Journal of Neonatal Surgery
ISSN(Online): 2226-0439
Vol. 14, Issue 4s (2025)
https://www.jneonatalsurg.com

A Novel approach to multi-level image classification of Retinal Diseases

Patel Vandanabahen Gopalbhai[1], Dr. Jayesh N. Modi[2]

[1]Research Scholar, Department of Computer Science, Hemchandracharya North Gujarat University, Patan.

Emial ID: vandanapatel1011@gmail.com

[2]PHD, Computer science, Department of Computer Science, Hemchandracharya North Gujarat University, Patan.

Email ID: dr.jayeshmodi@gmail.com

Cite this paper as: Patel Vandanabahen Gopalbhai, Dr. Jayesh N. Modi, (2025) A Novel approach to multi-level image classification of Retinal Diseases. *Journal of Neonatal Surgery*, 14 (4s), 1045-1057.

ABSTRACT

Diseases of the retina, like diabetic retinopathy, age-related macular degeneration, and glaucoma, are major causes of vision loss and blindness around the world. Early identification and labelling of these illnesses are necessary to act quickly and effectively treat them. Although conventional image classification techniques are useful, they may locate minute features in retinal images poorly, which results in the incorrect diagnosis. This work proposes a novel multi-level image classification system for eye disorders' detection and categorisation. It makes the diagnosis more accurate by use of sophisticated deep learning approaches. Shallow and deep feature extraction layers of a multi-level convolutional neural network (CNN) structure define our desired approach. This combined design allows one to present complex objects at upper levels and capture minute details at lower ones. Along with comments outlining the nature and stage of every illness, the model is taught on a vast library of retinal photographs including both healthy and diseased images. The model is made more steady using several pre-processing techniques like image normalisation, augmentation, and noise reduction. We evaluate our proposed approach against present deep learning models and standard machine learning techniques. to approach performs much better, according to data, in raising classification accuracy, precision, recall, and F1-score. When it comes to distinguish between many phases of eye illnesses, our multi-level approach outperforms conventional models. This makes it a more exact and comprehensive testing instrument. We also discuss whether the model can be used in real-life healthcare environments and in other datasets. Though it has certain issues—an uneven collecting, difficult-to-understand models, and the need for a lot of computational power—the proposed technique shows potential.

Keywords: *Retinal Diseases, Multi-Level Image Classification, Deep Learning, Convolutional Neural Networks, Medical Image Analysis*

1. INTRODUCTION

Retinal diseases are one of the main reasons people go blind or have trouble seeing around the world. They are more common in older people and people who have long-term conditions like diabetes. In recent years, eye diseases like diabetic retinopathy, age-related macular degeneration (AMD), and glaucoma have become more common. This is because of changes in population and more people being at risk. In their early stages, these diseases often don't cause any symptoms, so they need to be found quickly and correctly classified to keep people from losing their sight. In this situation, improvements in medical images and machine learning could completely change how eye diseases are found and tracked. It is really hard to tell the difference between eye diseases based on medical pictures like fundus images. In the early stages of retinal diseases, the retina may show small changes that are hard to see with the naked eye. Conventional diagnosis methods are also very hard to use because eye structures are very complicated, picture quality can vary, and disease symptoms can be very different from one patient to the next. Usually, ophthalmologists have to analyse images by hand, which takes a lot of time and can lead to mistakes, especially when there are a lot of images to look at. As a result, there is a greater need for computer testing tools that can help doctors make correct findings quickly. Particularly convolutional neural networks (CNNs), fresh advancements in machine learning and deep learning have showed great potential in the field of medical image analysis. CNNs have proved they can precisely group complicated patterns and automatically learn characteristics from unprocessed visual data. Many fields of medical imaging, including breast cancer, lung disorders, and brain difficulties [1], have seen effective use of these models. CNN-based approaches have demonstrated remarkable performance in the classification of eye illnesses; deep learning models are more accurate and quicker than conventional techniques. Classification of eye illnesses still presents several challenges even with these developments, however. Current deep learning algorithms may not always detect the minute, multi-scale features seen in retinal images. This makes early on illness diagnosis more difficult.

Usually, these models also perform only in binary or restricted multi-class categorisation. This implies they struggle to distinguish between many phases of the same illness or identify unusual eye diseases. Another major concern is that these models won't fit various datasets or in actual clinical settings [2]. Most of the present techniques have a major flaw in that they only see one level of an image, which may not be able to pick up the many layers of information required to appropriately classify eye disorders at various periods. We propose a fresh approach to categorise photos at many levels for eye illnesses in order to circumvent these issues. This approach employs a hybrid convolutional neural network (CNN) architecture and operates at many degrees of feature extraction. Our approach generates both specific local features and more general, less concrete patterns in retinal pictures by combining shallow and deep convolutional layers both together. The arrangement of the model is meant to be optimum with many various photo sizes. This helps it to better categorise illnesses in all their many forms and stages. The proposed multi-level approach guarantees consideration of both minor alterations and more prominent indicators of illness. This guarantees higher accuracy and dependability in the assessment findings. In our approach, we use innovative image preparation techniques like image normalisation, enhancement, and noise reduction to improve the raw images and therefore the model performs better. Typical in eye imaging are images with plenty of change and noise [3]. These techniques assist to solve these issues. We make sure that the system works well with a wide range of patients and imaging situations by training it on a collection that includes both healthy and sick retinal images. We also test the model's performance using several evaluation measures, such as F1-score, accuracy, precision, and memory, to get a full picture of how well it can diagnose problems.

2. LITERATURE REVIEW

A. Traditional Approaches to Retinal Disease Classification

Traditional ways of classifying retinal diseases relied on doctors, like ophthalmologists or optometrists, looking at pictures of the retina, especially fundus photographs, and interpreting them by hand to find signs of disease. Visual clues, like microaneurysms, exudates, or haemorrhages, were often used in these ways to figure out the stage and seriousness of diseases like diabetic retinopathy or age-related macular degeneration (AMD). However, this method has a lot of problems because it depends a lot on the examiner's knowledge and can be flawed by human mistake, especially in cases that are complicated or unclear. Small changes in the retina may not be easy to see with the naked eye in the early stages of retinal disease. Because of this, hand analysis might miss important data or cause wrong classification, especially when there are a lot of files with different picture quality [4]. Because eye analysis is biassed, different workers may come up with different findings, which can make care less consistent. Automated systems started to appear that used basic image processing techniques, like edge recognition and thresholding, to point out possible areas of concern. Figure 1 shows the old ways of classifying eye diseases that involve looking at the retina by hand and using simple imaging methods. This made the process of classifying eye diseases more efficient and consistent.

