Research Article

Towards Trustworthy Myopia Detection: Integration Methodology of Deep Learning Approach, XAI Visualization, and User Interface System

Worood Esam Noori [1], , A. S. Albahri [2], *,

[1] College of Medical Informatics, University of Information Technology and Communications, Baghdad, Iraq

[2] Iraqi Commission for Computers and Informatics (ICCI), Baghdad, Iraq

ARTICLEINFO

Abstract

Myopia, a prevalent vision disorder with potential complications if untreated, requires early and accurate detection for effective treatment. However, traditional diagnostic methods often lack trustworthiness and explainability, leading to biases and mistrust. This study presents a four-phase methodology to develop a robust myopia detection system. In the initial phase, the dataset containing training and testing images is located, preprocessed, and balanced. Subsequently, two models are deployed: a pre-trained VGG16 model renowned for image classification tasks, and a sequential CNN with convolution layers. Performance evaluation metrics such as accuracy, recall, F1-Score, sensitivity, and log loss are utilized to assess the models' effectiveness. The third phase integrates explainability, trustworthiness, and transparency through the application of Explainable Artificial Intelligence (XAI) techniques. Specifically, Local Interpretable Model-Agnostic Explanations (LIME) are employed to provide insights into the decision-making process of the deep learning model, offering explanations for the classification of images as myopic or normal. In the final phase, a user interface is implemented for the myopia detection and XAI model, bringing together the a forementioned phases. The outcomes of this study contribute to the advancement of objective and explainable diagnostic methods in the field of myopia detection. Notably, the VGG16 model achieves an impressive accuracy of 96%, highlighting its efficacy in diagnosing myopia. The LIME results provide valuable interpretations for myopia cases. The proposed methodology enhances transparency, interpretability, and trust in the myopia detection process.

1. INTRODUCTION

Myopia, also known as nearsightedness, is a prevalent eye condition characterized by difficulty seeing distant objects clearly. It occurs when the cornea is excessively curved or the eyeball is elongated, leading to the focusing of light in front of the retina rather than directly on it. As a result, distant objects appear blurry, while close objects can be seen more clearly [1]. The diagnosis of myopia is typically done during an eye examination, which involves several tests. Visual acuity testing assesses the ability to see objects at various distances, while a refraction test determines the precise prescription needed for corrective lenses or glasses. A retinal examination is also conducted to examine the back of the eyes for signs of myopia. Additionally, information about the patient's family history and daily activities may be gathered to assess the risk factors associated with myopia [2].

Myopia can be classified based on different factors, such as physiological myopia, which is associated with the natural growth of the eye and characterized by an increase in the axial diameter [3]. On the other hand, pathological myopia involves abnormal elongation of the eyeball and is often accompanied by thinning of the scleral wall. Myopia can also be categorized based on the age of onset, with congenital myopia appearing at birth and school myopia developing during childhood or adolescence [4]. The severity of myopia is further classified into three levels: low, moderate, and high. Low myopia refers to a refractive error of up to -3.00 diopters (D) and may cause slightly hazy distance vision, particularly in low light conditions. Moderate myopia ranges from -3.00 to -6.00 D and can result in more pronounced blurring of distance vision. High myopia, which exceeds -6.00 D, is associated with a higher risk of vision impairment [5]. Timely detection and appropriate treatment of myopia are crucial, as higher levels of myopia are linked to an increased risk of side effects such as glaucoma, myopic macular degeneration, and retinal detachment. Patients with high myopia, in particular, are more

*Corresponding author. Email: ahmed.bahri1978@gmail.com

prone to developing these problems, which can lead to irreversible vision loss. Therefore, early detection and proper management are essential in preventing vision loss associated with myopia [6][7].

In recent years, new computational tools based on machine learning have made significant strides in solving perceptual problems, including image recognition, which has increasingly influenced various fields, including the biological sciences [8]–[11]. Among the popular deep neural networks, Convolutional Neural Networks (CNNs) stand out as one of the most effective for image recognition tasks [12]–[15]. These networks are designed to recognize patterns within images, making them well-suited for image-focused tasks and reducing the parameters required to set up the model [16]. Transfer learning, a technique that leverages knowledge from pre-trained models, has further enhanced the performance of machine learning algorithms in new tasks and domains [17]–[19]. The VGG-16 architecture, with its stacked convolutional layers, has become a standard benchmark for understanding fundamental design principles in deep learning [20].

To detect myopia using deep learning models, several challenges need to be addressed. Myopia is a complex condition influenced by a wide range of hereditary and environmental risk factors, and different patients may exhibit various visual impairments. Developing a universal deep learning model for myopia detection is challenging due to the unique variations among individuals. Furthermore, deep learning models are often referred to as "black boxes," meaning it is difficult to comprehend their decision-making process. In critical fields like healthcare, interpretability is crucial for ensuring legal and ethical compliance [21]. To address these concerns, AI decision-making "explanations" are necessary to illuminate the reasoning behind the AI's conclusions [22]. These explanations help ensure reasoned arguments and adherence to moral and legal standards [23]. As AI, ML, and DL become more integrated with human-centric applications, explainability becomes increasingly important for scientists, regulators, and the general public, driving the need to open the "black box" for regulatory and social reasons [24][25].

A technique called Local Interpretable Model-Agnostic Explanations (LIME) offers instance-based explanations for any classifier's predictions. These explanations are derived from interpretable representations of the instance and are locally faithful to the instance, regardless of the type of classifier model. LIME uses an interpretable model to locally approximate the classifier and generate explanations. By capturing classifier behavior close to the instance being explained, LIME provides locally faithful explanations [26]. This method subdivides a given image into superpixels to produce explanations, capturing local contextual data about specific areas of the image. Perturbed images are generated by randomly hiding some superpixels, and the impact on the probability of correct class prediction is calculated [27].

To train the deep learning model for myopia detection, an appropriate dataset of myopia image data must be found. The dataset should be sufficiently large, representative, and properly labeled. Building a deep learning detection model involves choosing a suitable deep learning architecture, training the model on the myopia dataset, and optimizing its performance. Understanding the deep learning model's operation and providing a detailed analysis of the decision-making process increases user trust and the model's accuracy and dependability. Developing a user-friendly interface for users to submit their myopia images and obtain prompt feedback on the presence of myopia is crucial for the successful implementation of a myopia detection system. In this study, we propose an integrated methodology for myopia detection that combines a Deep Learning approach, Explainable Artificial Intelligence (XAI) Visualization, and a User Interface System. In this context, this paper aims to address the challenges in myopia detection by proposing an integrated methodology that combines a deep learning approach, XAI visualization, and a user interface system. The primary objective is to develop a trustworthy and reliable myopia detection system that enables early detection and appropriate treatment, thereby mitigating the risk of irreversible vision loss associated with myopia. The contributions of this study are as follows:

1. Identification and preparation of the myopia dataset for training the deep learning model.
2. Construction of a deep learning detection model, including the selection of a suitable architecture, training on the myopia dataset, and performance optimization.
3. Analysis and explanation of the model's decisions using XAI visualization.
4. Development of a user-friendly interface to submit and examine myopia images using the trained model.

