, a visual analytics application that analyses electronic medical records (EMRs) using interpretable and interactive recurrent neural networks (RNNs). The proposed technique extracts clinical features from EMRs to train an RNN. The RNN predicts hospital readmission and mortality by capturing clinical parameters' temporal connections. Doctors can interact with the RNN and examine its essential elements and temporal trends using RetainVis. The proposed technology uses attention processes and feature importance scores to help clinicians understand RNN predictions. On a real-world EMR dataset, the suggested tool showed good accuracy and interpretability.

The authors of the paper [16] proposed a glaucoma-detecting, explainable convolutional neural network. Colour fundus images are processed. Histogram and contrast-limited adaptive histogram equalization improve colour fundus imaging. Explainable convolutional neural networks use this augmented image data. Class Activation Mapping (CAM) heat maps explain CNN visual analysis, enabling the XAI. One of the three datasets, the ORIGA-Light retinal image dataset, has the greatest mean values with 93.5% accuracy, 93.8% precision, and 95.7% F1 score. This dataset has 650 retinal imaging data points, including 168 glaucoma cases and 482 normal cases. The authors of the paper [17] proposed a lightweight CNN model that can categorize COVID-19, pneumonia, and tuberculosis images. They also developed a framework for explaining model classifications. 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 reveals that the model and XAI can identify and classify lung illnesses. This model is simpler and better at classifying CXR images with XAI than other techniques. The authors of the paper [18] present an XAI-driven COVID-19 diagnosis decision-support system based on fused classification and segmentation. Image enhancement and a convolutional neural network are used to preprocess chest X-rays and separate lung regions. A fused classification and segmentation model predicts COVID-19 infection using several CNNs from the segmented images. Saliency maps and attention heatmaps highlight the X-ray image's most relevant parts, making the COVID-19 diagnosis interpretable. Feature importance scores help clinicians identify clinical factors that influence diagnosis. The proposed method was accurate and interpretable on a large dataset of COVID-19 and non-COVID-19 chest X-rays.

III. Dataset

For training and testing, we utilized the dataset "eye_diseases_classification [19]," which is focused on the classification of eye diseases. The dataset includes normal, diabetic retinopathy, cataract, and glaucoma retinal images, with roughly 1000 images in each class. The images were gathered from numerous sources, such as IDRiD, Oculur Recognition, HRF, and others. We only used cataract (1038 images) and normal (1074 images) eye images because our technique is designed to detect cataracts. The primary reason for using this dataset is that it contains high-resolution images that help train the VGG-19 model because high-resolution images contain more pixels and finer details, allowing the model to capture subtle features and intricate patterns that may be important for accurate classification. This is especially important for our proposed approach because small features can have a big impact on diagnostic accuracy. Furthermore, high-resolution photos provide additional spatial information, which aids the model in properly localizing and detecting specific regions of interest within the image. This is important for recognizing eye disorders since some conditions present in specific areas or structures of the eye. As a result, we have the best probability of accurately detecting cataracts and then improving the interpretability of the cataract detection technique using this dataset.

IV. Techniques

This section describes the main techniques and methodologies used in this paper.

A. VGG-19

VGG-19 is a CNN architecture and a variant of the VGG family of CNN models, which also includes VGG-16, VGG-11, and VGG-13. VGG-19 is a deep learning model with 19 layers, hence the name. The layers are stacked sequentially, with input images passing through each layer and undergoing various transformations until a final output is generated. Figure 1 [20] depicts the architecture of VGG-19, which is characterized by the following features:

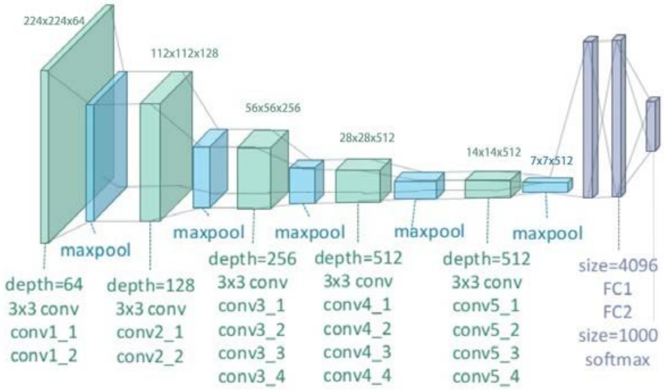

Figure 1. Architecture of VGG-19 [20]

