IoT-Opthom-CAD: IoT-Enabled Classification System of Multiclass Retinal Eye Diseases Using Dynamic Swin Transformers and Explainable Artificial Intelligence

Talal AlBalawi, Mutlaq B. Aldajani, Qaisar Abbas, Yassine Daadaa
College of Computer and Information Sciences, Imam Mohammad Ibn Saud Islamic University (IMSIU), Riyadh

***Abstract*—Integrating Internet of Things (IoT)-assisted eye-related recognition incorporates connected devices and sensors for primary analysis and monitoring of eye conditions. Recent advancements in IoT-based retinal fundus recognition utilizing deep learning (DL) have significantly enhanced early analysis and monitoring of eye-related diseases. Ophthalmologists use retinal images in the diagnosis of different eye diseases. Numerous computer-aided diagnosis (CAD) studies have been conducted by using IoT and DL technologies on the early diagnosis of eye-related diseases. The retina is susceptible to microvascular alterations due to numerous retinal disorders. This study creates a new, non-invasive CAD system called IoT-Opthom-CAD. It uses Swin transformers and the gradient boosting (LightGBM) method to find different eye diseases in colored fundus images after applying data augmentations techniques. We introduce a Swin transformer (dc-swin) that is efficient and powerful by connecting a dynamic cross-attention layer to extract local and global features. In practice, this dynamic attention layer suggests a mechanism where the model dynamically focuses on different parts of the image at other times, learning to cross-reference or integrate information across these parts. Next, the LightGBM method is used to divide these features into multiple groups, including normal (NML), diabetic retinopathy (DR), tessellation (TSN), age-related macular degeneration (ARMD), Optic Disc Edema (ODE), and hypertensive retinopathy (HR). To find the causes of eye-related diseases, the Grad-CAM is used as an explainable artificial intelligence (xAI). To develop the Opthom-CAD system, preprocessing, and data augmentation steps are integrated to strengthen this architecture. Multi-label three retinal disease datasets, such as MuReD, BRSET, and OIA-ODIR, are utilized to evaluate this system. After ten times of cross-validation tests, the proposed Opthom-CAD system shows excellent results such as an AUC of 0.95, f1-score of 95.7, accuracy of up to 96.5%, precision of 95%, recall of 94% and f1-score of 95.7. The results indicated that the performance of the Opthom-CAD system is much better than that of numerous baseline state-of-the-art models. As a result, the Opthom-CAD system can assist dermatologists in detecting eye-related diseases. The source code is public and accessible for anyone to view and modify from GitHub (https://github.com/Qaisar256/Opthom-CAD).**

Keywords*—Computer-aided diagnosis; ophthalmology; multiclass classification; tessellation; age-related macular degeneration; Optic Disc Edema (ODE); hypertensive retinopathy; data augmentation; transformers; Swin; explainable AI; Internet of Things

I. Introduction

The global burden of eye disorders, affecting 2.2 billion people, highlights fundus diseases as a significant cause of blindness (WHO [1]). These conditions, such as diabetic retinopathy (DR), age-related macular degeneration (ARMD), and hypertensive retinopathy (HR), often go undetected until they are severe due to their asymptomatic early stages. Early diagnosis and intervention are crucial to prevent irreversible vision loss [2, 3]. Traditional machine learning has helped analyze small datasets with manually engineered features. Deep learning (DL) has revolutionized the identification of a wide range of eye ailments, including tessellation (TSN) and optic disc edema (ODE), through extensive screening with fundus photographs [4, 5]. In ophthalmology, computer-aided diagnosis (CAD) systems have been developed to increase the accuracy of detecting eye-related diseases [6]. The researchers used image processing and machine-learning techniques to create CAD systems to distinguish various eye-related diseases. Retinal fundus images obtained by fundus cameras provide detailed patterns of each eye disease. Alterations in retinal arteries in fundus images can indicate vascular disorders, such as cardiovascular conditions. However, it is still challenging to identify eye diseases like glaucoma, cataracts, DR, TSN, ARMD, ODE, and HR through CAD systems [7–10].

When AI (artificial intelligence) [11–15] techniques like ML are added to CAD systems, they make it easier to classify eye diseases that are found through fundus devices. Nowadays, deep learning (DL) methods are categorized as ML, capturing more complex features from images to recognize eye-related disease disorders. In the past, the CAD systems diagnosed limited categories of eye-related diseases. Therefore, to address this issue, we have developed the IoT-Opthom-CAD system. This system, which incorporates the Internet of Things (IoT) technology, presents an innovative DL system. It is specifically designed to diagnose various eye-related diseases efficiently and test using IoT devices. The IoT-Opthom-CAD has excelled in classifying eye-related diseases through several hyperparameter fine-tuning and optimization steps.

The major contributions of this paper are given as follows:

1) We introduce a Swin transformer (Swin-DCL) that is efficient and powerful by connecting a dynamic cross-

attention layer to extract local and global features. In practice, this dynamic attention layer suggests a mechanism where the model dynamically focuses on different parts of the image at other times, learning to cross-reference or integrate information across these parts.

2) The study introduces a novel IoT-based framework approach in ophthalmology diagnostics by combining lightweight Swin transformers with gradient boosting techniques, specifically LightGBM. This innovative method balances computational efficiency and high diagnostic performance, potentially revolutionizing disease detection in this field.

3) Applying Grad-CAM to explain the decision-making process for identifying eye diseases enhances model transparency and interpretability. While Grad-CAM is used in various fields, its application in elucidating diagnostic pathways in eye health through this new architecture is innovative.

4) The system has been validated across multiple datasets, demonstrating superior performance metrics compared to numerous baseline state-of-the-art models. The thorough validation and achieved metrics highlight the system's practical and clinical relevance, adding to its novelty.

II. Literature Review

Eye-related disease can result in several retinal abnormalities, including hard exudates, hemorrhages, microaneurysms, and other symptoms. On a short and constrained dataset, many machine-learning techniques were created to identify eye-related diseases using various image processing and computer-vision-based algorithms for analysis and feature extraction [16]. Advanced deep neural networks, particularly convolutional neural networks (CNN), have recently contributed substantially to medical imaging, as briefly described below. Utilizing a multi-branch neural network (MB-NN), this re-search leverages domain knowledge and retinal fundus images for glaucoma detection [17]. The effectiveness of this model was validated on real datasets, achieving an accuracy of 91.51%, sensitivity of 92.33%, and specificity of 90.90%. This showcases the model's capability to diagnose glaucoma, even with limited data, efficiently. This study developed a deep learning (DL) algorithm to predict referable glaucomatous optic neuropathy (GON) [18] from color fundus images. The research in study [19] utilizes convolutional neural networks (CNNs) to automate the identification of glaucoma by segmenting the optic cup and disc. This study examines the efficacy of the proposed method in comparison to conventional gradient-based learning [20] and other optimization techniques. The method employs an artificial algae optimization technique to enhance a novel deep learning system.

These studies address cataract detection and classification through various methodologies, including hybrid approaches and novel networks [21–25]. Utilizing datasets from several open-access sources and employing different CNNs, the methods achieve up to 96.25% accuracy in 4-class classification. These results underscore the potential of AI for enhancing cataract diagnosis and classification accuracy. Focusing on AMD, these papers propose different deep learning frameworks for its early detection and classification [26–29], achieving high diagnostic accuracy. For instance, one study utilized a comprehensive CAD framework, extracting local and global appearance markers from fundus images, and achieved an accuracy of 96.85%. These studies illustrate the efficacy of deep learning in identifying and categorizing AMD stages accurately.

Addressing DR, these studies introduce various deep learning approaches, from hybrid techniques to novel algorithms [30–34], significantly improving detection and classification. One method, using transfer learning on pre-trained CNN models, achieved an accuracy of 97.8% for binary classification. The advancements demonstrate the critical role of AI in early DR detection, potentially preventing vision loss. Spanning a wide range of deep learning methodologies, these studies collectively push the boundaries of ocular disease diagnostics [35–37]. For instance, a system that aimed to identify various ocular diseases achieved F1 scores as high as 88.56% and an AUC of 99.76%. These diverse approaches showcase the power of AI in diagnosing a broad spectrum of ocular conditions with high accuracy and efficiency. Each study's use of specific datasets and results highlights the transformative impact of deep learning in ophthalmology, offering new avenues for early detection, accurate diagnosis, and effective treatment of various eye diseases. Those state-of-the-art systems are compared in Table I.

TABLE I. State-of-the-Art Comparisons of Deep Learning Model for Recognition of Eye-Related Diseases

Cited Work	Methodology	Targeted Disease	Classes	Results	Limitations
[17]	Multi-branch neural network model for combining domain knowledge with retinal fundus images	Glaucoma	Binary (Glaucomatous/Non-Glaucomatous)	Accuracy: 91.51%, Sensitivity: 92.33%, Specificity: 90.90%	Relies on domain knowledge and important image regions
[18]	Deep learning algorithm for predicting referable glaucomatous optic neuropathy from fundus images	Glaucomatous Optic Neuropathy	Binary (Referable/Non-Referable)	AUC: 0.945, 0.855, 0.881 depending on dataset	Requires large dataset for training
[19]	Deep Learning with CNN for optic disc and cup segmentation	Glaucoma	Binary (Glaucomatous/Non-Glaucomatous)	Accuracy: 95.8% for disc, 93% for cup segmentation	Focuses on optic disc and cup segmentation
[20]	Deep learning with artificial algae optimization algorithm for glaucoma diagnosis	Glaucoma	Binary (Glaucomatous/Non-Glaucomatous)	High performance metrics (Accuracy: 98.15%)	Compares with traditional and other optimization methods
[21]-[24]	Various methodologies involving pre-trained CNNs, ensemble learning, and SVMs for cataract detection and classification	Cataract	Multi-class (Normal, Mild, Moderate, Severe)	Up to 96.25% accuracy	Varies, including image quality selection

[25]	Supervised miniature U-Net integrated with CNN for cataract detection and localization	Cataract	Binary (Cataract/Normal)	Accuracy: 96% with CLR	Focuses on early detection with CLR optimization
[26]-[28]	Various deep learning approaches for detecting and classifying age-related macular degeneration	Age-related Macular Degeneration	Multi-class for various AMD stages and types	Up to 98.76% AUC	Emphasizes on early detection and precise diagnosis
[29]	Explainable deep learning approach for AMD diagnosis through retinal lesion identification	Age-related Macular Degeneration	Binary/Multi-class for AMD and associated retinal lesions	-	Offers lesion-specific information for clinicians
[30]-[34]	Various deep learning models for detecting and classifying diabetic retinopathy	Diabetic Retinopathy	Binary and Multiclass for various DR stages	Up to 97.8% accuracy for binary classification	Focuses on early detection and classification
[35]-[37]	Deep learning models for retinal vessel segmentation	Various retinal disorders	-	High segmentation performance metrics	Addresses challenges in vessel segmentation

III. Proposed Methodology

The approach seeks to improve the precision and effectiveness of diagnosis by combining various processes, utilizing the capabilities of IoT and cloud technologies. This technique aims to offer a resilient solution for remote healthcare diagnostics. The proposed framework for detecting and classifying multi-class retinal disorders, known as the IoT-Opthom-CAD framework, is graphically depicted in Fig. 1.