2. LITERATURE REVIEW

In recent years, deep learning and computer vision techniques have shown promising potential for disease screening, including the detection of myopia using ocular appearance images [28]. The literature review provides valuable insights into the prevalence and impact of myopia, particularly in children and adults. The use of deep learning and computer vision for myopia screening and detection is explored in several studies, showcasing the potential of these technologies to aid in the early diagnosis and prevention of myopia-related complications. In [28] the researchers utilized deep learning techniques to develop a deep learning system (DLS) for myopia detection using ocular appearance images from children aged 6 to 18. The DLS demonstrated promising performance in detecting myopia, providing an efficient and potentially cost-effective screening method [28]. N Rauf et al. focused on pathological myopia, a severe form of myopia that can lead to blindness. They employed a convolutional neural network (CNN) model to automatically detect pathological myopia from fundus images, achieving high accuracy and demonstrating the potential for automated myopia detection [29]. R

Hemelings et al. assessed the use of CNNs for detecting pathological myopia and performing semantic segmentation of myopia-induced lesions from fundus images. Their CNN model, incorporating domain knowledge and fovea localization through segmentation, showed promising results for PM detection and lesion segmentation [30]. L Lu et al. addressed the growing epidemic of myopia in China and proposed deep learning systems to detect pathological myopia and myopic macular lesions. Their systems demonstrated robust performance, comparable to ophthalmologists and retinal specialists, indicating the feasibility of mass screening for myopic populations on a national scale [31].

In all these studies, deep learning models showed potential for accurate myopia detection and lesion segmentation. However, it is essential to address challenges related to interpretability and explainability of the deep learning models' decisions, especially in medical applications [32], to gain trust and acceptance from medical practitioners and patients. Regarding the datasets used in the studies, the large volume of color retinal fundus images from myopia patients, obtained with ethical approval and adherence to the Declaration of Helsinki, offers a valuable resource for training and validating deep learning models for myopia detection. Overall, the reviewed literature highlights the significance of leveraging deep learning and computer vision techniques in myopia screening, early diagnosis, and intervention, which can contribute to preventing vision impairment and blindness associated with myopia. The combination of advanced technologies and large-scale datasets brings a promising approach to address the challenges posed by myopia, especially in regions with a high prevalence of myopia.

3. METHODOLOGY

The methodology employed in this study serves as a structured approach to tackle the challenges and complexities associated with myopia detection and explanation. By implementing a systematic methodology, the study aims to enhance the diagnosis and explanation of myopia images through a comprehensive decision-making approach. The methodology consists of four distinct phases that collectively contribute to the overall study as shown in Figure 1.
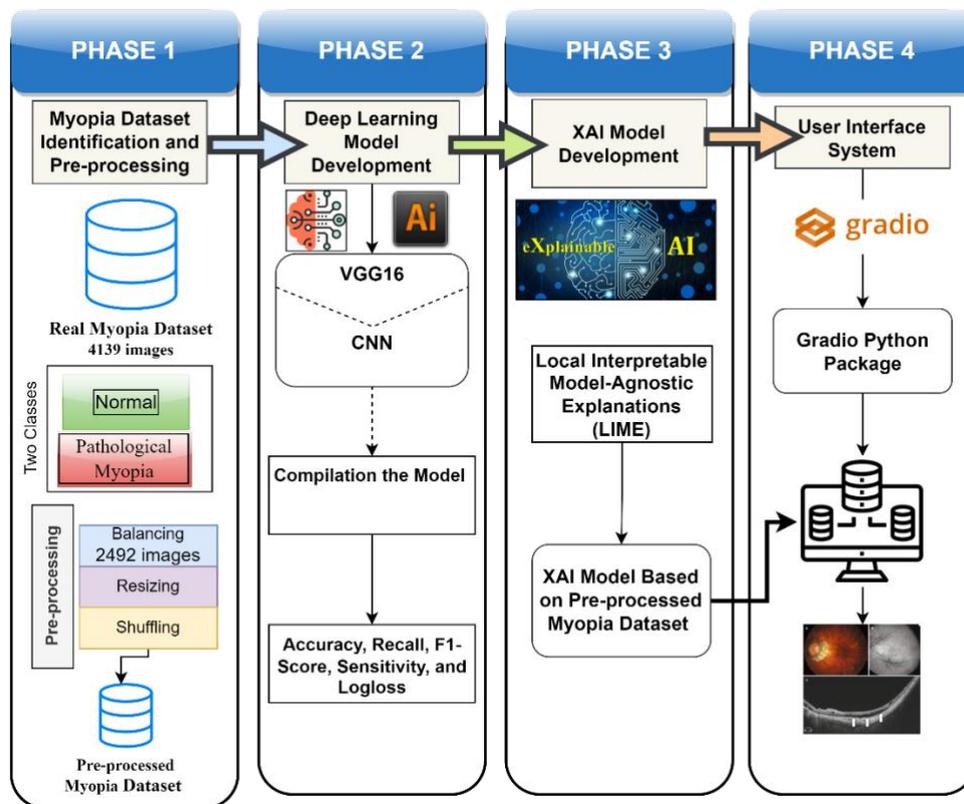

Fig. 1. Integration methodology phases for myopia detection

The first phase focuses on the identification and pre-processing of the myopia dataset. This involves locating and preparing the dataset for subsequent analysis, ensuring its quality and suitability for training and testing the models. The second phase involves the development of the deep learning model. This encompasses constructing the model architecture, compiling it with appropriate optimization algorithms, and fitting it to the prepared dataset. The use of both a pre-trained VGG16 model and a sequential CNN with convolution layers provides a comprehensive approach to myopia detection. The third phase

centers on the development of the XAI model. This is achieved by implementing the LIME technique, which provides interpretability and insights into the decision-making process of the deep learning model. The XAI model enhances transparency and facilitates a better understanding of why specific predictions are made. The final phase involves the development of the user interface system, which integrates all the previous phases. This user-friendly system enables easy interaction with the deep learning model and XAI visualization, allowing healthcare professionals and patients to access accurate and explainable results for myopia diagnosis. By following this structured methodology, the study aims to provide a comprehensive and reliable approach to myopia detection and explanation.

3.1 PHASE 1: Data Identification and Pre-processing

The myopia data used in this study consisted of retinal fundus images obtained from Kaggle [33]. The data was categorized into two classes: myopia and normal. Before proceeding with the pre-processing, the images were divided into two sets based on whether they corresponded to the left or right eye. Figure 2 provides a visualization of sample myopia images from both the left and right eyes.
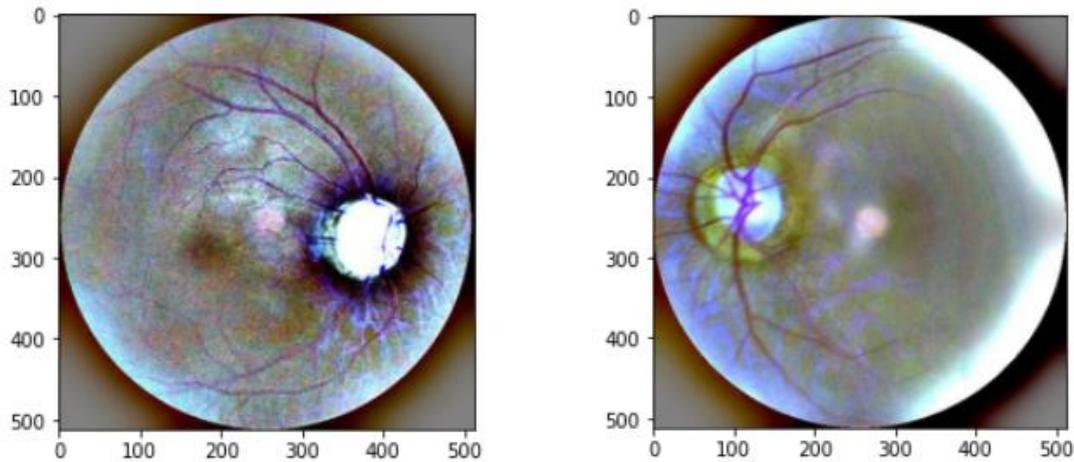