1) Convolutional Layers: The network starts with a series of convolutional layers that extract features from the images and use them as inputs to the next layer. VGG-19 has 16 convolutional layers, each with a kernel size of 3x3 and a stride of 1 pixel.
2) Max-Pooling Layers: There is a max-pooling layer after each convolutional layer, which decreases the spatial dimensions of the feature maps. VGG-19 has 5 max-pooling layers, each with a pooling size of 2x2 and a stride of 2 pixels.
3) Fully-Connected Layers: The network's final layers are fully-connected layers that are responsible for making final predictions. VGG-19 has three completely connected layers, the last of which is a softmax layer that generates the final output.
4) ReLU Activation Function: After each convolutional and fully-connected layer, the Rectified Linear Unit (ReLU) activation function is utilized. This activation function adds nonlinearity to the network, allowing it to learn more complex features.
5) Dropout Regularization: Dropout regularization is applied to the fully-connected layers to prevent overfitting. This strategy randomly removes a particular number of neurons during training, preventing the network from becoming overly reliant on any one attribute.

Because of its high accuracy and simplicity, VGG-19 has been widely employed in a number of computer vision applications, such as object recognition, image classification, and image segmentation. It has achieved cutting-edge performance on a variety of benchmark datasets and has served as a foundation for the development of more advanced CNN architectures. The following are the key reasons why we chose VGG-19 for our proposed system:

1) High Accuracy: VGG-19 has shown cutting-edge performance on a number of benchmark datasets, such as ImageNet, which is a large-scale task for classifying images. This makes it an excellent choice for situations requiring high accuracy.
2) Transfer Learning: VGG-19 has been pre-trained on huge datasets like ImageNet, making it an excellent starting point for transfer learning. Transfer learning is the process of fine-tuning a model that has already been trained on a whole new set of data. This can save time and improve performance.
3) Simple Architecture: The architecture of VGG-19 is very simple compared to more modern CNN models like ResNet and Inception.This simplifies understanding and implementation.

In our method, which involves classifying images, using a model that has already been trained, like VGG-19, is much better than making a deep learning CNN from scratch. First of all, VGG-19 has been trained on large datasets like ImageNet, so it knows how to recognize a wide range of image characteristics that help with image classification. When compared to a CNN built from scratch, this can lead to better accuracy and less overfitting, especially when dealing with limited training data. In our case, we only had a small number of medical images to train and test our model on because medical image data is very sensitive and not widely available. Second, the VGG-19 model's pre-trained weights can be used for transfer learning, which lets us fine-tune our model for certain classification tasks with less training data. This can greatly cut down on the amount of time and computing power needed to train a CNN from scratch. Finally, because VGG-19 has been widely embraced and well-tested by the deep learning community, it is supported by a range of tools and frameworks, making it easy to use and incorporate into current projects.

B. Grad-CAM

Grad-CAM is a XAI technique for visualizing portions of an input image that are used for generating a prediction by CNN. It aids in identifying the critical portions of an image that are most crucial to the model's final classification decision. Grad-CAM generates a heat map that highlights the most important parts of an input image that contribute to a given prediction provided by a CNN model.

Grad-CAM is model-agnostic, which means it may be applied to any CNN model regardless of architecture. Grad-CAM only requires the model's final convolutional layer, which is included in practically all CNN architectures. The heat map, also known as a class activation map, is produced by weighting the feature maps of the CNN's final convolutional layer with the gradients of the target class score in relation to the feature maps. The weighted feature maps that arise are subsequently passed through a global average pooling operation to produce a single activation map that is the same size as the final convolutional layer. The activation map is then upsampled to the original input image size and superimposed on it to visualize the portions of the input image that contributed the most to the prediction. The heat map highlights the image portions that are most essential to the target class, with warmer colours indicating greater value. Grad-CAM can thus be used to visualize the relevant regions of an image for any CNN model, as long as the final convolutional layer is available.