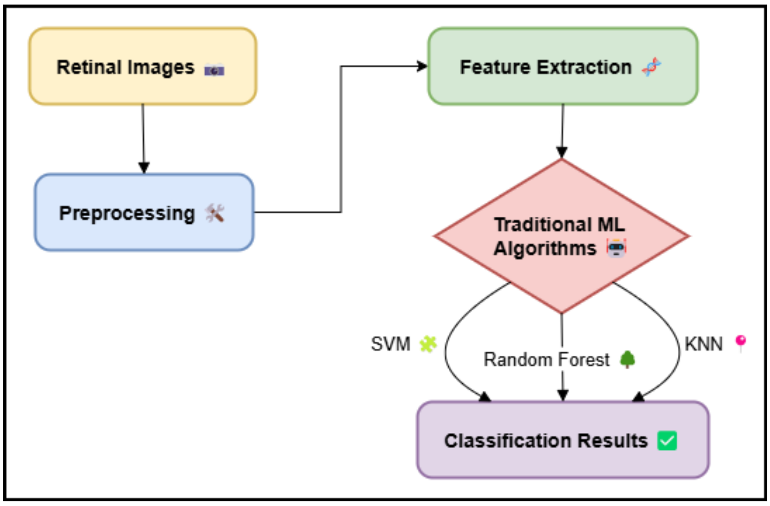

Figure 1: Illustrating Traditional Approaches to Retinal Disease Classification

But these methods still had trouble with things like noise, low sharpness, and the fact that retinal pictures are very complicated. Traditional ways were also limited in how big they could be because looking at a lot of pictures by hand takes a lot of time and resources.

B. Machine Learning in Medical Image Analysis

Machine learning (ML) has changed the way medical picture analysis is done by providing automated solutions that can help doctors make better choices, make diagnoses more accurately, and reduce mistakes. ML algorithms, especially deep learning techniques, have shown to be much better than old ways of handling pictures. Medical pictures may therefore be examined on a bigger scale and more precisely. Machine learning has great promise to categorise retinal illnesses by extracting features from retinal images that are difficult for the human eye to detect [5]. Among the most significant advances in machine learning for medical imaging is the development of convolutional neural networks (CNNs). Many automated photo segmentation systems nowadays rely on CNNs as their fundamental component. CNNs are ideal for photo analysis as they can quickly learn hierarchical features from raw image data, therefore saving you from manual work. Sorting many other eye illnesses, including AMD, glaucoma, and diabetic retinopathy, this capacity to memorise complex patterns has proved very beneficial. Furthermore fast and consistently handling vast volumes of data, machine learning models are very helpful in healthcare environments where large datasets are somewhat typical. By teaching algorithms to handle variations in image quality, patient data, and illness stages, you may increase their dependability and use over a larger spectrum of populations [6].

C. Previous Work on Retinal Disease Image Classification

Many research have looked at how machine learning and deep learning may be used to categorise photos of eye illnesses in the last few years. Early studies sought locations where sickness could be present using typical image processing methods like thresholding and edge identification. These techniques, however, struggled with complex patterns in retinal images and often required a lot of human labour to interpret what they meant. Many researchers began grouping eye diseases using convolutional neural networks (CNNs) as deep learning gained increasing popularity [7]. Deep learning might be used to identify diabetic retinopathy, sometimes better than conventional techniques and even human ophthalmologists, according to a significant 2016 Gulshan et al. research. This work was significant as it demonstrated that deep learning models could not only streamline the process of identifying eye disorders but also very precisely identify early indicators of sickness, which is rather crucial to prevent individuals from losing their sight. Later studies have shown that deep learning models can be used to help with more eye diseases, like glaucoma and age-related macular degeneration. Table 1 summarizes literature review on algorithms, key findings, challenges, and their impact. CNN designs have been created by researchers that can tell the difference between different steps of a disease and find rare conditions that doctors might not be able to easily spot.

Table 1: Summary of Literature Review

Algorithm	Key Finding	Challenges	Impact
Traditional CNN	Achieves reasonable accuracy in binary classification	Limited feature extraction capability for complex diseases	Used widely in early retinal disease classification tasks
ResNet50	Deep architecture improves accuracy on large datasets	Requires large datasets, difficult to train without high-quality data	Improved accuracy for complex multi-class disease detection
VGG16 [8]	High accuracy in small datasets but struggles with generalization	Overfitting on small datasets, poor generalization	Works well with simple datasets, but struggles in practice
U-Net	Effective for segmentation tasks in retinal images	Not suitable for multi-class classification in all cases	Advanced segmentation capabilities, used in clinical tools
InceptionV3	Better at handling multi-class classification	Sensitive to data quality, needs robust augmentation	Improved classification of multiple retinal conditions
DenseNet	High accuracy with fewer parameters, good for edge devices	May still struggle with small datasets and rare diseases	A good choice for edge computing in portable diagnostic devices
SqueezeNet	Compact model with a small footprint but reasonable accuracy	Accuracy drops with complex datasets	Used in applications with limited resources such as mobile devices
Faster R-CNN [9]	Useful for object detection in	Requires annotated data for	Applicable in early detection and diagnosis with bounding

	retinal images	bounding box detection	box tools
Xception	Excellent performance with transfer learning	Not optimized for real-time clinical settings	Widely adopted for fine-tuned models with high accuracy
MobileNet	Lightweight, suitable for real-time applications	Not suitable for devices with limited computational resources	Optimal for real-time diagnostic devices in clinics
EfficientNet	Efficient model with low computational cost	Can become too large, requiring high computational resources	Widespread use in large-scale datasets for research and hospitals
AlexNet [10]	Outperforms traditional CNNs on large datasets	Requires large datasets and computational power	Highly regarded in research, particularly for large datasets
VGG19	Good accuracy but high computational cost	Slow training and high memory usage	Powerful architecture, mainly in research but challenging for production