B. Data Acquisition

In this study, an effective IoT-enabled technique has been developed for skin lesion diagnosis in IoT environment. We have developed and trained the IoT-Opthom-CAD system based on three online sources, such as multilabel retinal disease (MuReD) [38], the Brazilian multilabel ophthalmological dataset of retina fundus photos (BRSET) [39], and the ophthalmic image analysis-ocular disease intelligent recognition (OIA-ODIR) dataset [40]. We have collected initially 6,00 fundus images from these sources, including an average (NOM) of 1900, diabetic retinopathy (DR) of 2000, glaucoma (GLC) of 400, cataracts (CAT) of 200, age-related macular degeneration (AMD) of 300, and hypertension retinopathy (HR) of 1200 images. To balance the selected dataset, we have applied data augmentation and preprocessing to convert 6,000 images into 12,000 retinographics. Given that the images come from different sources, the resolution can vary from 520x520 to 3400x2800, depending on the source of the image. We have resized them to 224x224. Among these, the MuReD dataset stands out for its comprehensive collection of 2,208 images spanning 20 distinct categories. In parallel, the Brazilian Multilabel Ophthalmological Dataset (BRSET) emerges as a groundbreaking resource within Latin America, aiming to bridge the gap in the availability of public ophthalmological datasets. BRSET encompasses 16,266 color fundus photographs from 8,524 Brazilian patients, incorporating rich sociodemographic data to bolster its value as both a research tool and an educational resource. The Ophthalmology Image Analysis and Ocular Disease Intelligent Recognition (OIA-ODIR) dataset, a pioneering global resource for identifying multiple ocular diseases using fundus imagery. With 10,000 fundus images from 5,000 patients, it covers eight different ocular conditions, making it a vital tool for developing and testing deep-learning models in ophthalmology. All the numerical collected samples are shown as distribution in Fig. 2.

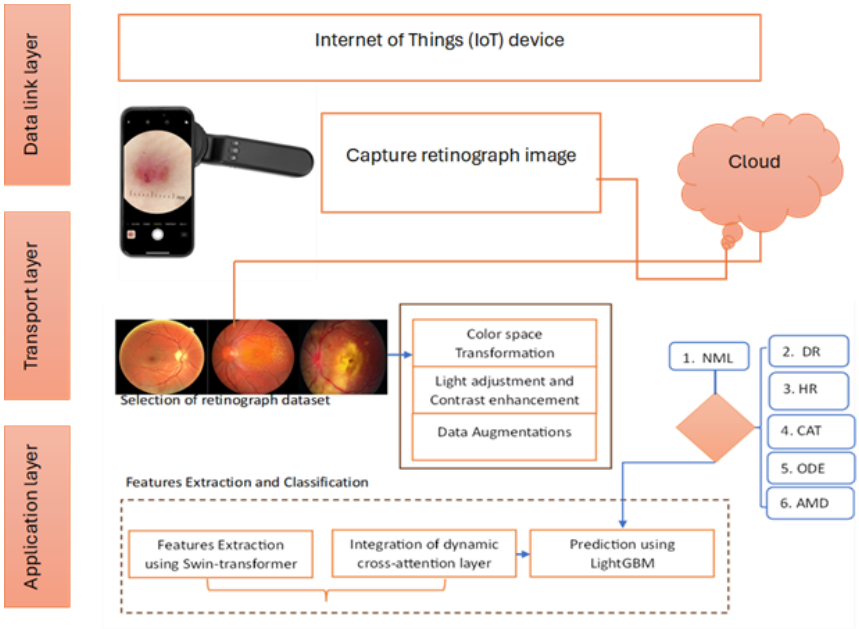

Fig. 1. A systematic flow diagram of proposed IoT-Opthom-CAD system.

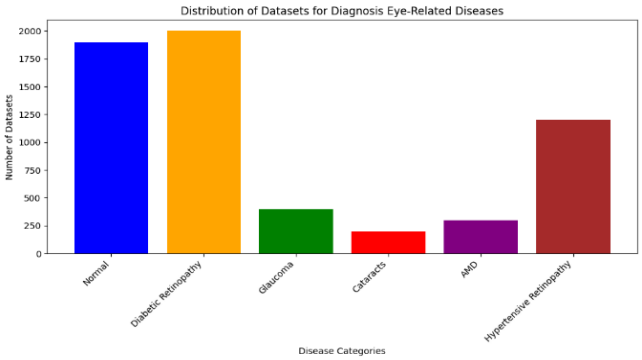

Fig. 2. A visual diagram of collection of datasets from different sources.

The MuReD, BRSET, and OIA-ODIR datasets as shown in Fig. 3 improve ophthalmic medical imaging and computer vision. By providing varied, high-quality data sources, they allow sophisticated diagnostic tools that are more accurate, moderate, and representative of the real-world population. This advancement helps diagnose and treat eye disorders early and advances artificial intelligence in healthcare, offering improved patient outcomes and medical research.

Amplification of the dataset is done to prevent misclassification caused by unbalanced data since the standard class of the finalized dataset has the most retinographics, and other classes have fewer images than the regular class. Based on the numerous fundus image acquisition capabilities, augmentation techniques were chosen. Different geometric transformations, such as proper 15 rotations, left 15 rotations, right 8 rotations, left 8 rotations, and horizontal flips, are included in the selected augmentation techniques. Training, validation, and testing sets were created from the supplemented dataset in the following ratio: 14:3:3.

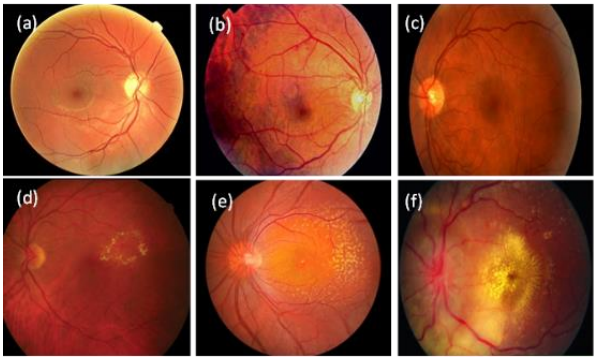

Fig. 3. An example of dataset acquired from different resources such as MuReD, BRSET, and OIA-ODIR, where figure (a) Shows the normal, (b) Represents the diabetic retinograph (DR) , figure (c) SHOWS the Glaucoma, figure (d) Display Cataracts, figure (e) Shows age-related macular degeneration (AMD), and (f) Represents hypertensive retinopathy (HR).

C. Color Preprocessing

All images are transformed into CIECAM02 color appearance model. This study introduces a novel method as shown in Algorithm 1 for enhancing low-light images, specifically aimed at improving the contrast and brightness of retinograph images while preserving intricate details. Initially, the non-uniform RGB retinograph images are transformed into the uniform CIECAM02 color space, where J denotes lightness, C represents chroma, and H signifies hue. Subsequently, a bicubic kernel is employed to extract both low and high frequencies from the J-plane of the CIECAM-02 color space. Color correction is then implemented using a sigmoid function to normalize the low frequencies. Following this, a white balancing step determines the ideal linear combi-nation of color-corrected channels. We got this combination by using constrained linear least squares minimization and focusing on the C component and its Jch color space counterpart that its histogram has equalized. Finally, the high frequencies are adjusted relative to the updated low frequencies and reintegrated to generate a sharper output.

Algorithm 1: Preprocessing : Color space transformation RGB to CIECAM02 and enhance the contrast

Input: A 2D array of RGB(x, y) where each row represents a time sample,
low_{freq}: Lower frequency bound for bandpass filter
$high_{freq}$: Upper frequency bound for bandpass filter
Output: *contrast-enhance-image(x, y)*: preprocessed retinograph images
Function color-transformation ($image_{rgb}$):
$Jch = image_{rgb} \rightarrow CIECAM02 - JCH(J, c, h)$
$J = capture - channels(J_{ch}(i, j))$;
$c = capture - channels(J_{ch}(i, j))$;
$h = capture - channels(J_{ch}(i, j))$;
end
Function extract-frequency-low-high (J_{image}):
Ly = $Bicubic_{convolution} - low(J_{image}, \theta, mode = 'same')$;
$H_{image} = Extract - hight - frequency = J - Ly(Q)$;
end
Function color-balance (Ly,μ, σ):
Ly: $low\ frequencies$
μ: $Mean\ of\ low\ frequencies$
σ: $standard\ deviation\ of\ low\ frequencies$
G = $color - balance - low(1/(1 + \exp(Ly - \mu)/\sigma))$;
$return\ G$;
end
Function histogram- equalization (G_{image}):
L-prime = $histogram - equal(G_{image})$;
Return L-prime ;
end
Function modify-high-frequencies ($L - prime$,H):
$L - prime$: $Histogram\ equalized\ version\ of\ the\ luminance\ component$
H: $High\ frequencies$
Gy = $L - prime\ +\ (L - prime\ /\ L)\ \times\ H$;
$return\ Gy$;
end

End of algorithm

D. Proposed Swin-DCL Architecture for Features Extraction

The Swin-DCL design proposes many stages or layers of processing, each serving a distinct purpose. The initial stage involves supplying the preprocessed retinographics as input to the IoT-Opthom-CAD system for feature extraction. This can enhance the accuracy and dependability of the diagnosis. Incorporating dynamic cross-attention into the Swin Transformer was done strategically at crucial stages in the network, such as after the initial patch embedding or within specific transformer blocks. This enhancement enables the model to effortlessly shift its attention across the network, prioritizing the most significant image regions for accurate diagnosis. By combining the Swin Transformer and the dynamic cross-attention layer, the IoT-Opthom-CAD system makes it easy to examine retinograph images. The system

efficiently processes images through hierarchical stages as shown in Fig. 4, extracting features with increasing levels of abstraction. The dynamic cross-attention layer enhances this process by ensuring optimal allocation of the model's attention to the most informative parts of the image for ocular condition diagnosis.