In our proposed approach, we take an image as input from the system's user. When working with image data, we selected Grad-CAM XAI because it allows us to visualize and comprehend the decision-making process of a CNN model. It helps understand which features the model is focusing on and what visual patterns it is learning to make decisions by providing a heat map of the most critical parts of an input image that contribute to a certain prediction. This is especially crucial when the model's decision-making process is difficult to understand, as in the case of VGG-19. Grad-CAM can provide insights into how the model uses multiple layers and feature maps to arrive at its final prediction in such circumstances. Grad-CAM can also be used to identify and visualize the parts of an image that the model is most sensitive to. This can be helpful for figuring out where the model might be weak or biased. Overall, Grad-CAM is a useful tool for analyzing and evaluating the behaviour of CNN models in general, making it an invaluable asset when working with image data.

V. Methodology and Working

Figure 2 describes the architecture of our proposed system, and its functioning is as follows:

A. Fundus Image

To begin, our system takes a fundus image from the system's user. A fundus image is a medical image of the inside of the eye, particularly of the retina at the back of the eye. The retina is responsible for capturing light and transmitting visual information to the brain. A fundus image can assist doctors in diagnosing and monitoring a wide range of eye problems. The image is obtained with the use of a fundus camera, which takes a high-resolution photograph of the retina. Fundus images can also be used to track the evolution of eye diseases over time and assess treatment effectiveness.

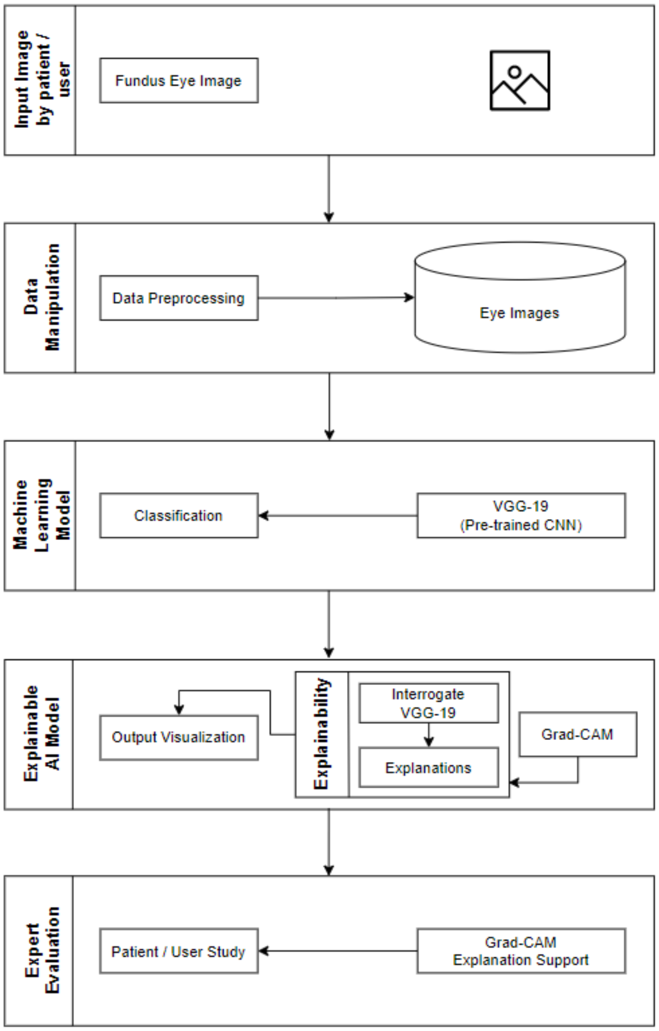

Figure 2. Architecture of the Proposed System

B. Image Pre-processing

Image pre-processing begins with scaling and cropping images to standardize the size of all images in the training dataset and removing undesirable regions or objects to decrease noise and focus on the area of interest. Following that, image normalization is performed to alter the brightness and contrast of the image in order to improve its quality and make it easier to analyze. Finally, filters such as smoothing and sharpening are used to reduce noise and enhance image features.