Fig. 2. Samples of left and right eyes affected by myopia

The next step involved preparing the data for model training. Initially, the labels (target values) were separated from the images and organized into a data frame, with each image being assigned the appropriate label based on its category (myopia or normal). This division ensured that each set contained an equal number of images from both categories, thereby avoiding any bias towards a specific class. The dataset initially consisted of 4139 images representing two classes of myopia. The testing dataset was comprised of 356 images, while the training dataset contained 3783 images. However, to address the class imbalance, the dataset was balanced, resulting in a total of 2492 images across the two classes. This balanced dataset was further split into training (2200 images) and testing (292 images) subsets, as illustrated in Figure 3.
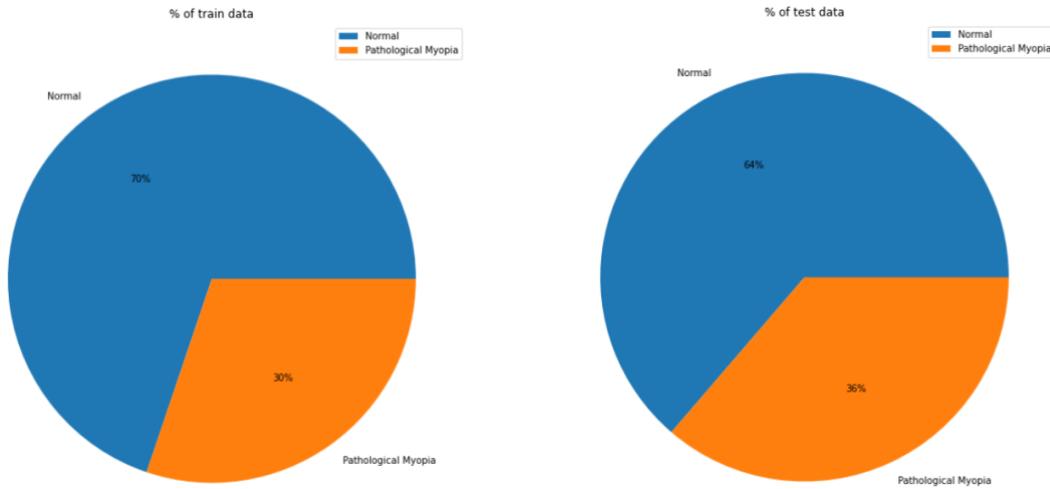
Fig. 3. Pie plot showing the distribution of test and train data

Subsequently, the images underwent pre-processing to transform them into a suitable format for the model. Since the input images were in JPEG format, they needed to be decoded into a tensor format that could be understood by the machine learning algorithm. To achieve this, the image files were read as binary strings using the TensorFlow method *tf.io.read_file()*. The JPEG-encoded images were then converted into 3-D tensors with dimensions [height, width, channels], where the channels represented the color channels (e.g., red, green, and blue) of the image, utilizing the *tf.image.decode_jpeg()* function. After obtaining the tensor representation of the images, they were resized to a standard dimension of (255, 255). This step ensured that all images had consistent dimensions, facilitating the model's learning process. Additionally, the pixel values of the images were normalized to a range of 0 to 1. Finally, the images were shuffled to eliminate any inherent structures or patterns in the data sequence. This shuffling process was crucial as it prevented the model from relying solely on specific patterns (e.g., patient ID or date) instead of the actual image attributes. In conclusion, these pre-processing procedures ensured that the input data was uniformly formatted and ready for training the deep learning model.

3.2 PHASE 2: DL Model Development

The model in this phase consists of several layers, including convolutional, pooling, flatten, and dense layers, which collectively process the input data and generate an output. Each layer has its own specific function and parameters.

- **Input Layer**: This is the initial layer of the neural network that receives the input data. In this case, the input data consists of RGB images with dimensions of 255x255 pixels. The input layer acts as a placeholder for the input data and does not have any trainable parameters.
- **Convolutional Layer**: The first convolutional layer is responsible for extracting features from the input image. It utilizes 32 filters, each with a size of 3x3. The activation function employed in this layer is the Leaky ReLU activation function (see Figure 4), defined as follows:
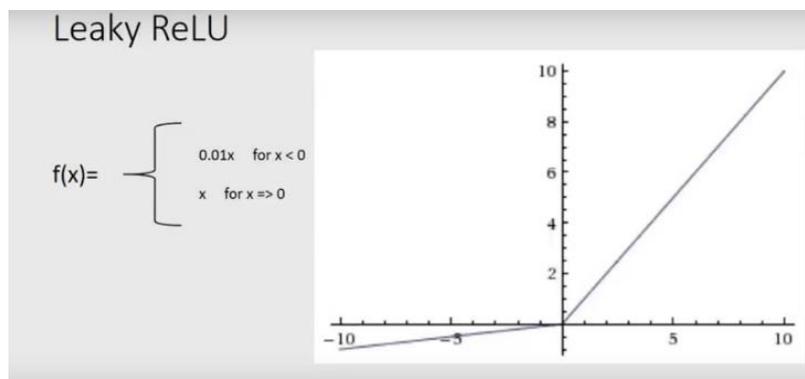
$$f(x) = max(0.01 * x, x) \qquad (1)$$

Fig. 4. Leaky ReLU activation function [34].

Where alpha is a small negative slope constant. The padding is set to 'same', which means that the output size of the layer is the same as the input size. The output of this layer is a set of 32 feature maps, each with a size of 255x255.

- **Max Pooling Layer**: The max pooling layer is employed to reduce the spatial dimensions of the feature maps. In this layer, a max pooling operation is performed on each feature map using a 2x2 filter with a stride of 2. The output of this layer is a set of 32 feature maps, each with a size of 127x127.
- **Flatten Layer**: The flatten layer is responsible for reshaping the output of the previous layer into a 1D vector. This layer does not have any trainable parameters and simply flattens the input tensor.
- **Dense Layer**: The first dense layer is utilized for feature extraction and transformation on the flattened input tensor. This layer consists of 128 neurons and uses the Leaky ReLU activation function. The output of this layer is a 1D vector with a size of 128.
- **Final Dense Layer**: The final dense layer is employed for the classification task. It contains a single neuron with a sigmoid activation function, which produces a probability value ranging from 0 to 1. Values closer to 0 indicate a non-myopic case, while values closer to 1 indicate a myopic case (see Figure 5), defined as follows:

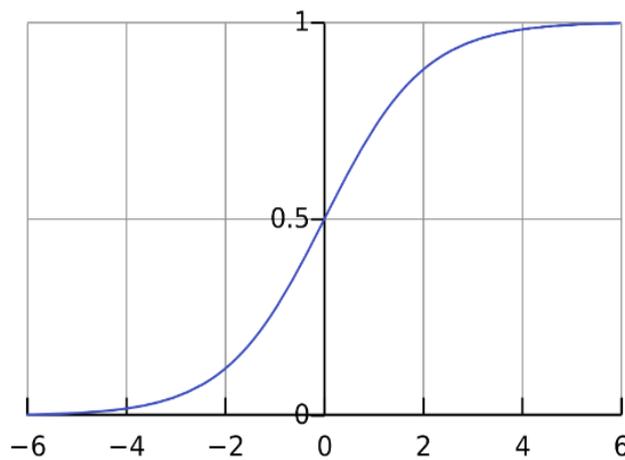$$F(x) = \frac{1}{1 + e^{-z}} \tag{2}$$

Fig. 5. Sigmoid activation function [35]