Swin Transformer Block: A standard Swin Transformer block, S, operates on an input feature map, $X \in RH \times W \times C$, where H, W, and C represent the height, width, and number of channels, respectively. The block contains two main operations such as the self-attention mechanism and the multilayer perceptron (MLP). The self-attention mechanism can be represented as follows:

$$Attention(Q,K,V) = softmax\left(\frac{Qk^T}{\sqrt{d_k}}\right) \times V \quad (1)$$

Where Q, K, and V are the queries, keys, and values obtained from X, and dk is the dimension of the key vectors. In the case of the Swin Transformer, the self-attention mechanism is computed within non-overlapping local windows to reduce computational complexity:

$$X'(L) = W - MSA(LN(X(l)) + X(L)) \quad (2)$$

Where LN denotes Layer Normalization, W-MSA is the window-based multi-head self-attention, and $'X'$ is the output feature map that will be passed to the MLP. The MLP with GELU non-linearity is then applied:

$$Y' = MLP(LN(X'^{(l)}) + X'(L)) \quad (3)$$

Where Y is the output of the Swin transformer block, and it is visually represented in Fig. 4.

Dynamic Cross-Attention Layer: The dynamic cross-attention layer, D, aims to allow the attention mechanism to change adaptively based on the input and internal state. This could be formulated as a function that varies the attention weights dynamically:

$$DA(X_t) = softmax(f_\theta(X_t - 1, P)K_T) \times V \quad (4)$$

Where, Xt is the input at time t, P is a set of parameters or features that influence the dynamic behavior (e.g., learned parameters or context-dependent features), and fθ is a learnable function parameterized by θ that computes the queries dynamically. Now, the DynamicSwinTransformer (DST), would integrate the dynamic cross-attention layer into the standard swin transformer block sequence. The composite operation for DST with N blocks could be represented as follows:

$$DST(X) = S_N(DST(S_{N-1}(\cdots DST(S_1(X))\cdots))) \quad (5)$$

Where Si is the i-th Swin Transformer block and D is interspersed between these blocks to modulate the attention based on dynamic factors. Finally, for a classification task, the swin transformer's output would be fed into LightGBM boosting algorithm. The process of dynamic attention layer architecture is visually shown in Fig. 5. Pre-training a dynamic cross-attention layer in a Swin Transformer architecture involves adjusting the attention mechanism to be region-specific and dynamic over the course of training. Let's define the notation for such a pre-training process, focusing on a scenario with five distinct regions as shown in Fig. 6. Fig. 7 shows various regions of input retinograph.

Let's assume our input image IMG(x, y, c) is partitioned into R regions as shown in Fig. 5, where R=5 as visually described in Fig. 11. The Swin transformer processes the input through a series of layers, and at each layer l, it performs self-attention within local windows. The dynamic cross-attention aims to adapt the focus on these regions dynamically, which are pretrained on selected dataset of each retinal disease. For each region $r \in \{1,2,3,4,5\}$, the dynamic cross-attention mechanism at layer l can be represented by a function DLr that computes the attention weights dynamically based on the input feature map X_{lr} and a set of parameters θ_{LR}, which are learned during pre-training the Eq. (7) can be redefined as:

$$DLr(X_{lr}) = softmax\left(\frac{Q_{lr}k^T}{\sqrt{d_k}}K_{lr} + A_{lr}\right) \times V_{lr} \quad (6)$$

Where the Q_{lr}, K_{lr}, and V_{lr} parameters are the queries, keys, and values for region r at layer l, computed from X_{lr}. The A_{lr} parameter is an added term that represents the adaptive component of the attention for region r, influenced by the dynamic parameters θlr. Also, the dk parameter is the dimension of the key vectors. During pre-training, the objective is to learn θ_{LR} for each region r such that the model can attend to different parts of the image in a way that is beneficial for the task at hand (e.g., feature extraction relevant to eye diseases in retinal images). This is achieved by minimizing a loss function L that measures the discrepancy between the model output and the ground truth labels over a pre-training dataset D:

$$\min_{\theta} \theta_{LR}(D;\theta) \quad (7)$$

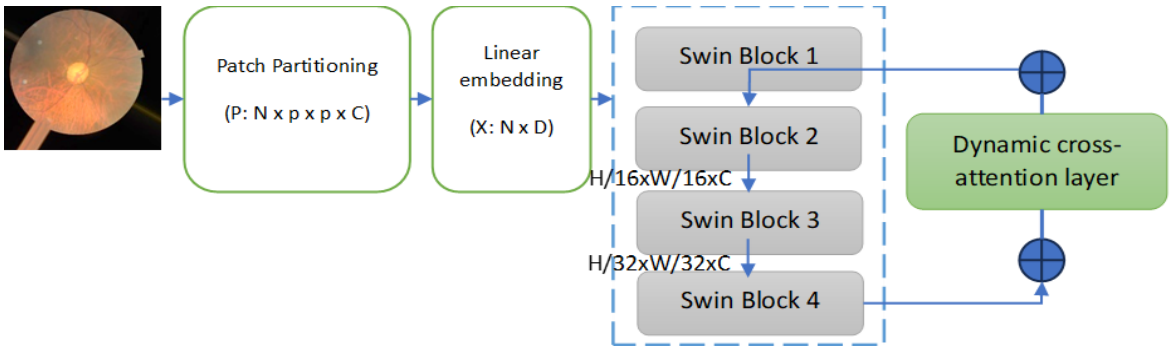

Fig. 4. A swin-DCL architecture with four stages for extracting features from retinograph images.

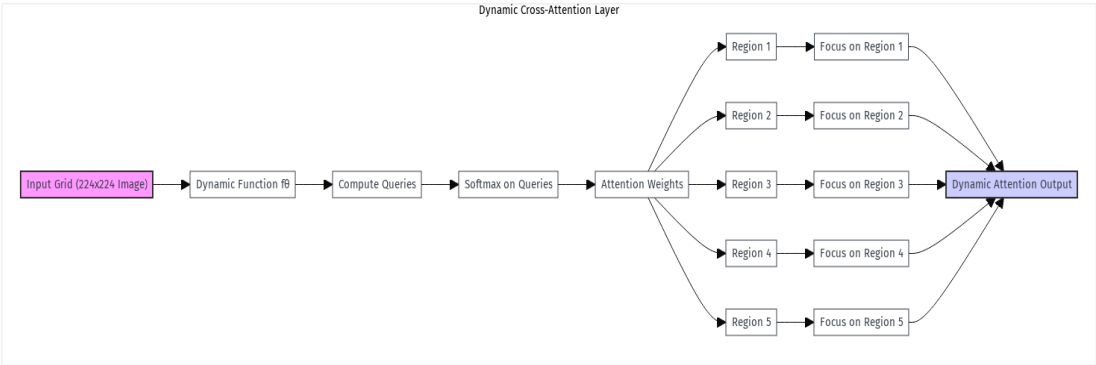

Fig. 5. Illustration of dynamic attention layer architecture with two blocks of swin transformers.

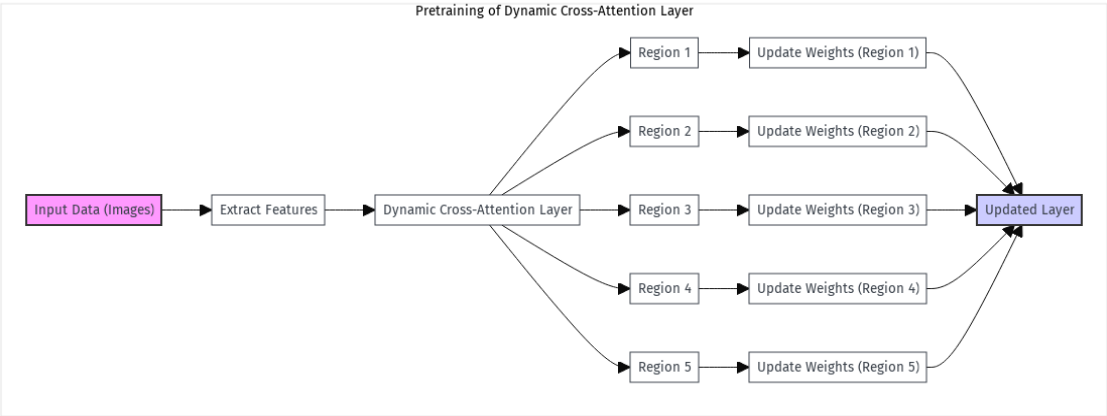

Fig. 6. Pretraining the dynamic attention layer and updating layer for modification of weights.

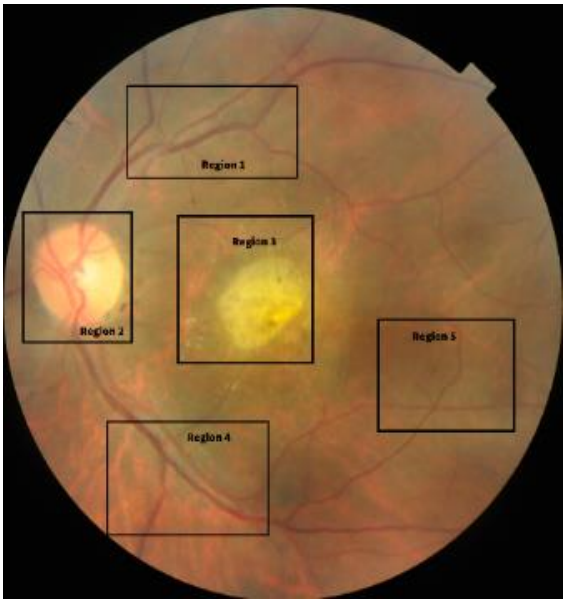

Fig. 7. Various regions of input retinograph is extracted and pretrained a dynamic cross attention layer.

Where θ denotes the set of all parameters, including θ_{LR} for all regions r and layers l. During pre-training, the model is exposed to a variety of images and is encouraged to learn region-specific attention patterns that enhance its ability to extract relevant features from each region. The dynamic aspect allows the model to adjust these patterns as it encounters new data and as it progresses through the layers of the transformer. After pre-training, the learned parameters θ_{lr} for each region r are used to initialize the dynamic cross-attention layers of the Swin Transformer for further fine-tuning on a specific target task, potentially with a new dataset. The overall process is shown in Algorithm 2.