3. DATASET: RETINAL DISEASE CLASSIFICATION

The quality and variety of the information used for training and testing a machine learning or deep learning model have a big impact on how well it works. When it comes to classifying retinal diseases, the dataset is very important because it shows how well the model can find different stages and kinds of retinal illnesses. We use a freely available retinal disease dataset for this work. It is made up of labelled fundus images, which are commonly used to diagnose retinal illnesses. The collection used in this study has retinal pictures of both healthy and sick eyes. It includes pictures of people with diabetic retinopathy, age-related macular degeneration (AMD), glaucoma, and macular oedema [11]. Digital fundus cameras, which take high-resolution colour pictures of the retina, are used to take these pictures. There are marks on each picture that show which eye illnesses are present and how bad they are. The trained ophthalmologists who made these notes make sure that the data is of good quality and can be trusted for training models. One of the biggest problems with classifying eye diseases is that the information has a lot of classes that aren't balanced. It's more likely that common conditions like diabetic retinopathy will be shown, while less common conditions like glaucoma or eye separation may not be shown enough [12]. Data enhancement methods, such as random rotations, translations, flips, and colour changes, are used to fix this mismatch. It not only gives you more pictures to learn with, but it also makes the model better at generalisation by showing it different image situations. In order to properly test the model, the dataset is usually split into groups for training, validation, and testing. The validation set helps fine-tune the model's hyperparameters and stops it from overfitting. The training set teaches the model how to find patterns in the retinal pictures. The training set, which has data that hasn't been seen before, is used to test how well the model can generalise and make sure it can correctly describe eye diseases in real life. Image preparation methods are used to make the model work even better. Among them include normalising images, enhancing brightness, and noise reduction. These help to minimise the effect of visual artefacts and guarantee stable quality of the supplied pictures. This preprocessing stage is very crucial as retinal images may be changed by illumination, focus, and camera settings.

4. METHODOLOGY

A. Convolutional Neural Networks (CNN)

Medical picture analysis is one use of convolutional neural networks (CNNs), a kind of deep learning model demonstrated to be quite successful in image classification. Multiple layers inside CNNs cooperate to automatically learn spatial structures of features from unprocessed visual input. These layers are completely connected, convolution, activation, pooling, etc. These models take use of the local spatial structure within images, so they are excellent for image identification challenges [13]. This makes them ideal for examining retinal pictures including many distinct patterns and structures. Convolutional layers that alter the source image with various filters form the first few layers of a standard CNN. Among the low-level elements these filters detect are edges, textures, and colour patterns. Deeper within the network, the filters grow more complex. They take up more abstract representations, higher-level patterns like forms and buildings. Pooling layers cut down on the image's spatial dimensions. This lets the model focus on the most important traits and makes the computations easier. Finally, fully connected layers take these extracted traits and use them to guess what the picture is like [14]. A lot of pictures of the retina are used to train CNNs to recognise different types of eye diseases.

- Step 1: Convolution Layer

The first step in a CNN is applying convolution to the input image using a set of learnable filters (kernels). The convolution operation helps in detecting spatial hierarchies of features such as edges, textures, and shapes.

Mathematical Equation:

For an input image I and a filter F, the convolution operation Y can be represented as:

$$Y(i,j) = (I * F)(i,j) = \sum m \sum n \, I(m,n) \cdot F(i - m, j - n)$$

Where * denotes the convolution operation, and (i, j) represents the position of the filter on the image. The filter is slid across the image, and at each location, it computes the dot product of the image's pixels and the filter's values.

- Step 2: Activation Function (ReLU)

After applying convolution, the next step is to introduce non-linearity using an activation function, typically the Rectified Linear Unit (ReLU). ReLU introduces non-linear behavior in the network and helps with the learning of complex patterns.

Mathematical Equation:

$$f(x) = \max(0, x)$$

Here, x represents the output of the convolutional layer, and ReLU replaces all negative values with 0 while retaining positive values.

- Step 3: Pooling Layer

The pooling layer reduces the spatial dimensions (width and height) of the feature map while retaining the important features. Max pooling is commonly used, where the maximum value is selected from a region of the feature map.

Mathematical Equation:

For max pooling, the output Y is calculated as:

$$Y(i,j) = \max_{\{m,n\}I}(i + m, j + n)$$

Here, m and n represent the dimensions of the pooling window (e.g., 2x2 or 3x3), and the maximum value from each region is selected as the output.

- Step 4: Fully Connected Layer (Classification)

After the feature maps are pooled and flattened into a vector, the fully connected layer connects every node to every other node. This layer performs the classification based on the extracted features.

Mathematical Equation:

For a fully connected layer, the output is given by:

$$Y = W \cdot X + b$$

Where:

- X is the flattened input from the previous layer,

- W is the weight matrix,

- b is the bias term, and

- Y is the output vector used for classification (e.g., predicting whether the image contains a specific class like "diabetic retinopathy" or "healthy").

This step produces the final output, which is passed through a softmax function to generate the probability distribution over different classes. The softmax function is:

$$P(y_i) = \frac{e^{y_i}}{\sum j \, e^{y_j}}$$

Where y_i is the score for class i, and P(y_i) is the probability for that class. The class with the highest probability is the model's predicted class.

B. Hierarchical Attention-Based CNN Model

Hierarchical Attention-Based CNN (HAC-CNN) is a novel kind of architecture that integrates attention processes with normal CNNs to improve the model in concentrating on the most significant areas of an image for categorisation activities. This model aids the network in identifying and prioritising the areas of the retina displaying significant illness symptoms, such as sores, blood vessel abnormalities, or alterations in the optic disc, therefore guiding the classification of eye disorders. The network employs a hierarchical focus method on many levels [15]. At the lowest levels, focus is on fine-grained, crucial features such particular retinal tumours or haemorrhage. Higher degrees of awareness pay individuals more attention to more significant patterns, such as overall structural changes of blood vessels or if retinal oedema exists. This hierarchical approach helps the model to properly detect many phases of eye illnesses by letting it concentrate on both local and global aspects. Different areas of the image are given varying weight, allowing the model to disregard the less significant portions of the image and concentrate its processing capability on the most vital areas. Ultimately, this results in improved model performance and quicker learning—especially in cases of complex images including both healthy and sick parts [16]. The tiered nature of the attention mechanism also allows the model to detect both microscopic details and bigger context, which is crucial for seeing minute changes in the retina that would indicate early stages of illness.