C. Classification

The input to our classification model is a fundus image with a predetermined size of 224x224 pixels. After that, the image goes through the convolutional layers; VGG-19 has 16, and after each one comes the ReLU activation function. With the help of convolutional layers, information is extracted from the input image at different scales and levels of abstraction. It takes an input image and runs it through a series of filters, or kernels, to produce a set of feature maps. The filters are learned to detect specific patterns or features in the training data. As more convolutional layers are added, the returned features become increasingly complicated and abstract. Low-level features, such as edges and corners, are captured by the first few layers, whereas high-level features, such as shapes and objects, are captured by the subsequent layers. The feature maps generated by the convolutional layers are passed on to the subsequent layers for further analysis and classification. Each convolutional layer employs a stride of 1, and its filters are of 3x3 size. VGG-19 can accurately capture fine-grained information due to the utilization of small filter sizes and stride values.

Then, there are five max-pooling layers in VGG-19. The convolutional layers' output is sent to the max-pooling layers, which decrease the feature maps' spatial dimensions while keeping the most important details. Because of this, the network can learn faster and more accurately. A 2x2 fixed filter with a stride of 2 is used in each max-pooling layer. This process involves partitioning the input feature map into non-overlapping 2x2 sections and keeping just the highest value within each section. This process effectively downsamples the feature map and introduces translation invariance, both of which make the network more robust against noise in the input. The computational cost is reduced, and overfitting is avoided by producing a feature map that is half the width and height of the input feature map.

After that, VGG-19 uses three fully-connected layers to identify whether or not the eye in the input image has cataracts. The many parameters available in the fully-connected layers enable the network to acquire complex, non-linear mappings from input to output. High-level information is extracted by the convolutional and max-pooling layers, and the fully-connected layers transform it into a format suitable for classification. The flattened output of the last max-pooling layer is fed into the fully-connected layers, where it undergoes a series of linear transformations. Learnable weights and biases are used to fine-tune these modifications during training via backpropagation. The fully-connected layers of VGG-19 have a total of 4096 units, and the activation function ReLU is applied to the layer outputs to introduce nonlinearity. The vanishing gradient problem is avoided, and more complex features can be learned by the network when ReLU activation is used. Finally, a softmax layer generates the final probability distribution across the output classes in VGG-19 after the final fully-connected layer. The softmax function normalizes the outputs of the last fully-connected layer to generate a probability distribution over all possible output classes.

D. XAI Model

In our proposed system, we use Grad-CAM XAI to explain the output of the black-box classification model VGG-19. Grad-CAM is a technique for visualizing the portions of the input image that are most relevant for the network's prediction in order to explain the output of VGG-19. The technique produces a class activation map that highlights the parts of the input image that are most significant to the network's prediction by using the gradients of the output class score with respect to the feature maps created by the last convolutional layer. This visualization explains how the network makes predictions and gives us insight into its decision-making process.

Grad-CAM initially passes the input image through the network during forward propagation, and the activations of the last convolutional layer and the fully connected layers are computed. The predicted class's class score is calculated using the output of the final fully connected layer and the softmax function. The class score shows the network's level of confidence in its prediction. The gradient of the output class score with respect to the activations of the final convolutional layer is determined during backward propagation. This gradient denotes the significance of each activation in predicting the target class. The gradient of the output class score with respect to the activations of the final convolutional layer is then multiplied by the activations, yielding a weighted activation map. This map displays the most relevant parts of the input image for the network's prediction. After that, the weighted activation map is averaged over the spatial dimensions to generate a single activation value for each feature map. This yields a class activation vector, which represents the significance of each feature map in predicting the target class. Finally, the class activation vector is used to generate a heat map that shows which portions of the input image are most relevant for the network's prediction. This heat map is superimposed on top of the input image to create a visualization of the regions most significant to the network's prediction. This heat map explains why VGG-19 classified a particular eye as a cataract or a normal eye by highlighting the parts of the image that enabled VGG-19 to reach that conclusion.

E. Expert Evaluation

Expert doctor evaluation is critical when working with XAI, particularly in the medical field, because it provides a means to validate and verify the AI model's output. XAI approaches can provide useful insights into how an AI model makes predictions, but they are not perfect and require specialized domain knowledge to interpret. Our system generates heat maps to illustrate the output of VGG-19, but most of the time, only medical professionals can comprehend and analyze them. As a result, only professionals can determine the accuracy of the output, and this expert validation is required because no system is 100% accurate, and even a single erroneous decision in this particular scenario can have serious consequences.