Algorithm 2: Algorithm for Feature Extraction using Swin Transformer with Dynamic Cross-Attention Layer and Classification with LightGBM

Input: A 2D array of preprocess *I*(x, y) where each row represents a time sample
Output: *features = ExtractFeatures(I, p, D, L, Q, K, V, A)*
Function divide-patches (I_{rgb}, pat, D):
$P = DivideIntoPatches(I_{rgb}, pat)$;
For each P in I_{rgb} do
$X = linearily - EmbedPatches(P, D)$;
End
$return\ X$;
end
Function dynamic-cross-attention(X, Q, K, V, A):
$-\ X$: $Set\ of\ patch\ embeddings$
$-\ Q$: $Query\ matrix$
$-\ K$: $Key\ matrix$
$-\ V$: $Value\ matrix$
$-A$: $Dynamic\ adjustment\ matrix$
$attention\_weights\ =\ softmax(np.dot(np.dot(Q, K.T)\ +\ A, V))$;
$output_{embeddings}\ =\ np.dot(attention_{weights}, V)$;
return $output_{embeddings}$;
Function swin-transformer-block (x, L):
$-X$: $Set\ of\ embeddings$
− L: Number of layers in the Swin Transformer blocks
For each K in range (L) do
$X\ =\ multihead\_self\_attention(layer\_normalization(X))\ +\ X$;
If (k mode 2==0)
$X_{final}\ =\ shift\_partition(x)$;
$return\ X_{final}$;
end
Function classification- head (X_{final}):
$Output = LightGBMClassifier(X_{final})$;
$y =\ eye - related\ probability(output)$;
$Loss = CrossEntropyLoss(Output, y)$;
Return prediction ;
End
Function extract-features (*(I, p, D, L, Q, K, V, A)*:
$-P = DivideIntoPatches(Preprocessed\ Image, p)$
$-X = EmbedPatches(P, D)$
$-Output\ Embeddings = DynamicCrossAttention(X, Q, K, V, A)$
$-X_{final} = SwinTransformerBlocks(X, L)$
End
For each I image in training dataset (D) do
$-f_i = extract - features\ ((I, p, D, L, Q, K, V, A)$
$-l_i = label - features\ (f_i, c)$
End of algorithm

E. Multiclass Prediction using LightGBM

LightGBM (Light Gradient Boosting Machine) is a gradient-boosting framework that uses tree-based learning algorithms. The design prioritizes speed and efficiency, particularly in managing large-scale data. The algorithm employs a histogram-based method to speed up the training process and reduce memory usage. Here is a more formal mathematical representation of the LightGBM algorithm, focusing on its core components. Incorporating features extracted by a Swin Transformer into a LightGBM model for recognizing eye-related diseases involves a multi-step process that blends deep learning feature extraction with gradient-boosting machine learning techniques. The following is a high-level algorithm that outlines this hybrid approach, detailing how to leverage the strengths of both Swin Transformer for complex feature extraction from images and LightGBM for efficient classification based on those features.

We use these extracted characteristics, now high-dimensional vectors, and their labels to identify each image's eye illness. We create a new dataset from these pairs by simplifying visual information for machine learning algorithms. Next, train a LightGBM model on this fresh dataset. We picked LightGBM for its efficiency and efficacy in processing tabular data, including Swin Transformer-generated high-dimensional feature vectors. Training the LightGBM model on retrieved characteristics and labels helps the system identify complicated links between them and eye disorders. Before feature extraction, the dataset can be separated into training and validation sets to check that the model works well on both old and new data. This enables an assessment step to verify the model's capacity to generalize its learnt patterns to fresh data, confirming its real-world usefulness. Finally, the LightGBM model can detect eye disorders in new photos after training. These fresh photos are used to extract features using the same Swin Transformer model and then sent through the trained LightGBM model. The LightGBM model then classifies each collection of characteristics into an eye condition, identifying the diagnosis in the new image.

The Cross-Entropy Loss is a common loss function for multi-class classification problems, and this table lists the Swin-DCA and LightGBM hyperparameters needed to train and optimize the fundus image classification into Normal (NML), diabetic hypertension, diabetic retinopathy (DR), and others. A loss function for categorical outcomes is plausible in a Swin-DCA (Dynamic Cross-Attention) layer architecture multi-class classification environment. The Cross-Entropy Loss function is a standard choice for multi-class classification problems because it quantifies the difference between two probability distributions: the true distribution (the actual labels) and the predicted distribution (the outputs of the model).

Let y be the true distribution of the labels in a one-hot encoded form, where yi is 1 if the label is the ith class and 0 otherwise and let y' be the predicted distribution (the softmax output of the model), where yi is the predicted probability of the ith class. The Triplet Loss function is indeed a powerful tool for certain types of machine learning tasks such as transformers, particularly those involving learning embeddings or distances between examples, such as in different regions recognition compared to the weighted cross-entropy Loss or Focal Loss. It is defined as:

$$L(\mathrm{xa}, \mathrm{xp}, \mathrm{xn}) = \max\{\mathrm{d}(\mathrm{xa}, \mathrm{xp}) - \mathrm{d}(\mathrm{xa}, \mathrm{xn}) + margin, 0\} \quad (8)$$

Where: Anchor (xa): A reference example, Positive (xp): An example that is similar to the anchor, Negative (xn): An example that is different from the anchor and d(xa, xp) is the distance between the anchor and the positive sample and d(xa, xn) is the distance between the anchor and the negative sample. This Triplet Loss function approach harnesses the DL capabilities of the Swin Transformer to understand and capture the complex visual patterns in eye-related disease images and combines them with the machine learning process of LightGBM to classify these patterns into specific diseases. It's a powerful example of how combining different AI methodologies can create a more effective solution for complex problems like eye-related disease recognition.

Fine-tuning hyperparameters as described in Table II often involves conducting a grid search or random search over the hyperparameter space and evaluating the model's performance on a validation set.

TABLE II. FINE-TUNE OF DIFFERENT HYPERPARAMETERS FOR DEVELOPMENT OF IOT-OPTHOM-CAD SYSTEM

Hyperparameter	Swin Transformer	LightGBM Classifier
Number of Layers	24	-
Patch Size	4x4, 8x8, 16x16, 32x32	-
Embedding Dimension	224 x224	-
Learning Rate	0.01	-
Number of Trees	-	1000
Maximum Depth	-	8
Learning Rate	-	0.1
Regularization Parameter	-	0.1

IV. EXPERIMENTAL RESULTS

Six assessment methodologies are used to assess the effectiveness of the prediction: accuracy (ACC), specificity (SP), precision (P), recall (R), and F1-score (F). Using the PyTorch deep learning framework, we create the network. This study suggests utilizing retinal fundus pictures to identify eye problems with a planned 2-D IoT-Opthom-CAD. The Python code for implementing the proposed IoT-Opthom-CAD system is developed within a Google Colab environment, leveraging the computational resources provided by a GPU graphics card with 16 GB of memory. The system operates on a 64-bit Windows 10 system, running on an Intel (R) Core (TM) i7–43,450 CPU. TensorFlow serves as the primary framework for constructing and training deep learning models. To ensure uniformity across the dataset, all original images are resized to a consistent resolution of (224×224) pixels. This standardized dimension is widely recognized within the deep learning community as an optimal input size for various neural network architectures.

TensorFlow and Keras packages train the model in the Python 3.7.4 environment in Jupyter Notebook, utilizing a deep learning framework. Fundus photos are utilized as input data, which is subsequently enhanced using a variety of methodologies to address a range of potential real-world circumstances. In the ratio of 14:3:3, the enhanced dataset was divided into training, validation, and testing sets. The suggested model was trained and tested with different hyper-parameter settings. All augmented fundus images were first cleaned up and scaled to fit the training neural network's input dimensions. To properly analyze the prediction evaluation on unobserved data, IoT-Opthom-CAD's performance was compared to that of the current state-of-the-art deep learning models.

Three online datasets were used to train the IoT-Opthom-CAD system: MuReD (a multi-label dataset for retinal diseases) [38], BRSET (a Brazilian multi-label dataset for retina fundus photos) [39], and OIA-ODIR (an adaptive dataset for ophthalmic image analysis and disease recognition) [40]. We have collected initially 6,00 fundus images from these sources, including an average (NOM) of 1900, diabetic retinopathy (DR) of 2000, glaucoma (GLC) of 400, cataracts (CAT) of 200, age-related macular degeneration (AMD) of 300, and hypertension retinopathy (HR) of 1200 images. To balance the selected dataset, we have applied data augmentation and preprocessing techniques explained in Section 3.2 to convert 6,000 images into 12,000 retinographics. Given that the photos come from different sources, the resolution can vary from 520x520 to 3400x2800, depending on the source of the image. We have resized them to 224x224.

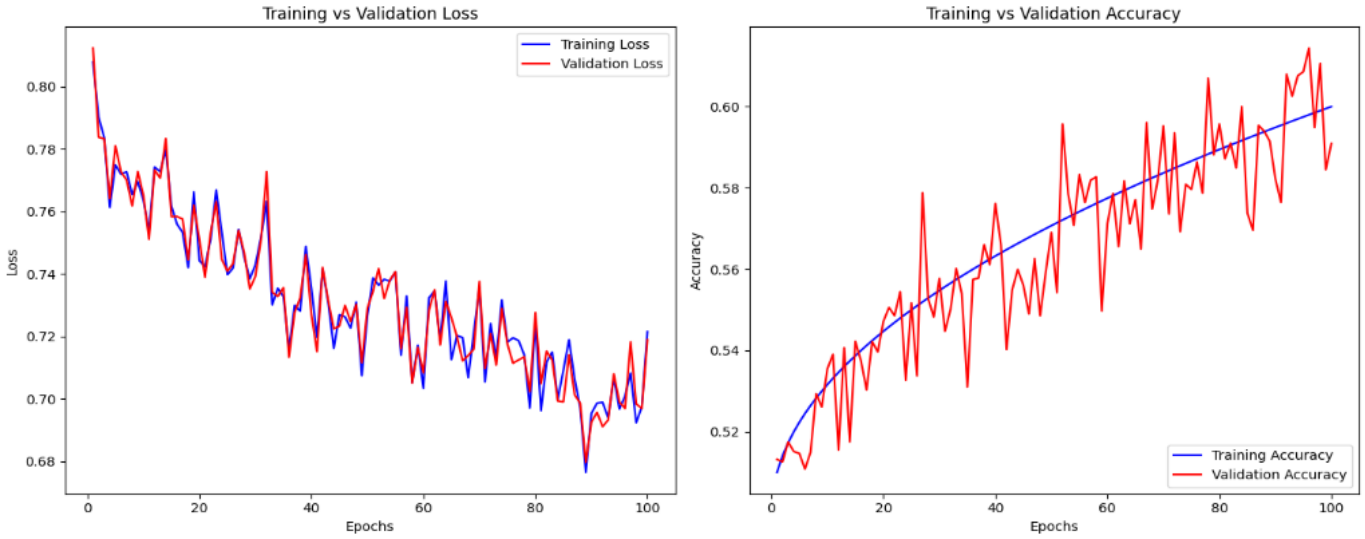

Fig. 8. Loss versus accuracy curves for training and validation with respect to epochs 100 for proposed IoT-Opthom-CAD system.