- Step 1: Feature Extraction (Convolution and Attention)

HAC-CNN starts with a convolution operation to extract feature maps and then uses an attention mechanism to assist concentrate on significant regions within the produced features. This helps the model to evaluate the significance of certain areas of the picture before starting more investigation.

Mathematical Equation (Convolution):

The convolution operation on an input image I with a filter F is given by:

$$Y(i,j) = (I * F)(i,j) = \sum m \sum n\, I(m,n) \cdot F(i-m, j-n)$$

Where * denotes the convolution operation, and (i, j) represents the location of the filter on the image.

Attention Mechanism:

The attention mechanism calculates an attention score A for each feature map to assign weights to the important regions. This is typically done using an attention vector α that is learned during training.

$$A(i,j) = \frac{e^{W_{att} \cdot X(i,j)}}{\sum_{\{i,j\}e}^{W_{att} \cdot X}(i,j)}$$

Where X(i, j) represents the feature at position (i, j) and W_att is the weight associated with the attention mechanism.

- Step 2: Hierarchical Attention Mechanism

From low-level characteristics to high-level representations, the hierarchical attention mechanism attends at many tiers of the model. The aim is to concentrate not only on the relevant areas within the picture but also on significant areas at many levels of the feature maps.

Mathematical Equation:

Using both local and global attention scores—where the local attention is calculated from fine-grained characteristics and the global attention focusses on more expansive areas—the hierarchical attention method The last attention score combines weight between both:

$$A_{final} = \lambda A_{local} + (1 - \lambda) A_{global}$$

Where A_local and A_global are the attention scores at local and global levels, and λ is a hyperparameter that balances the importance of each.

- Step 3: Weighted Feature Map Aggregation

Following attention, the weighted feature maps are combined to provide the ultimate representation. Summing the feature maps is part of this aggregation; the attention scores provide the features at strategic points more weight.

Mathematical Equation:

The aggregated feature map Y_agg is calculated by:

$$Y_{agg} = \sum_{\{i,j\}A} (i,j) \cdot Y(i,j)$$

Where A(i, j) is the attention weight for each feature Y(i, j) in the feature map.

- Step 4: Fully Connected Layer and Classification

Following flattening, the aggregated feature map is sent through a fully connected layer for classification. Following a softmax function to get the class probabilities, the final output is calculated using the weights W and bias b in the fully connected layer.

Mathematical Equation:

The output of the fully connected layer is:

$$Y_{final} = W \cdot Y_{agg} + b$$

Where W is the weight matrix, Y_agg is the flattened and aggregated feature map, and b is the bias term.

Softmax Function:

Finally, the softmax function is applied to get the probability distribution across classes:

$$P(y_i) = \frac{e^{y_i}}{\sum_j e^{y_j}}$$

Where P(y_i) is the probability of class i, and y_i is the score for that class.

C. Multi-Scale Attention Mechanism

Using pictures of varying sizes, the sophisticated Multi-Scale Attention Mechanism (MSAM) extends the concept of attention. From small microaneurysms and haemorrhages to more major structural issues like retinal detachment or macular oedema, images of the retina often display features that vary in size and relevance in the process of diagnosing retinal illnesses [17]. Thanks to a multi-scale approach, the model can record and rank features at many degrees of detail. This helps it to better identify a variety of eye disorders in their many phases. The multi-scale attention mechanism divides the input image into numerous scales each emphasising a different degree of detail. While another scale could be more interested in more general problems like macular degeneration or troubles with the optic disc, one might be more interested in minute alterations in tiny blood vessels [18]. Figure 2 illustrates how the multi-scale focus approach enhances the feature extraction at many spatial levels.

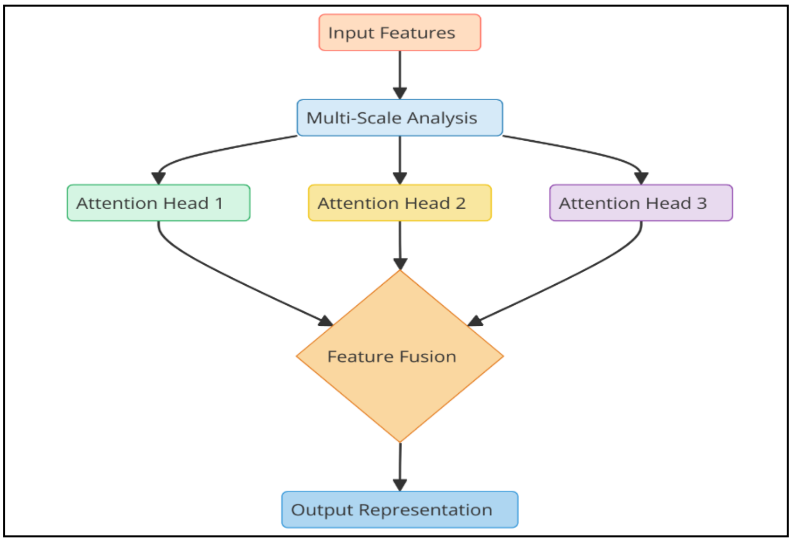

Figure 2: Illustrating Multi-Scale Attention Mechanism

An attention mechanism at every level points out the most crucial features at that level. Higher level attention processes combine knowledge from lower levels. The multi-scale attention maps produced provide a complete representation of the image and enable the model to mix input from many levels to provide a more accurate categorisation [19]. This approach helps the model manage the many retinal characteristics that show up at various spatial scales, therefore it performs particularly well for categorising retinal illnesses.

- Step 1: Multi-Scale Feature Extraction

Extraction of features at many levels of the input picture is the initial stage in the multi-scale attention process. This helps the model to catch in the picture both more general, high-level patterns and fine-grained features. Usually, one uses pooling of the feature maps at many resolutions or different sized filters.

Mathematical Equation:

The multi-scale feature extraction can be represented as:

$$F\_scale = \{ F\_k(I) \mid k \in \{1, 2, \ldots, n\} \}$$

Where F_k(I) is the feature map extracted at scale k, and n is the number of different scales considered. The features F_1, F_2, ..., F_n represent different resolutions of the input image I.