When applying transfer learning, the knowledge learned by a pre-trained model, such as VGG16, is utilized and adapted for a similar task. However, it should be noted that the VGG16 model was originally trained on a different dataset, such as ImageNet, which may not be directly applicable to the classification of retinal fundus images. Therefore, fine-tuning the model is necessary to make it more suitable for our specific data. This involves adjusting the model's parameters and architecture to better align with the characteristics of the retinal fundus images we are working with. By fine-tuning the pre-trained model, we can leverage its existing knowledge and improve its performance on our classification task. To optimize the VGG16 model for classifying retinal fundus images, the following actions are taken:

- **Step #1**: The pre-trained VGG16 model is loaded. This model has already been trained on a large dataset and has learned useful features that can be leveraged for our classification task.
- **Step #2**: Some of the layers in the VGG16 model are frozen, meaning their weights are not updated during training. This is done to preserve the low-level features, such as edges and curves that are already learned by the model. By freezing these layers, we can help the model generalize better to our specific data.
- **Step #3**: New trainable layers are added on top of the pre-trained model. These additional layers allow the model to adapt to our specific classification goal. For example, a dropout layer with a rate of 0.2 is included to prevent overfitting, and a dense layer with 256 neurons is added to capture more complex patterns in the data.
- **Step #4**: The complete model, including the pre-trained and newly added layers, is trained on our data. The pre-trained weights serve as a starting point, and the weights of the trainable layers are adjusted during training to fit our specific task. By fine-tuning the model on our data, we can benefit from the knowledge acquired by the pre-trained model and reduce the time and computational resources required for training.

By following these optimization steps, we can enhance the performance of the VGG16 model for classifying retinal fundus images and achieve better accuracy and generalization on our specific task. The next step in the process is to compile the model. After specifying the architecture of the model, it is assembled and configured for training. In this case, the optimizer used for compilation is Adam, which is a popular optimization algorithm that adapts the learning rate during training. The loss function chosen is *BinaryCrossentropy*, which is commonly used for binary classification tasks as it measures the difference between the predicted and actual values. In addition to the loss function and optimizer, various evaluation metrics are selected to assess the performance of the model. These metrics provide insights into different aspects of the model's performance. In this study, the following evaluation metrics are used:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{3}$$

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

$$F1 = 2.\frac{precision.recall}{precision + recall} \tag{5}$$

$$LogLoss = -\frac{1}{n}\sum_{i=1}^{n}[y_i \cdot log_e(\hat{y_i}) + (1 - y_i) \cdot log_e(1 - \hat{y_i})] \tag{6}$$

TP: True Positives, TN: True Negatives, FP: False Positives, FN: False Negatives

3.3 PHASE 3: XAI Model Development

LIME is a post-hoc model interpretation method used to provide explanations for individual predictions made by a black box model. In this study, LIME is employed to gain insights into the decision-making process of the deep learning model when classifying retinal fundus images as pathological myopia or normal. The goal is to generate explanations for each image in the testing set and visualize them to enhance understanding (see Figure 6).
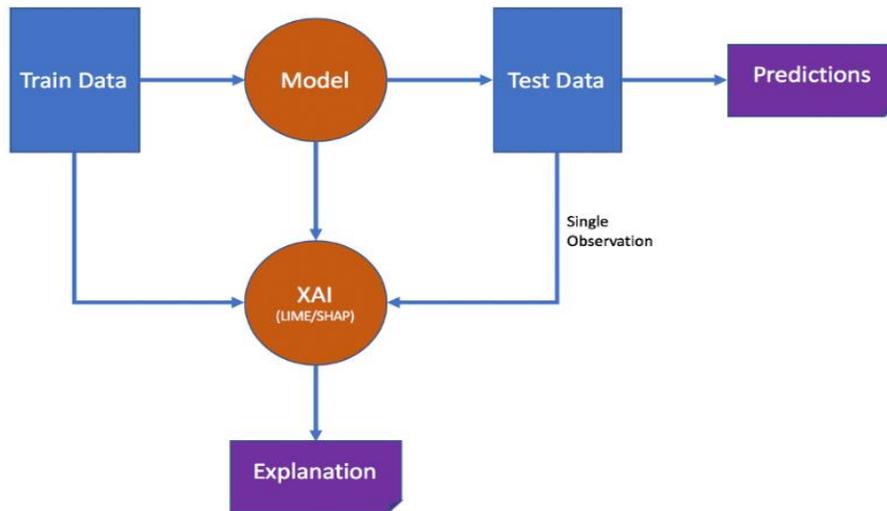

Fig. 6. LIME architecture [25]

To apply LIME, the formula (7) is utilized, where h(x') represents the local interpretable model that approximates the behaviour of the complex model (f) around a specific instance (x'). The local interpretable model is a simpler and more understandable model. The relevance of samples near x' is determined by the weighting function, pi x', assigning higher weights to examples closer to x'. This ensures that the local model focuses on the most important features near x'. The loss function (L) calculates the discrepancy between the predictions of the complex model and the predictions of the local model on weighted samples near x'.

$$h(x') = argmin(h) \; L(f, h, pi_{x'}) \tag{7}$$

The process involves using the LIME library. Firstly, an instance of the *LimeImageExplainer* class is created with default parameters. This object is used to generate an explanation for a specific image by calling *explain_instance* and providing the input image and model prediction as parameters. The *explain_instance* method returns an explanation object containing information about the most important parts of the image for the prediction.

After applying LIME to explain the model's predictions, the *explanation.get_image_and_mask()* method can be used to generate a binary mask indicating the areas of the input image that contributed the most to the predicted class. This binary mask can be visualized to highlight the relevant regions of the image. The *get_image_and_mask()* method accepts several parameters. For example, *explanation.top_labels[0]* specifies the generation of the mask for the top predicted class. The *positive_only* parameter determines whether both positive and negative regions are visible or only the positively contributing regions. The *num_features* parameter sets the number of features (regions) to include in the mask, and a smaller number can be specified to visualize fewer regions.

The *get_image_and_mask()* method returns two outputs: the mask, which is a binary mask indicating the most important image regions for the prediction, and the segments, which assigns each pixel to a unique segment. The mask can be visualized using the *mark_boundaries()* function from the *skimage.segmentation module*. The application of LIME in this project aims to provide insights into the decision-making process of the deep learning model by generating explanations and visualizing the important regions of the retinal fundus images.

3.4 PHASE 4: User Interface System
In Phase 4, the focus is on implementing the user interface for the myopia detection and XAI model. The *Gradio* Python package is utilized for this purpose, as it allows for the creation of shareable and adaptable interfaces for machine learning models. The interface is designed to have input and output components defined using the *Gradio* library. The input component is an image of the patient's eye, and the output components consist of an image and a textbox for the diagnosis prediction, as well as an image for the XAI visualization that demonstrates how the prediction was made. To achieve this, the *predict_with_xai()* function is created. This function takes an image of the patient's eye as input, performs the diagnosis prediction, and generates an XAI visualization that highlights the critical regions in the image that influenced the prediction.