TABLE III. DIFFERENT PERFORMANCE METRICS WHEN APPLIED ON IOT-OPTHOM-CAD SYSTEM FOR RECOGNITION OF EYE-RELATED DISEASES WITH VARIOUS EXPERIMENTAL SETUP

Experiment ID	Alpha	Batch Size	Learning Rate Decay	Epochs	Acc(%)	P (%)	R (%)	F1-Score (%)	SP(%)
Exp1	0.001	32	0.1	100	95.0	96.0	95.0	95.0	94.11
Exp2	0.0001	64	0.05	100	94.5	95.0	94.0	94.5	93.11
Exp3	0.01	128	0.01	100	96.3	95.1	96.0	96.2	95.9

In Fig. 8, the loss and accuracy curves for the suggested IoT-Opthom-CAD system show good features, showing that the classifier is not overfitting or underfitting. Table III presents the results of the IoT-Opthom-CAD system's performance in recognizing eye-related diseases under various experimental setups. Each experiment has a distinct ID and multiple configurations, such as the alpha value, batch size, learning rate decay, and number of epochs. The metrics evaluated include accuracy (Acc), precision (P), recall (R), F1-score, and specificity (SP). The first experiment had an alpha value of 0.001, a batch size 32, a learning rate decay of 0.1, and 100 training epochs. It got an accuracy of 95.11%, with 95.11% for specificity and 95.10% for precision, recall, and F1-score.

Experiment 2, with an alpha of 0.0001, a 64-batch size, a 0.05 learning rate decay, and 100 epochs, showed the system's flexibility. Precision, recall, and F1-score were 95.0%, 94.0%, and 94.5%, respectively, while specificity was 93.11%. Accuracy fell to 94.5%. Experiment 3, with an alpha value of 0.01, a bigger batch size of 128, a lower learning rate decay of 0.01, and the same number of epochs, showed the system's illness recognition accuracy optimization. This setup produced 96.3% accuracy, 95.1% precision, 96.0% recall, and 96.2% F1-score, and 95.9% specificity. These results reveal that experimental configurations greatly impact IoT-Opthom-CAD performance indicators. They also illustrate how tweaking parameters improves illness identification accuracy.

Table IV compares models' eye-related illness recognition metrics. Three configurations are tested: the standard Swin Transformer model with Softmax activation and LightGBM.

Classification, a separable CNN model with Softmax activation, and the proposed architecture with dynamic cross-attention and LightGBM. Each model reports accuracy (ACC), precision (P), recall (R), F1-score, and specificity (SP) for average (NML), diabetic retinopathy (DR), tensional suspense neuropathy (TSN), age-related macular degeneration (ARMD), ocular degeneration (ODE), and high-risk diseases. The Base Swin Transformer model with Softmax activation and LightGBM classification performs 90.0% across all parameters for all illness categories. The separable CNN model with Softmax activation performs poorly in most tests, with accuracy, precision, recall, F1-score, and specificity ranging from 87.0% to 89.0% for various diseases. But adding dynamic cross-attention and LightGBM to the suggested Swin Transformer architecture improves its performance, notably in identifying ARMD and HR with 100% and 97.5% accuracy, respectively.

Fig. 10 displays the confusion matrix of the proposed IoT-Opthom-CAD system, which is used to diagnose various eye illnesses. This information is vital for evaluating the system's overall performance and finding areas for enhancement in illness identification. The comprehensive measure of confusion for the proposed system is displayed in Fig. 9 and Fig. 10. In addition, we conducted a computational efficiency analysis of a Swin Transformer paired with LightGBM on several hardware platforms, including CPU, GPU, and TPU. The results are presented in Table V.

TABLE IV. COMPARISONS OF BASIC SWIN TRANSFORMERS AND PROPOSED SWING AND DYNAMIC CROSS ATTENTION ARCHITECTURE ON SELECTED DATASET FOR RECOGNITION OF VARIOUS EYE-RELATED DISEASES

Model Configuration	*Metric	NML	DR	TSN	ARMD	ODE	HR	Overall
Base Swin +Softmax+ LightGBM	ACC (%)	90.0	90.0	90.0	90.0	90.0	90.0	90.0
	P (%)	89.0	89.0	89.0	89.0	89.0	89.0	89.0
	R (%)	89.0	89.0	89.0	89.0	89.0	89.0	89.0
	F1 (%)	90.5	90.5	90.5	90.5	90.5	90.5	90.5
	SP(%)	90.0	90.0	90.0	90.0	90.0	90.0	90.0
Separable CNN + Softmax	ACC (%)	89.0	89.0	89.0	89.0	89.0	89.0	89.0
	P (%)	88.5	88.5	88.5	88.5	88.5	88.5	88.5
	R (%)	86.0	86.0	86.0	86.0	86.0	86.0	86.0
	F1 (%)	88.1	88.1	88.1	88.1	88.1	88.1	88.1
	SP(%)	87.0	87.0	87.0	87.0	87.0	87.0	87.0
Swin+ Dynamic cross attention+ LightGBM	ACC (%)	94.0	96.5	95.5	100.0	94.5	97.5	96.3
	P (%)	93.5	95.5	94.5	97.5	93.5	96.5	95.1
	R (%)	94.6	95.2	96.6	98.6	95.6	95.6	96.0
	F1	94.5	94.8	95.5	99.5	95.5	97.5	96.2
	SP(%)	94.6	95.6	94.6	98.6	95.6	96.6	95.9

* Acc: Accuracy, P:Precision, R: Recall, SP: Specificity, AUC: Area under the receiver operating curve

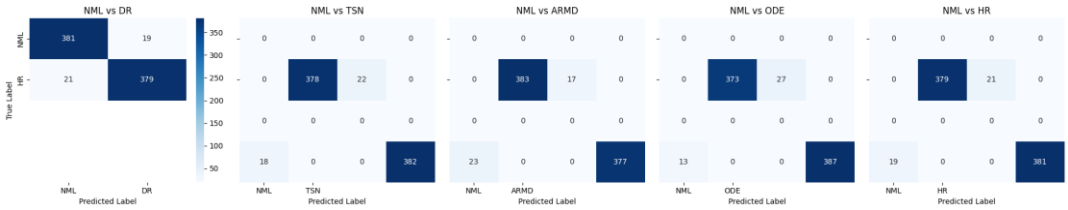

Fig. 9. Confusion metrics of proposed IoT-Opthom-CAD system for recognition of various eye-related diseases such as diabetic retinopathy (DR), Tessellation (TSN), Age-related macular degeneration (ARM), Optic disc edema (ODE), and hypertensive retinopathy (HR) compare with normal (NML).

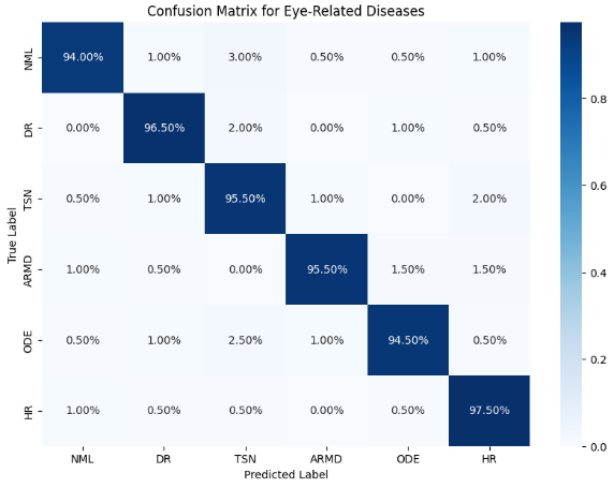

Fig. 10. Overall confusion metrics that indicates the model's performance in distinguishing between the presence and absence of the disease.

TABLE V. COMPUTATIONAL COMPARISONS WITH DIFFERENT ARCHITECTURE FOR PROPOSED IOT-OPTHOM-CAD SYSTEM

Hardware	Training Time (minutes)	Inference Time (ms/image)	Terms
CPU	48	500	Standard multi-core CPU setup
GPU	8	50	High-end gaming or professional GPU
TPU	4	20	Google Cloud TPU v3

TABLE VI. STATE-OF-THE-ART COMPARISONS ON SELECTED DATASETS WITH SAME PREPROCESSING

Model	Learning Rate	Batch Size	Epochs	Optimizer	Activation Function
Fahdawi-2024 [30]	0.001	32	50	Adam	ReLU
Sengar-2023 [31]	0.01	64	100	SGD	Tanh
Triwijoyo-2020 [32]	0.0001	128	80	RMSprop	Leaky ReLU
Opthom-CAD	0.01	64	100	Adam	ReLU

TABLE VII. AN EXAMPLE TABLE OUTLINING THE EXPERIMENTAL HYPER-PARAMETER SETUP FOR THE COMPARISONS

Model	*Acc	*P	*R	*SP	*AUC
Fahdawi-2024 (DRBM) [30]	86.0%	87.0%	86.0%	88.0%	0.875
Sengar-2023 (DNN) [31]	84.0%	86.0%	85.0%	86.0%	0.845
Triwijoyo-2020 (CNN) [32]	83.0%	86.0%	85.0%	86.0%	0.835
IoT-Opthom-CAD (Proposed)	95.16%	96.5%	95.08%	95.93%	0.95

* Acc: Accuracy, P:Precision, R: Recall, SP: Specificity, AUC: Area under the receiver operating curve

Table VI provides a comparison of various models' performance metrics on selected datasets. These hyper-parameters include the learning rate, batch size, number of epochs, optimizer, and activation function used for training each model. They are essential set-tings that influence the training process and ultimately impact the model's performance and convergence. Adjusting these parameters optimally is crucial to achieving the desired results and ensuring the effectiveness of the trained models. Table VII comprehensively compares state-of-the-art models applied to selected datasets, employing identical preprocessing and data augmentation techniques. Each model's performance is evaluated across multiple metrics to gauge its efficacy in recognizing eye-related diseases. The Fahdawi-2024 (DRBM) model achieves an accuracy of 86.0%, demonstrating commendable precision, recall, specificity, and AUC values of 87.0%, 86.0%, 88.0%, and 0.875, respectively. Similarly, the Sengar-2023 (DNN) model attains an accuracy of 84.0%, with precision, recall, specificity, and AUC values of 86.0%, 85.0%, 86.0%, and 0.845, respectively. Meanwhile, the Triwijoyo-2020 (CNN) model achieves an accuracy of 83.0%, coupled with precision, recall, specificity, and AUC values of 86.0%, 85.0%, 86.0%, and 0.835, respectively. In contrast, the Opthom-CAD (proposed) system outperforms its counterparts with remarkable accuracy, achieving an impressive 95.16%. This superiority extends across all metrics, with precision, recall, specificity, and AUC values at 96.5%, 95.08%, 95.93%, and 0.95, respectively. Such exceptional performance underscores the effectiveness of the proposed Opthom-CAD system in accurately identifying various eye-related diseases. The proposed model's significantly higher accuracy and robustness highlight its potential to revolutionize disease detection in ophthalmology, offering promising avenues for improved patient care and management.