Expert doctors in the medical field have years of training and experience diagnosing illnesses and analyzing imaging results. They are well-versed in the subtleties of various diseases and can provide valuable feedback on the accuracy and dependability of an AI model's predictions. Expert doctor evaluation can assist in identifying mistakes or biases in the AI model's predictions that XAI techniques may have overlooked. They can also assist in identifying circumstances when the AI model's predictions are true but require additional investigation or testing. Furthermore, expert doctor evaluation can aid in the development of confidence and adoption of AI models in the medical arena. Doctors are frequently the end-users of

AI systems in healthcare, and their buy-in and confidence are important for the systems' effective implementation.

VI. Results

We evaluated our system on around 425 images and achieved an extremely high accuracy of 97% using the VGG-19 algorithm for classifying images in two categories: a healthy eye or a cataract eye. Following that, we used two XAI, Grad-CAM and Grad-CAM++, to explain the result of the VGG-19 algorithm, namely why VGG-19 predicted a specific image as a cataract-infected eye or a healthy eye.

Grad-CAM and Grad-CAM++ are two methods of visualizing and comprehending deep neural network decisions. They are based on the class activation mapping concept, which seeks to highlight the portions of an input image that are most relevant to the network's prediction. The fundamental distinction between Grad-CAM and Grad-CAM++ is in how the gradients are computed and the class activation maps are generated. Grad-CAM computes the gradients of the target class in relation to the feature mappings in a CNN's final convolutional layer. The importance weights for each feature map are then calculated by globally averaging these gradients. The feature maps are multiplied by their weights and added together to form a weighted combination that represents the class activation map. Finally, for visualization, the class activation map is upsampled to the original input image size. Grad-CAM++ enhances Grad-CAM by including additional localization cues from multiple convolutional layers. Grad-CAM++ computes gradients from all convolutional layers in the network rather than just the last convolutional layer. The gradients are weighted according to their importance, which is defined by their global average pooling values. These weighted gradients are then integrated across layers to produce a detailed class activation map that takes into account many levels of abstraction. The class activation map, like Grad-CAM, is upsampled for visualization. Grad-CAM++, as a result, seeks to provide a more comprehensive understanding of the network's decision-making process by taking into account various levels of abstraction in the CNN architecture.

After generating the heatmaps, we masked them on the actual input image to acquire a better understanding of the regions of the eye that are most important in producing a particular result. Figure 2 shows the heatmap produced by Grad-CAM when an image of a healthy eye is used as input, and Figure 3 shows the heatmap produced by Grad-CAM when an image of a cataract-infected eye is used as input. The heatmap generated by Grad-CAM++ when the input is an image of a healthy eye is shown in Figure 5, and the heatmap generated by Grad-CAM++ when the input is an image of a cataract-infected eye is shown in Figure 6. The warmer portions of the image, particularly the red regions but also the yellow regions, show that these are the most crucial components that contributed to VGG-19 producing that specific outcome. As can be seen, Grad-CAM++ can more accurately interpret VGG-19 results and recognize more regions that are responsible for the specific output than regular Grad-CAM.