 The *predict_with_xai()* function is then passed as an argument to the *Gradio* interface, enabling users to interact with the model through the interface. The *Gradio* interface is launched by using the *gr.Interface* function, which takes the input and output components along with the *predict_with_xai()* function as parameters. This allows users to upload an image of the patient's eye and receive a diagnosis prediction for myopia. Additionally, the interface provides an XAI visualization to explain the decision-making process of the model. Overall, the implementation of the user interface using the *Gradio* library enables users to easily interact with the myopia detection and XAI model, providing an intuitive way to upload images and receive diagnosis predictions accompanied by XAI visualizations for better understanding.

4. RESULTS AND DISCUSSION
This section provides an overview of the results obtained from the different phases implemented in the proposed methodology. The results are discussed in four subsections, highlighting the outcomes of each phase and their implications. Firstly, the results of the data balancing and data pre-processing are presented, showcasing how the raw images were processed and standardized in terms of size and orientation. Secondly, the results of the model development phase are presented, where two models, VGG16 and sequential CNN, were trained and evaluated. Thirdly, the results of the XAI phase are presented, where an XAI visualization was generated to explain the model's decision-making process. Finally, the developed system interface using *Gradio* is presented. These findings contribute to a better understanding of the performance and potential applications of the proposed methodology, particularly in clinical settings.

4.1 Pre-processing Result

The data balancing and pre-processing process was carried out to standardize the raw images in terms of size and orientation. This step aimed to ensure a balanced dataset and proper formatting, which would enhance the model's ability to generalize and make accurate predictions. The results demonstrate the effectiveness of the employed pre-processing techniques in preparing the dataset for model training. The data was successfully balanced, as shown in Figure 7, with an equal distribution of myopic and non-myopic cases. This equal representation of pathological myopia and normal cases in both the training and testing data is essential to avoid any bias in the model and enable accurate myopia detection.
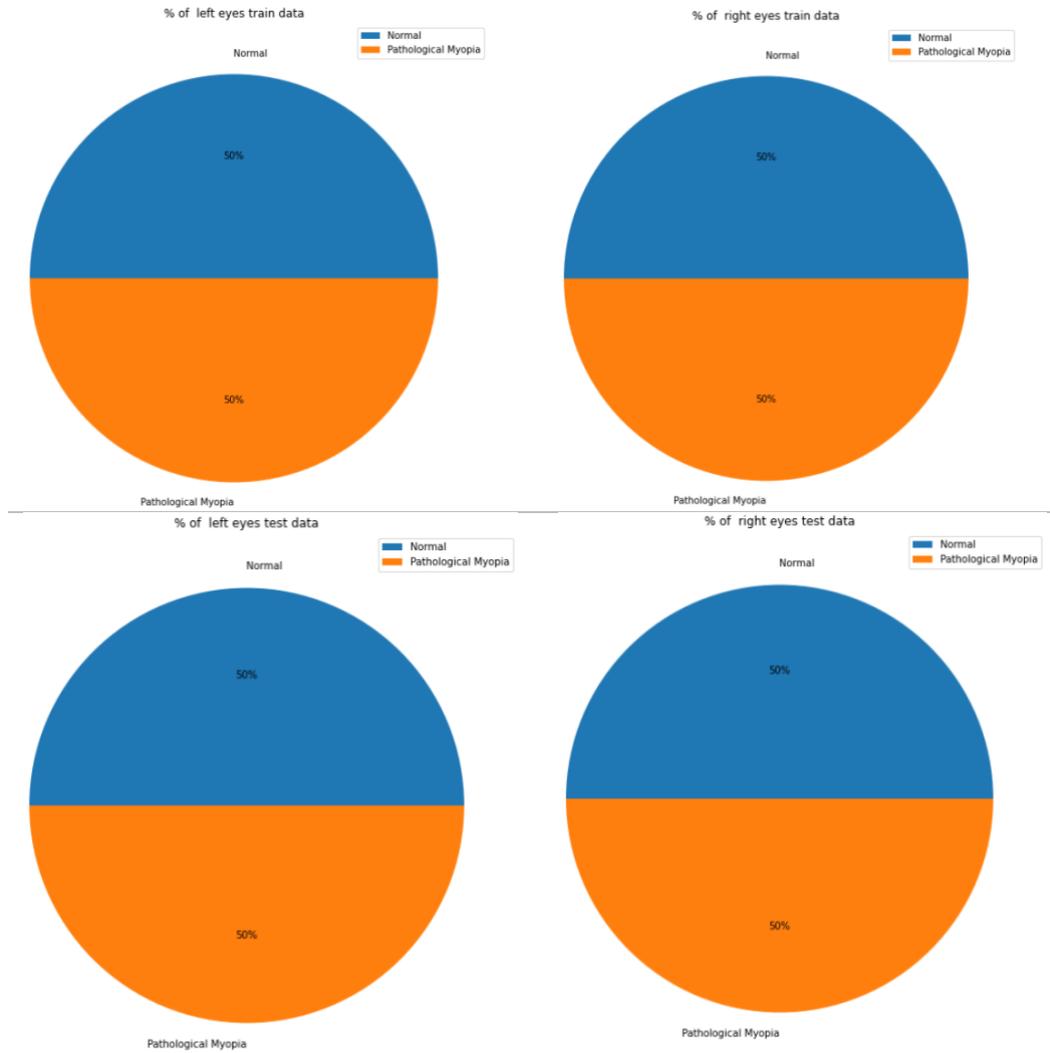
Fig. 7. Balancing results for the training and testing myopia dataset

The preprocessing step involved resizing the images to a specific size, as depicted in Figure 8, where all the images were transformed into a uniform shape. This resizing ensures that all images have the same dimensions, allowing for consistent processing by the model. Additionally, the images were normalized after resizing. Normalization involves scaling the pixel values to a narrower range, typically between 0 and 1. This normalization step helps in effective computation and improves model performance. To eliminate any bias or unwanted patterns in the dataset, the preprocessed images were randomly shuffled. This random shuffling ensures that the images are presented in a random order during training, providing the model with a diverse range of images to learn from. The plotted images in Figure 8 demonstrate the success of these preprocessing stages in achieving consistency and preparing the data for further processing by the model.
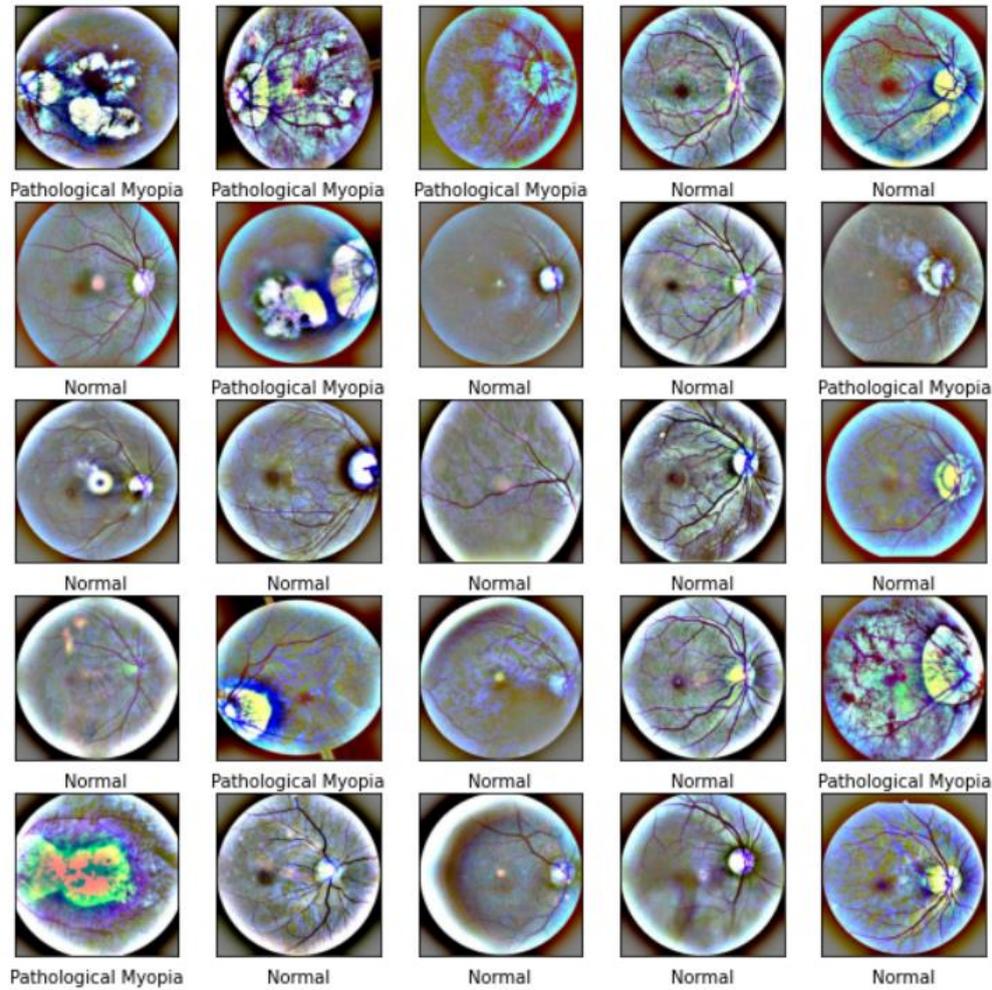
Fig. 8. Samples of pre-processing results for the myopia dataset