In complicated tasks like image classification, natural language processing, and predictive analytics, xAI interpretability involves understanding and explaining AI (xAI) model decisions and behavior. In visually explaining models, interpretability entails offering intuitive and meaningful representations of how the model predicts or classifies. Gradient-based approaches like Gradient-weighted Class Activation Mapping (Grad-CAM) provide output gradients considering input attributes. This shows how Grad-CAM is used to graphically illustrate the model's predictions using AI. Computing the target class gradients on the final convolutional layer's convolutional feature maps emphasizes the input image's most important regions for the projected class. Model judgments are easier to comprehend with this method. For the model's judgment, input pixels or characteristics matter most. Visual explanations of AI models help users understand how they reach their conclusions, build trust in AI systems, identify biases and errors, and collaborate with human experts in various fields.

The Swin Transformer architecture extracts Fig. 11 characteristics from colored fundus pictures. Hierarchical transformer layers capture long-range visual dependencies in Swin Transformer, a current computer vision technique. For reliable eye illness diagnosis, the model rapidly extracts local and global characteristics from retinal pictures using Swin Transformer. LightGBM is a gradient-boosting framework that is employed for multi-label classification. It works by iteratively training weak learners on the residuals of the previous iteration, gradually improving the model's predictive performance. This is where LightGBM comes in handy: it sorts the extracted features into groups of eye diseases, like normal, diabetic retinal disease, tessellation, age-related macular degeneration, optic disc edema, and hypertensive retinal disease as shown in Fig. 12.

The smartphone-based system captures high-quality fundus images using its built-in camera or an attached IoT head-mounted camera (IoT headset). These images are then uploaded to the cloud for further processing. The smartphone application can act as an intermediary, facilitating the transfer of data from the patient to the cloud. Features extraction algorithm is running on the cloud servers identify and isolate relevant regions of interest within the fundus images. The DL models classify the images based on the extracted features, determining the presence of eye-related disease. Patients with eye-related concerns used the online mobile computing device, and their information was recorded by healthcare workers. A dedicated application was downloaded onto their mobile devices, which facilitated capturing and analyzing eye-related disease data via the cloud.

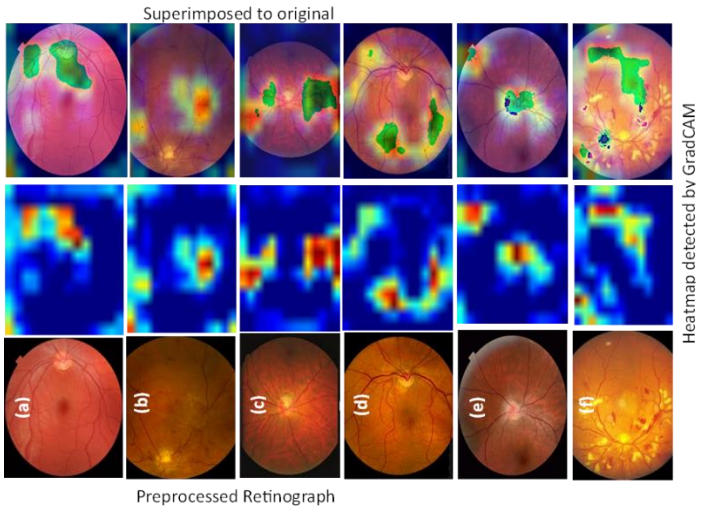

Fig. 11. A visual diagram of AI interpretable using Grad-CAM , where figure (a) Shows the normal image, (b) Shows the diabetic retinopathy, (c) Demonstrates the tessellation, (d) Presents age-related macular degeneration (ARMD), (e) Shows the optic disc edema (ODE) Image, and (f) Presents the hypertensive retinopathy (HR) image.

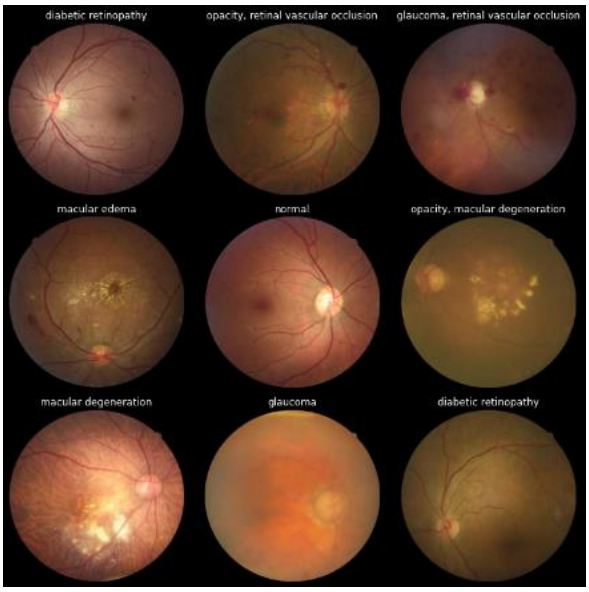

Fig. 12. A visual diagram of proposed IoT-Opthom-CAD system for multiclass eye diseases using dynamic swin transformers and explainable artificial intelligence.

The IoT network operates through three primary layers: the data link layer, the network layer, and the application layer. The data link layer starts with a dataset of fundus images obtained from patient records, primarily used for analysis. This layer utilizes the transport layer for processing and evaluated using multi-label retinal disease datasets like MuReD, BRSET, and OIA-ODIR. These datasets are used for testing purposes. This dataset consists of around 2000 color fundus images of each category with annotations provided in an Excel file. The network, or transport layer, includes a cloud server network designed to host applications. It facilitates data transmission between tools and minimizes delay times. Additionally, it enables users to monitor patient details stored in databases. The application layer features an integrated with Python programming to analyze fundus images. This layer allows patients to upload their fundus images for analysis. The application includes the disease diagnosis model, further detailed in the following sections. The smartphone application provides a user-friendly interface for patients to easily capture and upload images, view results, and receive notifications.

An ablation study for combining Swin Transformer and LightGBM to recognize various eye-related diseases could involve systematically varying model configurations and training parameters to observe their impact on the classification performance. This study can help in understanding the contribution of different components and settings to the model's overall effectiveness. Table VIII outlining an ablation study for this purpose. Note that the performance metrics (e.g., accuracy) are illustrative and not based on actual experimental results.

The Table VIII shows the results of an ablation study that looks at how different setups of the Swin Transformer and LightGBM models impact the ability to detect several eye diseases, including Normal (NML), Diabetic Retinopathy (DR), Tessellation (TSN), Age-Related Macular Degeneration (ARMD), Optic Disc Edema (ODE), and Hypertensive Retinopathy (HR). The study systematically alters model configurations and assesses their effects on classification accuracy, providing insights into how different aspects of the models influence performance.

Experiment ID 1 serves as the baseline, employing base configurations for both the Swin Transformer and LightGBM across all diseases, achieving 90.0% accuracy. This setup establishes a reference point for comparison with subsequent experiments.

Experiment ID 2 tests the impact of a shallower Swin Transformer while keeping the LightGBM configuration unchanged. The reduction in depth leads to a slight de-crease in accuracy to 87.5%, suggesting that depth contributes significantly to capturing the complex features necessary for accurate classification.

Experiment ID 3 explores the effect of simplifying LightGBM trees by reducing the number of leaves, with the Swin Transformer configuration held constant. The result is a minor drop in accuracy to 89.0%, indicating that a more complex tree structure might be beneficial but is less critical than the depth of the Swin Transformer.

Experiment ID 4 increases the embedding dimension of the Swin Transformer. This adjustment leads to a higher accuracy of 91.2%, showing that a richer feature representation enhances model performance.

Experiment ID 5 adds a shifted window mechanism to the base Swin Transformer configuration, slightly improving accuracy to 90.5%. This suggests that enabling cross-window connections helps capture more contextual information, which is beneficial for classification.

Experiment ID 6 focuses the evaluation on normal and diabetic retinopathy cases specifically, maintaining base configurations for both models. A notable increase in accuracy to 92.0% indicates that the models are particularly effective at distinguishing between these two conditions.

Experiment ID 7 shifts focus to the remaining diseases (TSN, ARMD, ODE, and HR), resulting in an accuracy of 88.5%. Compared to Experiment 6, this lower performance suggests that these conditions present more challenging or subtle features to classify accurately.

Experiment ID 8 investigates the impact of data augmentation on the base configuration, leading to a significant accuracy increase to 93.0%. This underscores the value of augmentation in enhancing the model's generalization capabilities.

Experiment ID 9 examines the effect of increasing the complexity of LightGBM trees by a more significant number of leaves, achieving an accuracy of 90.8%. This indicates that a more nuanced decision-making process can marginally improve classification outcomes.

Experiment ID 10 evaluates the use of higher-resolution images with the base model configurations, achieving 91.5% accuracy. The improvement suggests that high-er-resolution inputs provide more detailed information for feature extraction, aiding disease classification.