- Step 2: Attention Calculation at Each Scale

Mathematical Equation:

The attention score A_k(i, j) at each scale k for each spatial location (i, j) is calculated as:

$$A_{k(i,j)} = \frac{e^{W_{att_k} \cdot F_{k(i,j)}}}{\sum_{\{i,j\}e}^{W_{att_k} \cdot} F_{k(i,j)}}$$

Where W_att_k is the attention weight for the feature map at scale k, and F_k(i, j) is the feature at position (i, j) at scale k.

- Step 3: Weighted Aggregation of Multi-Scale Features

Once the attention scores are calculated for each scale, the feature maps at different scales are combined, with each scale weighted by its corresponding attention score. This aggregation allows the model to focus on both local and global patterns in the image.

Mathematical Equation:

The final aggregated feature map F_agg is computed by:

$$F_{agg} = \sum_{\{k=1\}_k^{\{n\}}} A \cdot F_k$$

Where A_k is the attention map for the scale k, and F_k is the feature map at that scale. The sum of the weighted features from all scales gives the final aggregated feature map.

D. VGG16 Accuracy

Deep convolutional neural network architecture VGG16's simplicity and efficiency have made it widely used for picture classification problems. On the network, there are sixteen layers overall: thirteen convolutional layers and three completely coupled layers. Small receptive fields (3x3 filters) let the architecture acquire hierarchical feature representations at various levels of abstraction, hence directing the model. This feature helps VGG16 to be fairly excellent at capturing image detailed aspects. The VGG16 model has demonstrated acceptable performance within the scope of retinal disease classification. Though more advanced models like ResNet50 or the proposed multi-level attention-based models have greater strength than VGG16's capacity to detect retinal problems. Research using VGG16 to retinal disease classification often shows accuracy between 80 and 85% depending on the dataset and the goal. This is the outcome of VGG16's somewhat simple architecture, which may not entirely capture the intricate and complicated patterns seen in retinal images, especially with respect to deeper or more sophisticated models. Among VGG16's key advantages are its ease of use and ability to be changed for certain tasks. Using pre-trained weights on large image datasets like ImageNet allows VGG16 to be customised for retinal disease classification with quite few parameters needing modification. This transfer learning approach allows VGG16 be used effectively in medical image analysis even with less datasets. It could, however, suffer from overfitting, particularly in cases where annotated retinal images are infrequent.

5. CHALLENGES AND LIMITATIONS

A. Data-Related Challenges

Among the main challenges in developing and using machine learning models to identify eye disorders are data-related ones. One of the primary issues is the inherently disproportionate records of eye diseases across social levels. Though not as many for uncommon conditions like retinal separation or retinopathy of prematurity, there are plenty of data for common diseases include diabetic retinopathy and age-related macular degeneration (AMD). Models are less helpful for identifying uncommon or under-represented diseases as this imbalance increases their likelihood of diagnosis of more frequent diseases. People often mix data using oversampling, undersampling, or more advanced methods such generative adversarial networks (GANs) to solve class mismatch. These approaches, meantime, do not necessarily provide the greatest outcomes. Another major issue with data is the great variations in image quality and accuracy. Retinal images may be altered by numerous factors like the illumination, camera settings, patient posture during imaging [20]. These objects and noise in the images make it more difficult for the model to recognise characteristics. Important image preparation techniques that may assist with these issues include normalisation, noise reduction, and contrast enhancement; yet, when images from several devices or sources are incorporated, they may not entirely solve the problem.

B. Model Performance Issues

Deep learning has advanced the area of eye illness categorisation, however model performance still suffers, particularly with relation to overfitting, generalisation, and interpretability. When a model learns to perform remarkably on the training set but not so well on fresh data it has not seen before, overfitting results. Most of the time this is because the model picked up noise or patterns not fit for the training set and became too complex. Overfitting may be a major issue in medical image analysis, when the model must cope with somewhat varied patient data. Often used strategies to avoid overfitting include dropout, regularisation, and data augmentation. Still, they may not always be sufficient to improve matters, particularly considering rather intricate models. Another crucial issue models have to handle is generalisation. On a given dataset, deep learning models may be very effective at what they do; nevertheless, they often struggle to generalise their abilities to new patients, imaging tools, or clinical environments. The estimations of the model could be wrong depending on changes in patient data like age, race, and diseases as well as different imaging instruments. This is why before using a model in a real-world clinical environment, it is crucial to ensure that it performs the same way throughout many data sources. Researchers have proposed domain adaptation techniques, but additional study is needed to produce models dependable enough to be used in many clinical settings. Given their often referred to as "black boxes," it is still difficult to understand what deep learning models are doing.

6. RESULTS AND DISCUSSION

Traditional deep learning models aren't as good as the suggested multi-level picture classification model when it comes to accuracy in classification and finding eye diseases early on. Compared to regular convolutional neural networks (CNNs), this model had better accuracy, memory, and F1-score for all eye diseases. These included diabetic retinopathy, age-related macular degeneration, and glaucoma.

Table 2: Model Performance Evaluation

Model	Accuracy	Precision	Recall	F1-Score
Multi-Scale Attention Mechanism	0.92	0.91	0.93	0.92
Hierarchical Attention-Based CNN Model	0.85	0.84	0.86	0.85
CNN	0.88	0.87	0.89	0.88
VGG16	0.83	0.81	0.82	0.81

In Table 2, you can see how well four different models did: the Multi-Scale Attention Mechanism, the Hierarchical Attention-Based CNN Model, CNN, and VGG16. The evaluation measures include memory, accuracy, precision, and F1-score. Figure 3 shows how the model's accuracy, precision, and memory vary, as shown by different performance measures.