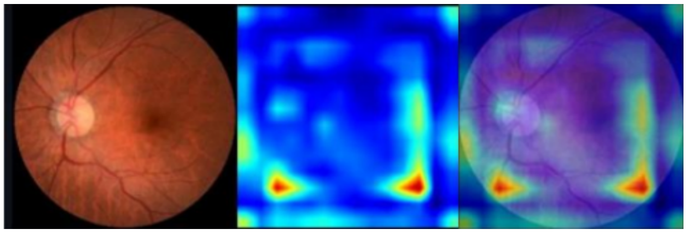

Figure 3. Grad-CAM output when the input image is of a healthy eye

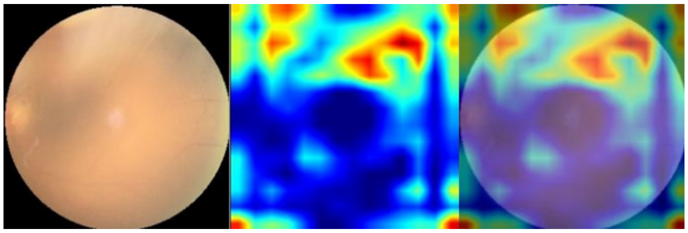

Figure 4. Grad-CAM output when the input image is of a cataract eye

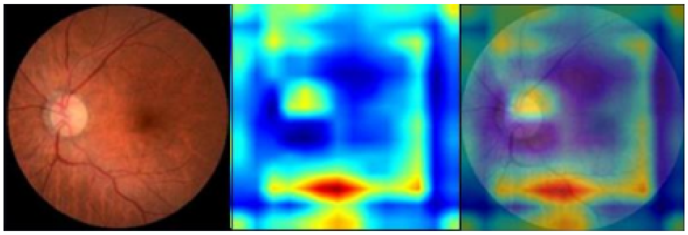

Figure 5. Grad-CAM++ output when the input image is of a healthy eye

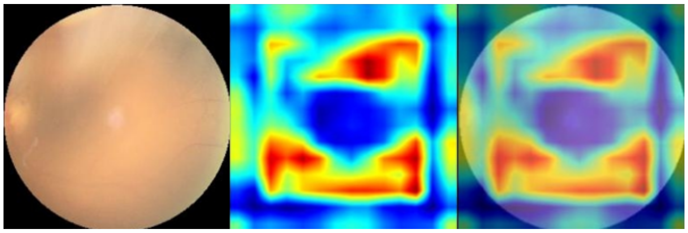

Figure 6. Grad-CAM++ output when the input image is of a cataract eye

VII. Conclusion

Overall, our findings show that employing the VGG-19 deep learning model for early cataract detection is effective. The model identified cataract from fundus images with 97% accuracy, demonstrating its potential as an important diagnostic tool for ophthalmologists and medical professionals.

One of the study's major findings is the usefulness of applying XAI techniques in medical diagnostics. Because of the increasing complexity of modern machine learning algorithms, it is frequently difficult to fully understand how a model arrives at a specific choice. XAI bridges this gap by providing clear insights into the decision-making process, which is especially crucial in the medical field, where decisions can have life-or-death implications. As a result, we were able to acquire insights into the model's decision-making process and visualize the regions of interest in the fundus images that contributed to its predictions by using Grad-CAM, a XAI approach. This additional layer of interpretability and transparency can aid in the model's performance and facilitate its adoption into clinical practise.

In conclusion, our research shows that deep learning and XAI techniques have the potential to improve the accuracy and interpretability of medical image analysis, thereby improving patient outcomes and minimizing healthcare expenditures. The results of our research can have important implications for the field of ophthalmology as well as the larger medical community. XAI-based diagnostic tools, with further development and refinement, could provide essential support for medical professionals, enabling faster and more accurate diagnoses and, eventually, better patient outcomes.

VIII. Future Work

While our proposed model for diagnosing cataracts using VGG-19 has shown good results, accuracy can still be improved. Additional data augmentation techniques, such as rotation, scaling, or noise injection, could also potentially increase the model's robustness and ability to generalize to previously unseen data. Other deep learning models, like ResNet or Inception, could also be looked at to see if they could make the system more accurate and reliable. Our proposed approach could also be enhanced to detect varied degrees of cataract severity.

Even though our model was trained and tested on a set of images, future research could look into whether the model could be validated in a real-world clinical setting. This could involve testing the model on images taken by doctors or using the model to find cataracts in a real-world setting. Real-world validation is important to show that the model is clinically relevant and can be used well in a clinical setting. The proposed model is a stand-alone cataract detection tool. Future studies could look into how the model could be integrated with clinical workflows to increase the efficiency and accuracy of cataract diagnosis. This could result in faster and more accurate diagnoses for patients.

References

[1] "Vision Impairment and blindness," World Health Organization, 13-Oct-2022. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment. [Accessed: 05-Apr-2023].

[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] 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.

[4] 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

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] 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.

[13] 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

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] "eye_diseases_classification," eye_diseases_classification — Kaggle. /datasets/gunavenkatdoddi/eye-diseases-classification

[20] Zheng, Yufeng & Yang, Clifford Merkulov, Aleksey. (2018). Breast cancer screening using convolutional neural network and follow-up digital mammography. 4. 10.1117/12.2304564.