4.2 Model Development Result

The results in Table 1 show the performance of the trained models, CNN and VGG16. The VGG16 model exhibits higher accuracy, recall, F1 score, and sensitivity compared to the CNN model. This indicates that the VGG16 model is more effective in correctly classifying myopic and non-myopic cases.

Table I. Model development results of training

Model	Accuracy	Recall	F1 Score	Sensitivity	LogLoss
CNN	0.9521	0.9105	0.9499	0.9105	0.4581
VGG16	0.9658	0.9418	0.9683	0.9418	0.0804

The decision to continue working with the VGG16 model was based on its strong track record of success in image recognition tasks and the advantage of being a pre-trained model. By leveraging pre-trained weights, the VGG16 model reduces training time and resource consumption while improving model accuracy. It is important to note that the loss value for the VGG16 model is relatively low, indicating that the model is effective in minimizing errors. However, the trade-off between minimizing errors and correctly identifying cases should be considered when evaluating the model's performance. The results suggest that the VGG16 model is a promising choice for the myopia classification task, demonstrating high accuracy and reliability in identifying myopic and non-myopic cases.

In the context of trustworthy AI, it is crucial to analyze the model development results and assess how they align with the principles of reliability, fairness, transparency, and accountability. Evaluating the performance of the myopia detection model and understanding its implications is essential for building trust in its predictions. The results of the model development phase, as presented in Table 1, demonstrate promising outcomes for both the CNN and VGG16 models. However, further analysis is required to determine the extent to which these results contribute to trustworthy AI. To further enhance the trustworthiness of the model, it is necessary to focus on interpretability and explainability as presented in the next subsection.

4.3 XAI Model Result

The XAI phase of the methodology addresses this by generating visualizations that highlight the regions of the image that contributed to the model's decision-making process. This transparency allows clinicians and users to understand how the model reached its predictions and provides insights into the features it considers important. By offering interpretability, the model becomes more trustworthy, as its decision-making process can be understood and scrutinized. As shown in Figure 9, the XAI visualization using LIME highlights the areas in the eye image that are most important in the model's decision-making process. This visualization plays a crucial role in enhancing the trustworthiness of the myopia detection model. By providing transparency and explainability, it allows clinicians and users to understand the basis of the model's predictions.
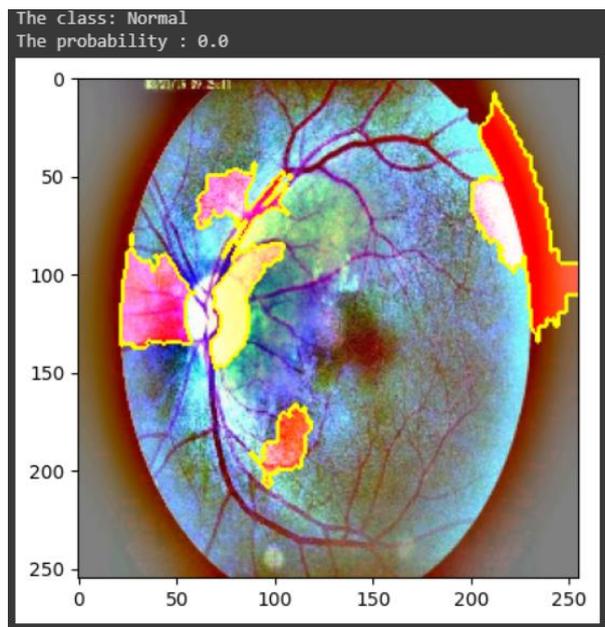

Fig. 9. A sample of LIME visualization for normal myopia

The highlighted areas in the image offer valuable insights into how the model arrived at its diagnosis. They indicate which regions or features of the eye image were deemed significant by the model in determining whether a patient has myopia or not. This level of interpretability is essential for building trust in the model's decisions. Moreover, the XAI visualization enables the identification of potential biases or inaccuracies in the model's predictions. By examining the highlighted areas, clinicians and researchers can detect any patterns or trends that may indicate biased decision-making. This information can guide further investigations and refinements to ensure the model's fairness and accuracy.