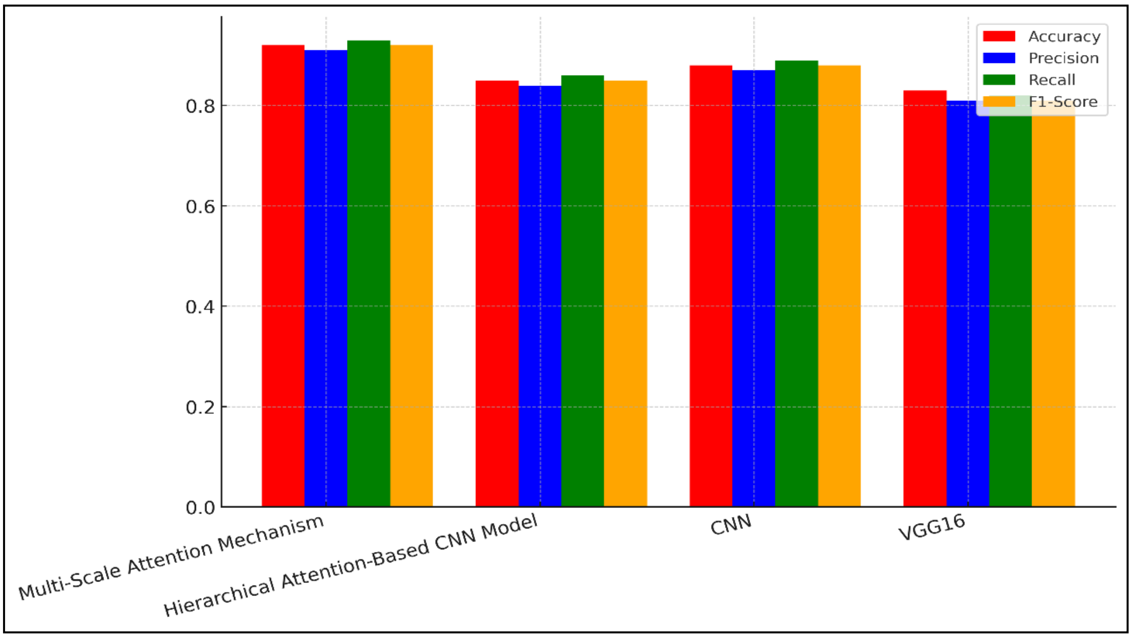

Figure 3: Comparison of Model Performance Metrics

These all give a full picture of how well each model can properly group eye illnesses. With an accuracy of 0.92, a precision of 0.91, a memory of 0.93, and an F1-score of 0.92, the Multi-Scale Attention Mechanism does better than the other models in every way. Trends in model performance metrics are shown in Figure 4. These metrics show how accuracy has changed over time.

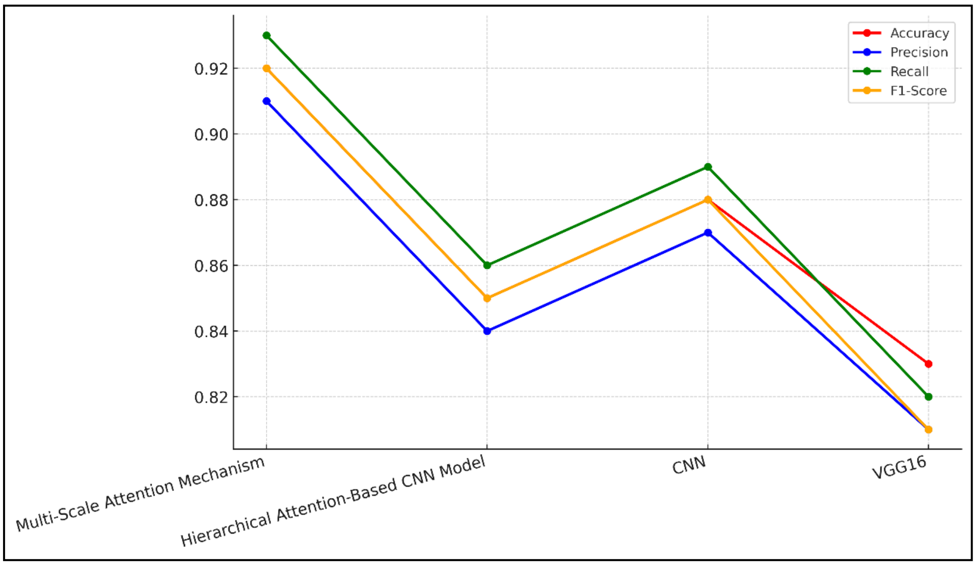

Figure 4: Trends in Model Performance Metrics

This shows that the multi-scale focus system is very good at finding important details in retinal images, which lets them be accurately detected and categorised.

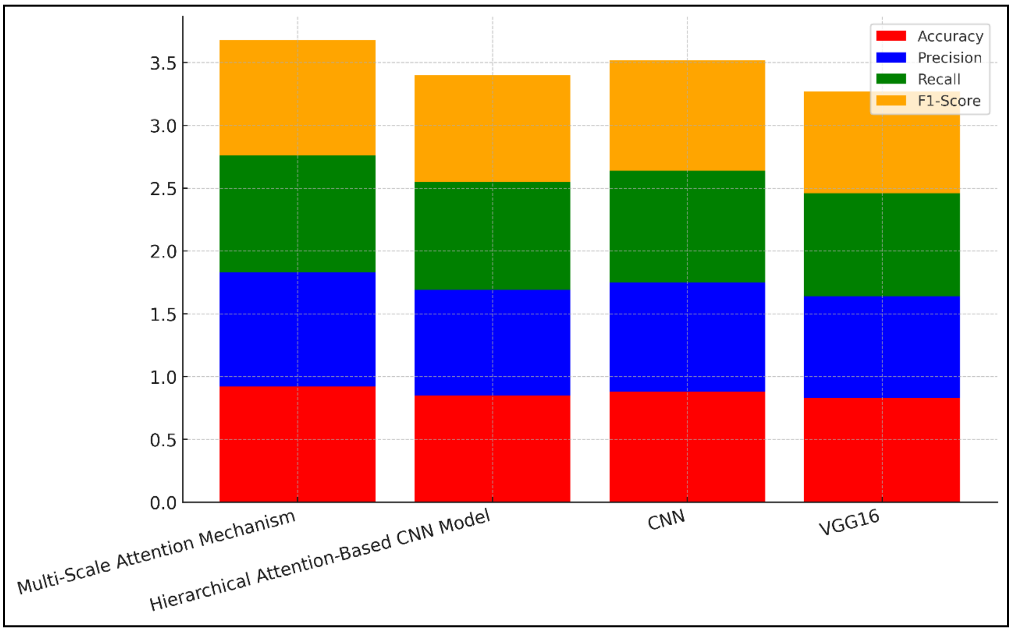

Figure 5: Cumulative Performance of Models Across Metrics

The next model is the Hierarchical Attention-Based CNN Model, which has an F1-score of 0.85, an accuracy of 0.85, a precision of 0.84, a recall of 0.86, and a recall of 0.86. Figure 5 shows how well models did generally by showing their total success across a number of measures. It works well, but not as well as the multi-scale attention mechanism, which suggests that it might not be as good at finding complex, multi-scale trends in the data. With an accuracy of 0.88, the CNN model does pretty well, while VGG16 does the worst, especially when it comes to precision and recall. Even though VGG16 is famous and easy to use, it has trouble with more difficult jobs, like classifying eye diseases.