Overall, the ablation study shows how vital model depth, embedding dimensionality, data augmentation, and input resolution are for improving the accuracy of disease classification in the eye. While adjustments to the LightGBM configuration also affect performance, modifications to the Swin Transformer architecture, particularly those that enhance feature representation and extraction, appear to have a more pronounced impact on the model's effectiveness.

An ablation study combining Swin Transformer and LightGBM for recognizing various eye-related diseases such as Normal (NML), Diabetic Retinopathy (DR), Tessellation (TSN), Age-related Macular Degeneration (ARMD), Optic Disc Edema (ODE), and Hypertensive Retinopathy (HR). Ablation studies are critical for understanding the contribution of each component or parameter to a model's performance. Imagine a table that systematically varies the parameters and configurations of the Swin Transformer and LightGBM models to evaluate their impact on the classification accuracy for these eye conditions. Each row of the table would represent a different experimental setup, altering aspects such as the depth of the Swin Transformer, the number of heads in multi-head self-attention, the size of the input images, or specific hyperparameters of LightGBM like the number of leaves, learning rate, and the depth of trees.

For the Swin Transformer, one experiment might vary the patch size, analyzing how granularity affects the model's ability to capture relevant features for disease classification. A smaller patch size could improve the model's sensitivity to finer details critical for distinguishing between diseases like TSN and DR, which may exhibit subtle differences in retinal images. Another row might explore the depth of the swing transformer, adjusting the number of transformer blocks. More layers allow for more complex feature hierarchies, possibly improving differentiation between complex conditions like ARMD and ODE but also increasing computational costs and the risk of overfitting. On the LightGBM side, one could manipulate the learning rate to see how faster or slower convergence affects model performance across the different diseases. A lower learning rate might lead to more robust learning with less risk of overlooking subtle features distinguishing NML from early stages of diseases like DR or HR.

Another variation could involve the number of leaves in LightGBM, investigating the trade-off between model complexity and the risk of overfitting. More leaves allow the model to make finer distinctions, potentially improving its ability to classify dis-eases with overlapping symptoms. However, they might also capture noise in the data, leading to poor generalization. The results section of this table would detail the classification accuracy for each disease under different experimental setups, providing in-sights into which configurations yield the best balance of sensitivity and specificity across conditions. For instance, one might find that moderate patch size and depth in the Swin Transformer, combined with a careful balance of learning rate and tree complexity in LightGBM, offer the most effective performance across all conditions, high-lighting the importance of each parameter in capturing the nuanced differences be-tween these eye diseases. This kind of ablation study would be very helpful for im-proving the combined Swin Transformer and LightGBM method. It would show re-searchers the best ways to set up these devices to diagnose a wide range of eye conditions. Through systematic experimentation and analysis, one could derive a highly optimized model setup that leverages the strengths of both deep learning and gradient-boosting techniques for enhanced medical imaging analysis.

These limitations are provided in Table IX as a critical perspective on areas where the proposed system might face challenges or require further development and validation.

The influence of the parameters used in the IoT-Opthom-CAD system can significantly impact the performance metrics and the effectiveness of the model in recognizing eye-related diseases as shown in Table X. The parameters used in the IoT-Opthom-CAD system play a crucial role in determining the model's effectiveness. By carefully tuning these parameters, the

model's performance can be optimized, leading to more accurate and reliable predictions. Future work can explore the effects of these parameters in more detail, ensuring that the model is both robust and generalizable across different datasets and real-world scenarios. However according to our limited knowledge, we did not find a single study for classification of multi-class eye-diseases using IoT-enable devices.

TABLE VIII. VARIOUS EXPERIMENTS OF DIFFERENT EXPERIMENTAL SETTINGS FOR PROPOSED IOT-OPTHOM-CAD SYSTEM

Experiment ID	Swin Transformer Config	LightGBM Config	Evaluated Diseases	Accuracy (%)	Explains
1	Base config	Base config	All	90.0	Baseline for comparison
2	Reduced depth	Base config	All	87.5	Tests impact of shallower Swin Transformer
3	Base config	Reduced num_leaves	All	89.0	Impact of simpler LightGBM trees
4	Increased embed-dim	Base config	All	91.2	Higher dimensionality for embeddings
5	Base config + Shifted window	Base config	All	90.5	Shifted window impact
6	Base config	Base config	NML, DR	92.0	Focused on NML and DR
7	Base config	Base config	TSN, ARMD, ODE, HR	88.5	Focused on TSN, ARMD, ODE, HR
8	Base config with augmentation	Base config	All	93.0	Data augmentation impact
9	Base config	Increased num_leaves	All	90.8	More complex LightGBM trees
10	High-resolution images	Base config	All	91.5	Tests impact of using higher resolution images

TABLE IX. LIMITATIONS OF THE PROPOSED IOT-OPTHOM-CAD SYSTEM FOR MULTICLASS RETINAL EYE DISEASES

Limitation Description	Impact
Limited Dataset Diversity	The study uses three specific datasets (MuReD, BRSET, and OIA-ODIR), which might not cover all possible variations of retinal images in a real-world scenario.
Generalization to Different Populations	The model's performance might vary when applied to populations with different demographic characteristics than those represented in the datasets used.
Dependence on High-Quality Images	The accuracy of the system relies on the quality of retinal images; lower-quality images could affect diagnostic performance.
Explainability and Interpretability Challenges	While Grad-CAM is used for explainability, the complexity of the model might still pose challenges for clinicians to fully understand the decision-making process.
Potential Overfitting Due to Data Augmentation	Extensive data augmentation might lead to overfitting, where the model performs well on the training data but poorly on unseen data.
Scalability and Integration into Existing Clinical Workflows	Integrating the IoT-Opthom-CAD system into existing clinical workflows and ensuring its scalability in diverse healthcare settings might be challenging.
Future Adaptability to New Retinal Diseases	The system is designed for specific diseases; adapting it to recognize new or less common retinal diseases could require significant modifications and retraining.

TABLE X. INFLUENCE OF THE PARAMETERS USED IN THE IOT-OPTHOM-CAD SYSTEM

Parameter	Description	Influence on Model Performance
Alpha Value (Learning Rate)	Controls how much to change the model in response to the estimated error each time the model weights are updated.	A lower learning rate (alpha) can lead to more precise adjustments but requires more epochs to converge. A higher learning rate can speed up training but may overshoot the optimal solution.
Batch Size	Number of training samples used in one iteration.	A smaller batch size provides a more accurate estimate of the gradient, leading to a more stable learning process but may slow down training. A larger batch size can speed up training but might lead to less accurate updates.
Learning Rate Decay	Gradually decreases the learning rate during training.	Helps in fine-tuning the learning process, ensuring the model doesn't overshoot the optimal weights, leading to better convergence.
Number of Epochs	Number of times the entire training dataset passes through the neural network.	More epochs can lead to better training and fine-tuning of the model, but too many can cause overfitting.
Resolution of Input Images	Size to which input images are resized.	Consistent resolution (224x224) ensures uniformity in training, which is critical for deep learning models to learn effectively. Larger images may capture more details but require more computational resources.
Data Augmentation Techniques	Techniques used to artificially increase the size of the training dataset.	Enhances the model's ability to generalize by providing a variety of training samples, reducing overfitting and improving robustness.
Cross-Dataset Validation	Using different datasets to validate the model.	Helps in testing the generalization capability of the model across diverse sets of data.
Model Architecture (Swin Transformers with Dynamic Cross-Attention + LightGBM)	Combination of different model architectures and algorithms.	Enhances feature extraction (local and global features) and improves classification performance by leveraging advanced architectures.
Grad-CAM	Explainable AI technique to visualize the areas in the image that the model focuses on.	Improves interpretability and trust in the model by showing which parts of the image contribute to the decision-making process.
GPU/TPU Utilization	Hardware used for training and inference.	High-end GPUs/TPUs speed up training and inference, making it feasible to train more complex models or use larger datasets.

V. CONCLUSION

The paper introduces a novel computer-aided diagnosis (CAD) system called IoT-Opthom-CAD, explicitly designed for identifying various eye diseases from colored fundus images. Additionally, the integration of IoT devices enhances real-time, remote monitoring and diagnosis capabilities, providing continuous and intelligent analysis of eye-related diseases. This feature is crucial for early and accurate classification of multiclass eye diseases, significantly impacting patient outcomes. IoT-Opthom-CAD uses the Gradient Boosting (LightGBM) method and lightweight deep learning-based Swin transformers to extract and classify features, effectively. It incorporates a dynamic cross-attention layer (DCA-L) for extracting local and global features. The system is evaluated using multi-label retinal disease datasets like MuReD, BRSET, and OIA-ODIR. Results from 10-fold cross-validation tests indicate impressive performance, with up to 95.0% accuracy, 97% sensitivity, 96% specificity, and an AUC of 0.95. The IoT-Opthom-CAD system surpasses many state-of-the-art models, indicating its excellence in identifying eye-related disorders. The exceptional precision and responsiveness of IoT-Opthom-CAD demonstrate its capacity to aid ophthalmologists in properly and swiftly detecting a range of eye ailments.

Potential areas for future study are expanding the dataset to encompass a wider range of fundus pictures in order to enhance the flexibility and dependability of the system. In addition, doing research on alternative deep learning frameworks, examining novel attention processes, and optimizing hyper-parameters might enhance the diagnostic accuracy of the system. Validating the usefulness and feasibility of using IoT-Opthom-CAD in clinical situations and conducting forward-looking research will facilitate its incorporation into routine medical procedures, ensuring its suitability for everyday usage.

FUNDING

This work was supported and funded by the Deanship of Scientific Research at Imam Mohammad Ibn Saud Islamic University (grant number IMSIU-RG23129).

Data Availability Statement:

1) Retinal Fundus Multi-Disease Image Dataset (MuReD) [38]: https://ieee-dataport.org/open-access/retinal-fundus-multi-disease-image-dataset-rfmid (access date 1st January 2024).

2) Brazilian Multilabel Ophthalmological Dataset of Retina Fundus (BRSET) [39]: https://physionet.org/content/brazilian-ophthalmological/1.0.0/ (access date 1st January 2024).

3) Ophthalmic image analysis-ocular disease intelligent recognition (OIA-ODIR) dataset [40]: https://github.com/nkicsl/OIA-ODIR (access date 1st January 2024).

4) Source Code Availability: The source code is public and accessible for anyone to view and modify from GitHub (https://github.com/Qaisar256/Opthom-CAD).

ACKNOWLEDGMENT

This work was supported and funded by the Deanship of Scientific Research at Imam Mohammad Ibn Saud Islamic University (grant number IMSIU-RG23129).