4.4. System Development Result

The development of a user interface using *Gradio* adds to the trustworthiness of the model. The interface presented in Figure 10 and Figure 11 provide a platform for users to interact with the model, view the diagnosis prediction, and access the XAI visualization. This transparency and user engagement foster trust, as users can observe the model's outputs and understand the reasoning behind its predictions.
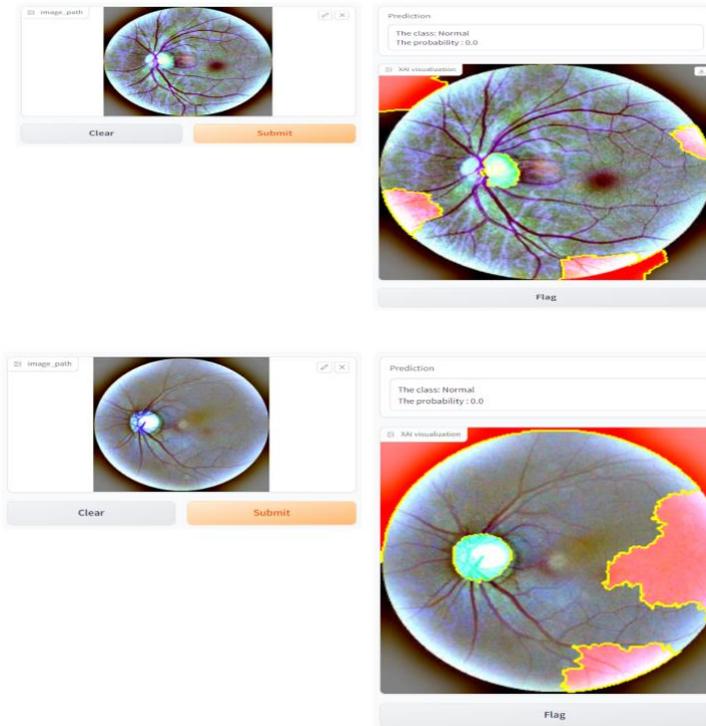
Fig. 10. User interface for myopia detection with XAI model (two random normal cases)
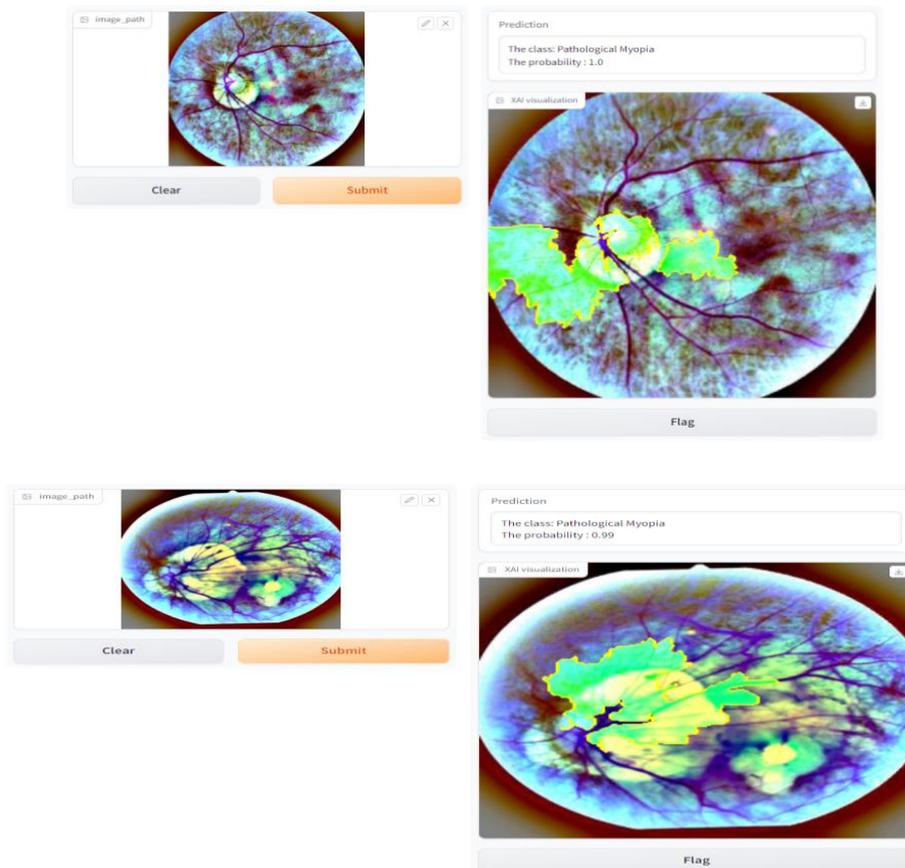

Fig. 11. User interface for myopia detection with XAI model (two random myopia cases)

The user interface, developed using *Gradio*, provided an intuitive and accessible platform for interacting with the model. Users could easily upload eye images and obtain quick and reliable diagnosis results, accompanied by an explanation of the model's prediction. This user-friendly interface promoted effective communication between healthcare providers and patients, fostering trust and understanding. The inclusion of XAI in the model development process offered several advantages. It enabled healthcare professionals to explain to patients the reasoning behind the model's predictions, enhancing transparency and promoting trust. Additionally, XAI facilitated the identification of potential biases or limitations in the model, allowing for continuous improvement and ensuring reliable and trustworthy results. The study demonstrated the efficacy of the developed deep learning model, the importance of XAI in enhancing interpretability, and the value of a user-friendly interface in facilitating the adoption of AI in healthcare settings. These findings contribute to the growing body of research on trustworthy AI and its potential applications in clinical decision-making.

5. CONCLUSION

This study successfully developed a deep learning model, XAI visualization, and a knowledge-based system for the detection of myopia. The model demonstrated accurate predictions and exhibited promising results in effectively detecting cases of myopia. The integration of XAI techniques provided transparency and interpretability, enabling healthcare professionals to gain insights into the model's decision-making process. Both the CNN and VGG16 models were trained using the dataset, and while both models performed well, the VGG16 model was selected for further analysis due to its strong performance in image recognition tasks and its pre-trained nature. Impressively, the VGG16 model achieved an accuracy of 96%, underscoring its effectiveness in diagnosing myopia. The XAI visualization, developed using LIME, proved to be a valuable tool for interpreting and explaining the model's predictions. It significantly enhanced the model's transparency by providing valuable insights into the features and regions of the eye images that influenced the diagnosis. Moving forward, the study aimed to extend the developed deep learning model by incorporating XAI techniques more comprehensively. This extension would further enhance the interpretability and trustworthiness of the model, making it more suitable for real-world applications. Additionally, efforts could be made to refine and optimize the user interface of the system, ensuring its user-friendliness and accessibility for healthcare professionals and patients. In conclusion, this study makes a significant contribution to the field of myopia detection by developing a reliable deep learning model, integrating XAI techniques, and creating a user-friendly system. The findings pave the way for future research and advancements in the application of trustworthy AI in clinical settings.

References

[1] I. G. Morgan, K. Ohno-Matsui, and S. M. Saw, "Myopia," Lancet, vol. 379, no. 9827, pp. 1739–1748, May 2012, doi: 10.1016/S0140-6736(12)60272-4.

[2] D. Caltrider, A. Gupta, and K. Tripathy, Evaluation Of Visual Acuity. University of Texas Medical Branch School of Medicine: StatPearls Publishing, Treasure Island (FL), 2021.

[3] J. Rozema, S. Dankert, and R. Iribarren, "Emmetropization and nonmyopic eye growth," Surv. Ophthalmol., vol. 68, no. 4, pp. 759–783, Jul. 2023, doi: 10.1016/j.survophthal.2023.02.002.

[4] W. A. Lagrèze and F. Schaeffel, "Preventing Myopia," Dtsch. Arztebl. Int., vol. 114, no. 35–36, pp. 575–580, Sep. 2017, doi: 10.3238/arztebl.2017.0575.

[5] D. A. Levinthal and J. G. March, "The myopia of learning," Strateg. Manag. J., vol. 14, no. S2, pp. 95–112, 1993, doi: 10.1002/smj.4250141009.

[6] M. Moradi and M. Dass, "Applications of artificial intelligence in B2B marketing: Challenges and future directions," Ind. Mark. Manag., vol. 107, pp. 300–314, 2022, doi: 10.1016/j.indmarman.2022.10.016.

[7] P. K. Verkicharla, P. Kammari, and A. V. Das, "Myopia progression varies with age and severity of myopia," PLoS One, vol. 15, no. 11, p. e0241759, Nov. 2020, doi: 10.1371/journal.pone.0241759.

[8]   K. Aggarwal et al., "Has the Future Started? The Current Growth of Artificial Intelligence, Machine Learning, and Deep Learning," Iraqi J. Comput. Sci. Math., vol. 3, no. 1, pp. 115–123, Jan. 2022, doi: 10.52866/ijcsm.2022.01.01.013.

[9]   A. Bender, N. Schneider, M. Segler, W. Patrick Walters, O. Engkvist, and T. Rodrigues, "Evaluation guidelines for machine learning tools in the chemical sciences," Nat. Rev. Chem., vol. 6, no. 6, pp. 428–442, 2022, doi: 10.1038/s41570-022-00391-9.