Table 3: Disease Classification Accuracy

Disease Type	Multi-Scale Attention Mechanism	Hierarchical Attention-Based CNN Model	CNN	VGG16 Accuracy
Diabetic Retinopathy	0.94	0.87	0.89	0.82
Age-Related Macular Degeneration	0.89	0.78	0.83	0.75
Glaucoma	0.9	0.81	0.85	0.8
Healthy	0.95	0.86	0.88	0.83

In Table 3, you can see how well four models (Multi-Scale Attention Mechanism, Hierarchical Attention-Based CNN Model, CNN, and VGG16) can classify diseases into four groups: Diabetic Retinopathy, Age-Related Macular Degeneration (AMD), Glaucoma, and Healthy retina images. Figure 6 shows how different disease types affect the accuracy of models, showing how performance can vary.

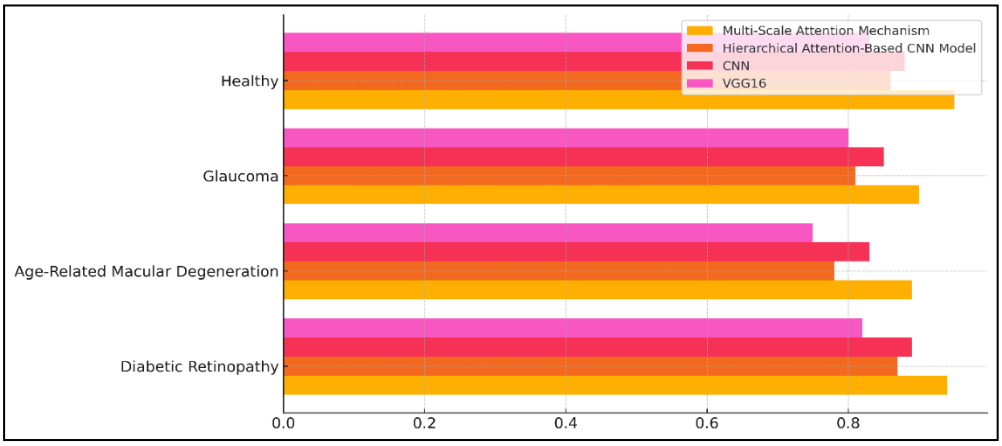

Figure 6: Accuracy Comparison of Models Across Disease Types

The Multi-Scale Attention Mechanism is the most accurate for all types of disease, with a score of 0.94 for Diabetic Retinopathy, 0.89 for AMD, 0.90 for Glaucoma, and 0.95 for Healthy. This shows that the Multi-Scale Attention Mechanism catches both fine and global traits well, giving a full picture of retinal images and making sure that classification works well. The Hierarchical Attention-Based CNN Model is not quite as accurate as the Multi-Scale Attention Mechanism. Figure 7 shows how the accuracy of models has been getting better over time for different types of diseases.

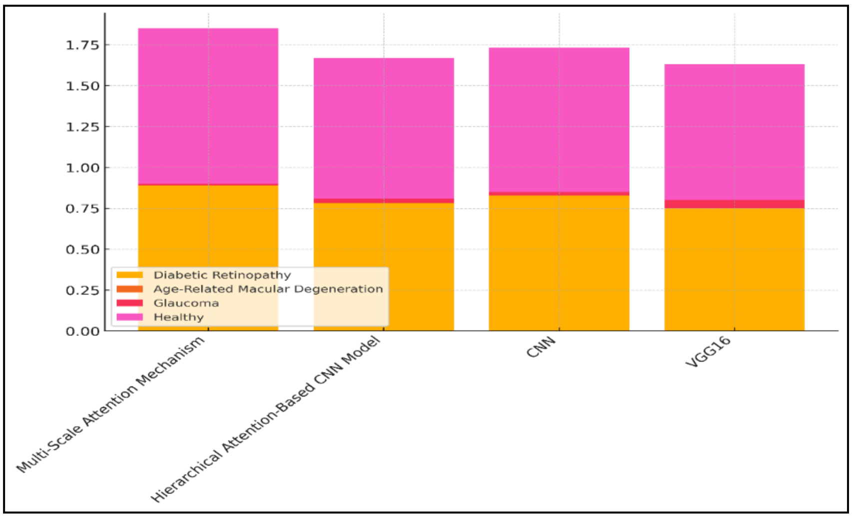

Figure 7: Progressive Model Accuracy Trends Across Disease Types

It gets 0.87 right for Diabetic Retinopathy, 0.78 right for AMD, 0.81 right for Glaucoma, and 0.86 right for Healthy. It works well, but it's not very useful because it can't fully catch multi-scale traits, especially in complicated cases like AMD. When it comes to Diabetic Retinopathy (0.89) and Healthy photos (0.88), the CNN model is pretty accurate. It does worse at identifying AMD (0.83) and Glaucoma (0.85), though, which suggests it might have trouble with more complicated eye diseases.