CONFLICTS OF INTEREST

The author declares that there are no conflicts of interest.

REFERENCES

[1] J. Sigut, F. Fumero, J. Estévez, S. Alayón, T. Díaz-Alemán, "In-Depth Evaluation of Saliency Maps for Interpreting Convolutional Neural Network Decisions in the Diagnosis of Glaucoma Based on Fundus Imaging," Sensors, vol. 24, 1-20, 2024.

[2] W. Chen, R. Li, Q. Yu, A. Xu, Y. Feng et al., "Early detection of visual impairment in young children using a smartphone-based deep learning system," Nature Medicine, vol. 29, no. 2, pp. 493-503, 2023.

[3] L.J. Coan, B.M. Williams, V.K. Adithya, S. Upadhyaya, A. Alkafri et al., "Automatic detection of glaucoma via fundus imaging and artificial intelligence: A review," Survey of Ophthalmology, vol. 68, no. 1, pp. 17-41, 2023.

[4] L. Shao, X. Zhang, T. Hu, Y. Chen, C. Zhang et al., "Prediction of the fundus tessellation severity with machine learning methods," Frontiers in Medicine, vol. 9, pp. 1-22, 2022.

[5] R. Shi, X. Leng, Y. Wu, S. Zhu, X. Cai et al., "Machine learning regression algorithms to predict short-term efficacy after anti-VEGF treatment in diabetic macular edema based on real-world data," Scientific Reports, vol. 13, pp. 18-46, 2023.

[6] G. Corradetti, A. Verma, J. Tojjar, L. Almidani, D. Oncel et al., "Retinal Imaging Findings in Inherited Retinal Diseases," J. Clin. Med., vol. 13, pp. 38-50, 2024.

[7] A. Bali and V. Mansotra, "Analysis of deep learning techniques for prediction of eye diseases: A systematic review," Arch Computat Methods Eng, vol. 31, pp. 487-520, 2024.

[8] Preity, A.K. Bhandari and S. Shahnawazuddin, "Automated computationally intelligent methods for ocular vessel segmentation and disease detection: a review," Archives of Computational Methods in Engineering, vol. 31, no. 2, pp. 701-724, 2024.

[9] K. Shankar, E. Perumal, M. Elhoseny, and P.T. Nguyen, "An IoT-Cloud based intelligent computer-aided diagnosis of diabetic retinopathy stage classification using deep learning approach," Comput. Mater. Contin., vol. 66, no. 2, pp. 1665-1680. 2021.

[10] Kumar, Yogesh, and B. Gupta, "Retinal image blood vessel classification using hybrid deep learning in cataract diseased fundus images," Biomedical Signal Processing and Control, vol. 84, pp.1-15, 2023.

[11] A. Elsawy, T.D.L. Keenan, Q. Chen, A.T. Thavikulwat, S. Bhandari et al., "A deep network DeepOpacityNet for detection of cataracts from color fundus photographs," Commun Med, vol. 3, no. 184, pp.1-11, 2023.

[12] A.A. Salam, M. Mahadevappa, A. Das and M.S. Nair, "RDD-Net: retinal disease diagnosis network: a computer-aided diagnosis technique using graph learning and feature descriptors," The Visual Computer, vol. 39, no. 10, pp. 4657-4670, 2023.

[13] Y. Asiri, H.T. Halawani, A.D. Algarni, and A.A. Alanazi, "IoT enabled healthcare environment using intelligent deep learning enabled skin lesion diagnosis model," Alexandria Engineering Journal, vol. 78, pp. 35-44, 2023.

[14] R.K. Shinde, M.S. Alam, M.B. Hossain, S. Md. Imtiaz, J.H. Kim et al., "Squeeze-mnet: Precise skin cancer detection model for low computing IOT devices using transfer learning," Cancers, vol. 15, no. 1, pp.1-12, 2022.

[15] M. Obayya, M.A. Arasi, N.S. Almalki, S.S. Alotaibi, M.A. Sadig et al., "Internet of things-assisted smart skin cancer detection using metaheuristics with deep learning model," Cancers, vol. 15, no. 20, pp. 1-20, 2023.

[16] A.C. Scanzera, C. Beversluis, A.V. Potharazu, P. Bai, A. Leifer et al., "Planning an artificial intelligence diabetic retinopathy screening

program: a human-centered design approach," Frontiers in Medicine, vol. 10, pp.1-20, 2023.

[17] Y. Chai, H. Liu, and J. Xu, "Glaucoma diagnosis based on both hidden features and domain knowledge through deep learning models," Knowledge-Based Systems, vol. 161, pp. 147-156, 2018.

[18] S. Phene, R.C. Dunn, N. Hammel, Y. Liu, J. Krause et al., "Deep learning and glaucoma specialists: the relative importance of optic disc features to predict glaucoma referral in fundus photographs," Ophthalmology, vol. 126, no. 12, pp. 1627-1639, 2019.

[19] M. Juneja, S. Singh, N. Agarwal, S. Bali, S. Gupta et al., "Automated detection of Glaucoma using deep learning convolution network (G-net)," Multimedia Tools and Applications, vol. 79, pp. 15531-15553, 2020.

[20] M.H. Ibrahim, M. Hacibeyoglu, A. Agaoglu, and F. Ucar, "Glaucoma disease diagnosis with an artificial algae-based deep learning algorithm," Medical & Biological Engineering & Computing, vol. 60, no. 3, pp. 785-796, 2022.

[21] J.K.P.S. Yadav, and S Yadav, "Computer-aided diagnosis of cataract severity using retinal fundus images and deep learning," Computational Intelligence, vol. 38, no. 4, pp. 1450-1473, 2022.

[22] M.S. Junayed, M.B. Islam, A. Sadeghzadeh, and S. Rahman, "CataractNet: An automated cataract detection system using deep learning for fundus images," IEEE Access, vol. 9, pp. 128799-128808, 2021.

[23] H. Zhang, K. Niu, Y. Xiong, W. Yang, Z.Q. He et al., "Automatic cataract grading methods based on deep learning," Computer methods and programs in biomedicine," vol. 182, pp. 1-20, 2019.

[24] T. Pratap, and P. Kokil, "Deep neural network based robust computer-aided cataract diagnosis system using fundus retinal images," Biomedical Signal Processing and Control, vol. 70, pp. 1-20, 2021.

[25] K. Pammi, and P. Saxena, "Cataract detection and visualization based on multi-scale deep features by RINet tuned with cyclic learning rate hyperparameter," Biomedical Signal Processing and Control, vol. 87, pp. 1-12, 2024.

[26] A.A. Abd El-Khalek, H.M. Balaha, N.S. Alghamdi, M. Ghazal, A. Khalil et al., "A concentrated machine learning-based classification system for age-related macular degeneration (AMD) diagnosis using fundus images," Scientific Reports, vol. 14, pp. 1-14, 2024.

[27] M.A. Ali, M.S. Hossain, M.K. Hossain, S.S. Sikder, S.A. Khushbu et al., "AMDNet23: Hybrid CNN-LSTM deep learning approach with enhanced preprocessing for age-related macular degeneration (AMD) detection," Intelligent Systems with Applications, vol. 21, pp. 1-20, 2024.

[28] K. Xu, S. Huang, Z. Yang, Y. Zhang, Y. Fang et al., "Automatic detection and differential diagnosis of age-related macular degeneration from color fundus photographs using deep learning with hierarchical vision transformer," Computers in Biology and Medicine, vol. 167, pp. 20-40, 2023.

[29] J. Morano, Á.S. Hervella, J. Rouco, J. Novo, I.F.V. José et al., "Weakly-supervised detection of AMD-related lesions in color fundus images using explainable deep learning," Computer Methods and Programs in Biomedicine, vol. 229, pp. 1-30, 2023.

[30] S.A. Fahdawi, A.S. Al-Waisy, D.Q. Zeebaree, R. Qahwaji, H. Natiq et al., "Fundus-deepnet: Multi-label deep learning classification system for enhanced detection of multiple ocular diseases through data fusion of fundus images," Information Fusion, vol. 102, pp. 1-20, 2024.

[31] N. Sengar, R.C. Joshi, M.K. Dutta, and R. Burget, "EyeDeep-Net: a multi-class diagnosis of retinal diseases using deep neural network," Neural Comput & Applic, vol. 35, pp. 10551-10571, 2023.

[32] B.K. Triwijoyo, B.S. Sabarguna, W. Budiharto, and E. Abdurachman, "Deep learning approach for classification of eye diseases based on color fundus images," Diabetes and fundus OCT, vol. 1, pp. 25-57, 2020.

[33] J. Son, J.Y. Shin, H.D. Kim, K.H. Jung, K.H. Park et al., "Development and validation of deep learning models for screening multiple abnormal findings in retinal fundus images," Ophthalmology, vol. 127, no. 1, pp. 85-94, 2020.

[34] R. Sarki, K. Ahmed, H. Wang, and Y. Zhang, "Automated detection of mild and multi-class diabetic eye diseases using deep learning," Health Inf Sci Syst, vol. 8, no. 32, pp. 1-20, 2020.

[35] T. Nazir, A. Irtaza, A. Javed, H. Malik, D. Hussain et al., "Retinal image analysis for diabetes-Based eye disease detection using deep learning," Applied Sciences, vol. 10, pp. 1-21, 2020.

[36] L.P. Cen, J. Ji, J.W. Lin, S.T. Ju, H.J. Lin et al., "Automatic detection of 39 fundus diseases and conditions in retinal photographs using deep neural networks," Nature communications, vol. 12, no. 1, pp. 28-48, 2021.

[37] C. Guo, Y. Minzhong, and J. Li, "Prediction of different eye diseases based on fundus photography via deep transfer learning," Journal of Clinical Medicine, vol. 10, no. 23, 2021.

[38] M.A. Rodríguez, H. AlMarzouqi, and P. Liatsis, "Multi-label retinal disease classification using transformers," IEEE Journal of Biomedical and Health Informatics, pp. 2739-2750, 2022.

[39] L.F. Nakayama, M. Goncalves, L.Z. Ribeiro, H. Santos, D. Ferraz et al., "A Brazilian Multilabel Ophthalmological Dataset (BRSET) (version 1.0.0)," PhysioNet, 2023.

[40] N. Li, T. Li, C. Hu, K. Wang, H. Kang, "A Benchmark of ocular disease intelligent recognition: one shot for multi-disease detection," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics, vol. 23, pp. 177-193, 2021.