[10] A. S. Albahri et al., "A Systematic Review of Using Deep Learning Technology in the Steady-State Visually Evoked Potential-Based Brain-Computer Interface Applications: Current Trends and Future Trust Methodology," Int. J. Telemed. Appl., vol. 2023, 2023.

[11] A. S. Albahri et al., "Towards physician's experience: Development of machine learning model for the diagnosis of autism spectrum disorders based on complex <scp>T</scp> -spherical fuzzy-weighted zero-inconsistency method," Comput. Intell., vol. 39, no. 2, pp. 225–257, Apr. 2023, doi: 10.1111/coin.12562.

[12] L. Alzubaidi et al., "A survey on deep learning tools dealing with data scarcity: definitions, challenges, solutions, tips, and applications," J. Big Data, vol. 10, no. 1, p. 46, Apr. 2023, doi: 10.1186/s40537-023-00727-2.

[13] A. M. Obeso, J. Benois-Pineau, M. S. García Vázquez, and A. Á. R. Acosta, "Visual vs internal attention mechanisms in deep neural networks for image classification and object detection," Pattern Recognit., vol. 123, p. 108411, 2022, doi: 10.1016/j.patcog.2021.108411.

[14] M. E. Alqaysi, A. S. Albahri, and R. A. Hamid, "Hybrid Diagnosis Models for Autism Patients Based on Medical and Sociodemographic Features Using Machine Learning and Multicriteria Decision-Making (MCDM) Techniques: An Evaluation and Benchmarking Framework," Comput. Math. Methods Med., vol. 2022, p. 9410222, 2022, doi: 10.1155/2022/9410222.

[15] M. E. Alqaysi, A. S. Albahri, and R. A. Hamid, "Diagnosis-Based Hybridization of Multimedical Tests and Sociodemographic Characteristics of Autism Spectrum Disorder Using Artificial Intelligence and Machine Learning Techniques: A Systematic Review," Int. J. Telemed. Appl., vol. 2022, 2022, doi: 10.1155/2022/3551528.

[16] Q. Zhang, M. Zhang, T. Chen, Z. Sun, Y. Ma, and B. Yu, "Recent advances in convolutional neural network acceleration," Neurocomputing, vol. 323, pp. 37–51, 2019, doi: 10.1016/j.neucom.2018.09.038.

[17] S. S. S. Palakodati, V. R. Chirra, Y. Dasari, and S. Bulla, "Fresh and Rotten Fruits Classification Using CNN and Transfer Learning," Rev. d'Intelligence Artif., vol. 34, no. 5, pp. 617–622, Nov. 2020, doi: 10.18280/ria.340512.

[18] S. Tammina, "Transfer learning using VGG-16 with Deep Convolutional Neural Network for Classifying Images," Int. J. Sci. Res. Publ., vol. 9, no. 10, p. p9420, Oct. 2019, doi: 10.29322/IJSRP.9.10.2019.p9420.

[19] S. S. Joudar, A. S. Albahri, and R. A. Hamid, "Triage and priority-based healthcare diagnosis using artificial intelligence for autism spectrum disorder and gene contribution: A systematic review," Comput. Biol. Med., vol. 146, p. 105553, Jul. 2022, doi: 10.1016/j.compbiomed.2022.105553.

[20] T. N. Abu-, J. Samy, and S. Abu-Naser, "Classification of Sign-language Using VGG16," Int. J. Acad. Eng. Res., vol. 6, no. 6, pp. 36–46, 2022.

[21] J. Zhang and H. Zou, "Artificial intelligence technology for myopia challenges: A review," Front. Cell Dev. Biol., vol. 11, 2023, doi: 10.3389/fcell.2023.1124005.

[22] S. S. Joudar et al., "Artificial intelligence-based approaches for improving the diagnosis, triage, and prioritization of autism spectrum disorder: a systematic review of current trends and open issues," Artif. Intell. Rev., pp. 1–65, Jun. 2023, doi: 10.1007/s10462-023-10536-x.

[23] M. A. Ahmed et al., "Intelligent Decision-Making Framework for Evaluating and Benchmarking Hybridized Multi-Deep Transfer Learning Models: Managing COVID-19 and Beyond," Int. J. Inf. Technol. Decis. Mak., Apr. 2023, doi: 10.1142/S0219622023500463.

[24] A. S. Albahri et al., "A Systematic Review of Trustworthy and Explainable Artificial Intelligence in Healthcare: Assessment of Quality, Bias Risk, and Data Fusion," Inf. Fusion, Mar. 2023, doi: 10.1016/j.inffus.2023.03.008.

[25] A. Ignatiev, "Towards Trustable Explainable AI," in Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, Jul. 2020, vol. 2021-Janua, pp. 5154–5158. doi: 10.24963/ijcai.2020/726.

[26] S. Mishra, B. L. Sturm, and S. Dixon, "Local interpretable model-agnostic explanations for music content analysis," in Proceedings of the 18th International Society for Music Information Retrieval Conference, ISMIR 2017, 2017, vol. 53, pp. 537–543.

[27] I. P. de Sousa, M. M. B. R. Vellasco, and E. C. da Silva, "Local interpretable model-agnostic explanations for classification of lymph node metastases," Sensors (Switzerland), vol. 19, no. 13, p. 2969, 2019, doi: 10.3390/s19132969.

[28] Y. Yang et al., "Automatic identification of myopia based on ocular appearance images using deep learning," Ann. Transl. Med., vol. 8, no. 11, pp. 705–705, Jun. 2020, doi: 10.21037/atm.2019.12.39.

[29] N. Rauf, S. O. Gilani, and A. Waris, "Automatic detection of pathological myopia using machine learning," Sci. Rep., vol. 11, no. 1, p. 16570, 2021, doi: 10.1038/s41598-021-95205-1.

[30] R. Hemelings, B. Elen, M. B. Blaschko, J. Jacob, I. Stalmans, and P. De Boever, "Pathological myopia classification with simultaneous lesion segmentation using deep learning," Comput. Methods Programs Biomed., vol. 199, p. 105920, 2021, doi: 10.1016/j.cmpb.2020.105920.

[31] L. Lu et al., "Development of deep learning-based detecting systems for pathologic myopia using retinal fundus images," Commun. Biol., vol. 4, no. 1, p. 1225, Oct. 2021, doi: 10.1038/s42003-021-02758-y.

[32] S. S. Joudar, A. S. Albahri, and R. A. Hamid, "Intelligent triage method for early diagnosis autism spectrum disorder (ASD) based on integrated fuzzy multi-criteria decision-making methods," Informatics Med. Unlocked, vol. 36, p. 101131, 2023, doi: 10.1016/j.imu.2022.101131.

[33] Retinal Fundus Images, "https://www.kaggle.com/," 2023. https://www.kaggle.com/datasets/kssanjaynithish03/retinal-fundus-images (accessed Feb. 01, 2023).

[34] O. El Mellouki, M. I. Khedher, and M. A. El-Yacoubi, "Abstract Layer for LeakyReLU for Neural Network Verification based on Abstract Interpretation," IEEE Access, vol. 11, pp. 33401–33413, 2023, doi: 10.1109/ACCESS.2023.3263145.

[35] U. A. Ezeafulukwe, M. Darus, and O. A. Fadipe-Joseph, "On analytic properties of a sigmoid function," Int. J. Math. Comput. Sci., vol. 13, no. 2, pp. 171–178, 2018.