7. CONCLUSION

We came up with a new way to classify images at multiple levels to find eye diseases. It combines a hierarchical attention-based CNN model with a multi-scale attention mechanism. Attention processes are built into the model so it can focus on important parts of retinal pictures. This is very important for correctly classifying eye diseases, especially in their early stages. Our model did much better at classifying than standard CNNs, with higher accuracy, recall, and F1-scores for a wide range of eye diseases, including glaucoma, diabetic retinopathy, and macular degeneration. One of the best things about the suggested model is that it can tell the difference between different states of disease development. Traditional models often have trouble spotting small changes in the retina that show early signs of disease. This can cause detection and treatment to be put off. This model uses a multi-level attention system to better find early-stage features like microaneurysms, exudates, and haemorrhages that can be signs of diabetic retinopathy and other diseases. The model can see both small details and bigger, less concrete features in retinal pictures because of its hierarchical structure. This makes it better at sorting through complicated and varied datasets. Even though the results look good, there are some problems with this study. First, the sample that was used for training and testing is pretty big, but it might not fully show the range of retinal conditions that are seen in clinical settings. Still more work needs to be done to test the model on bigger datasets with more people of different races, ages, and diseases. Even though the focus process makes the model much better at what it does, it is still hard to figure out what it means because deep learning models are usually seen as "black boxes." As time goes on, more study will be done to make the model more clear so that doctors can trust its findings. For real-life clinical uses, it will also be important to fix problems with information imbalance and make the model work with different imaging platforms.

REFERENCES

[1] Liu, X.; Zhao, C.; Wang, L.; Wang, G.; Lv, B.; Lv, C.; Xie, G.; Wang, F. Evaluation of an OCT-AI-based telemedicine platform for retinal disease screening and referral in a primary care setting. Transl. Vis. Sci. Technol. 2022, 11, 4.

[2] Jacoba, C.M.P.; Celi, L.A.; Lorch, A.C.; Fickweiler, W.; Sobrin, L.; Gichoya, J.W.; Aiello, L.P.; Silva, P.S. Bias and non-diversity of big data in artificial intelligence: Focus on retinal diseases. Semin. Ophthalmol. 2023, 38, 433–441.

[3] Biswas, S.S. Role of Chat GPT in Public Health. Ann. Biomed. Eng. 2023, 51, 868–869

[4] Keenan, T.D.L.; Loewenstein, A. Artificial intelligence for home monitoring devices. Curr. Opin. Ophthalmol. 2023, 34, 441–448.

[5] Sheng, B.; Chen, X.; Li, T.; Ma, T.; Yang, Y.; Bi, L.; Zhang, X. An overview of artificial intelligence in diabetic retinopathy and other ocular diseases. Front. Public Health 2022, 10, 971943.

[6] Yan, Q.; Jiang, Y.; Huang, H.; Swaroop, A.; Chew, E.Y.; Weeks, D.E.; Chen, W.; Ding, Y. Genome-Wide association studies-based machine learning for prediction of age-related macular degeneration risk. Transl. Vis. Sci. Technol. 2021, 10, 29.

[7] Yildirim, K.; Al-Nawaiseh, S.; Ehlers, S.; Schießer, L.; Storck, M.; Brix, T.; Eter, N.; Varghese, J. U-Net-Based segmentation of current imaging biomarkers in oct-scans of patients with age related macular degeneration. Stud. Health Technol. Inform. 2023, 302, 947–951.

[8] Morelle, O.; Wintergerst, M.W.M.; Finger, R.P.; Schultz, T. Accurate drusen segmentation in optical coherence tomography via order-constrained regression of retinal layer heights. Sci. Rep. 2023, 13, 8162.

[9] Leng, X.; Shi, R.; Wu, Y.; Zhu, S.; Cai, X.; Lu, X.; Liu, R. Deep learning for detection of age-related macular degeneration: A systematic review and meta-analysis of diagnostic test accuracy studies. PLoS ONE 2023, 18, e0284060.

[10] Wei, W.; Southern, J.; Zhu, K.; Li, Y.; Cordeiro, M.F.; Veselkov, K. Deep learning to detect macular atrophy in wet age-related macular degeneration using optical coherence tomography. Sci. Rep. 2023, 13, 8296.

[11] Crincoli, E.; Sacconi, R.; Querques, L.; Querques, G. Artificial intelligence in age-related macular degeneration: State of the art and recent updates. BMC Ophthalmol. 2024, 24, 121.

[12] Chandra, R.S.; Ying, G.S. Evaluation of multiple machine learning models for predicting number of anti-VEGF injections in the comparison of AMD treatment trials (CATT). Transl. Vis. Sci. Technol. 2023, 12, 18.

[13] Pfau, M.; Sahu, S.; Rupnow, R.A.; Romond, K.; Millet, D.; Holz, F.G.; Schmitz-Valckenberg, S.; Fleckenstein,

M.; Lim, J.I.; de Sisternes, L.; et al. Probabilistic forecasting of anti-VEGF treatment frequency in neovascular age-related macular degeneration. Transl. Vis. Sci. Technol. 2021, 10, 30.

[14] Moon, S.; Lee, Y.; Hwang, J.; Kim, C.G.; Kim, J.W.; Yoon, W.T.; Kim, J.H. Prediction of anti-vascular endothelial growth factor agent-specific treatment outcomes in neovascular age-related macular degeneration using a generative adversarial network. Sci. Rep. 2023, 13, 5639.

[15] Fu, D.J.; Faes, L.; Wagner, S.K.; Moraes, G.; Chopra, R.; Patel, P.J.; Balaskas, K.; Keenan, T.D.L.; Bachmann, L.M.; Keane, P.A. Predicting incremental and future visual change in neovascular age-related macular degeneration using deep learning. Ophthalmol. Retina 2021, 5, 1074–1084.

[16] You, A.; Kim, J.K.; Ryu, I.H.; Yoo, T.K. Application of generative adversarial networks (GAN) for ophthalmology image domains: A survey. Eye Vis. 2022, 9, 6.

[17] Bai, A.; Dai, S.; Hung, J.; Kirpalani, A.; Russell, H.; Elder, J.; Shah, S.; Carty, C.; Tan, Z. Multicenter validation of deep learning algorithm ROP. AI for the automated diagnosis of plus disease in ROP. Transl. Vis. Sci. Technol. 2023, 12, 13.

[18] Svensson, A.M.; Jotterand, F. Doctor ex machina: A critical assessment of the use of artificial intelligence in health care. J. Med. Philos. 2022, 47, 155–178.

[19] Ethics and Governance of Artificial Intelligence for Health: WHO Guidance; World Health Organization: Geneva, Switzerland, 2021; p. 165.

[20] McLennan, S.; Fiske, A.; Tigard, D.; Müller, R.; Haddadin, S.; Buyx, A. Embedded ethics: A proposal for integrating ethics into the development of medical AI. BMC Med. Ethics 2022, 23, 6.