Retinal Disease Diagnosis Reimagined: A Deep Transfer Learning and Multi-Scale Attention Network Algorithm for Precision Ophthalmology

Ramanathan S
Department of Computer Science And Engineering,
Thiagarajar College of Engineering,
Madurai, Tamil Nadu, India
ramanathanswami12@gmail.com

Arun Kumar M
Department of Computer Science and Engineering,
Thiagarajar College of Engineering,
Madurai, Tamil Nadu, India
arunkumarnavamani@gmail.com

Susmanth Srinivas A
Department of Computer Science and Engineering,
Thiagarajar College of Engineering,
Madurai, Tamil Nadu, India
susmanth@student.tce.edu

Shanmukha Naveen K
Department of Information Technology,
SSN College of Engineering,
Chennai, Tamil Nadu, India
shanmukhanaveen2010809@ssn.edu.in

SenthilKumar C
Department of Computer Science and Engineering,
Thiagarajar College of Engineering,
Madurai, Tamil Nadu, India
cskcse@tce.edu

***Abstract*—Retinal diseases are critical in nature as they progress quite quickly and lead to catastrophic damages in retina and therefore leading to loss of vision. The availability of proper healthcare systems to all sections of the society remains uncertain. A retinal disease detection model is created to solve these issues. The Fundus dataset used here consists of 1548 exhaustive collection of images in 3 different classes. InceptionV3 was among the other pre-trained model that the VGG16 substantially surpassed with a recognition accuracy of 93.4%. The Custom Multi-Scale Attention Network model yielded a competent 91% accuracy score.**

***Index Terms*—Retinal Disease, Fundus dataset, Deep Learning, Pretrained Model, InceptionV3, VGG16, Custom MSAN, DME, DR, Healthy**

I. INTRODUCTION:

The delicate yet intricate set of parts that make up the human eye, a wonder of biological engineering, is necessary for us to comprehend the world around us. Among its various components, the retina plays a pivotal role in vision, transforming incoming light into electrical signals that are transmitted to the brain for interpretation. But this crucial and critical organ is not immune to illness, and when it is present with any disease, it can cause severe vision loss and even blindness in certain specific scenarios. Among the most prevalent and threatening of these diseases are [1] Diabetic Macular Edema (DME) and [2] Diabetic Retinopathy (DR), both of which are complications of diabetes mellitus.

The core region of the retina, the macula, which is important for clear, detailed vision, accumulated fluids in the event of diabetic macular edema. This disorder develops as a result of fluid leakage into the surrounding tissue caused by diabetes related damage to the blood vessels in the retina. If left untreated, DME can result in significant visual impairment, making early detection and intervention crucial for preserving sight.

Diabetic Retinopathy, nevertheless, is a more general phrase that conveys and encompasses a plethora of retinal complications resulting from diabetes. It can manifest as [3] Non-Proliferative Diabetic Retinopathy marked by the development of microaneurysms and hemorrhages in the retina, or as [4] Proliferative Diabetic Retinopathy (PDR) is a condition in which aberrant blood vessels form and have the potential to burst, severely impairing vision. To halt the course of DR and lessen its potentially dangerous effects on vision, early detection and prompt management are essential.

The prevalence of diabetes is on the rise globally, and with it, the incidence of DME and DR continues to escalate. Effective and efficient methods for diagnosing and predicting these conditions are thus of paramount importance in preventing vision impairment among those afflicted by diabetes.

This research paper delves into the utilization of advanced medical imaging techniques, specifically Eye Fundus and Macular Optical Coherence Tomography (OCT) images, to address the challenges posed by DME and DR. By leveraging the power of multiscale attention networks and transfer learning, we aim to develop predictive models that can assist clinicians in identifying and managing these ocular complications more effectively.

To facilitate our investigation, we make use of the [5] OCT-AND-EYE-FUNDUS-DATASET, a rich repository comprising 1548 Eye Fundus images and 1113 Macular OCT images captured between 2015 and 2022. This dataset is a product of collaboration between CONACYT CF-2019-1759 grant, PA-PIIT IN 205420, IMO (Instituto Mexicano de Oftalmología), APEC (Asociación Para Evitar la Ceguera), and INDEREB (Instituto de la Retina del Bajío). It represents a valuable resource for training and evaluating our predictive models.

In the sections that follow, we will examine the specifics of our suggested strategy, the design of our multiscale attention networks, and the insights gained through [6] transfer learning. We will also present our results, discuss their implications, and underscore the significance of accurate DME and DR prediction in the context of diabetes management and ocular health. In a world where diabetes continues to affect millions, and where the preservation of sight is a paramount concern, the research presented herein holds the potential to make a tremendous effect of the early detection and treatment of diabetic retinal diseases like macular edema.

II. METHODOLOGY

In this study, we harnessed the power of transfer learning, a well-established technique in deep learning, to leverage pre-trained convolutional neural network [7] (CNN) architectures. Transfer learning allows us to capitalize on the knowledge and representations learned from massive datasets, such as ImageNet, and adapt them to our specific fundus image classification task.

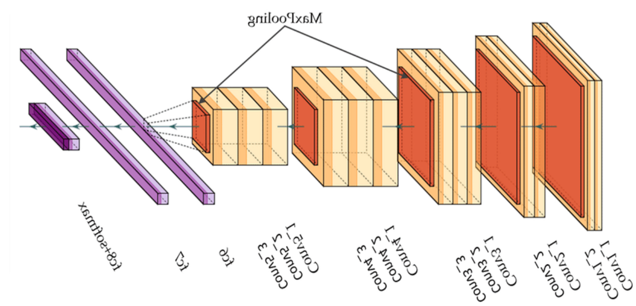

Fig. 1. The Architecture and Implementation of VGG-16 [8]

For deep learning models to perform well, an immense amount of labeled data and processing power are needed. Acquiring such resources in the context of medical image analysis can be tricky. The issue at hand gets solved by transfer learning (TL), a technique that allows us to use architectures of neural networks that have been pretrained on large amounts of data for broad visual recognition tasks.

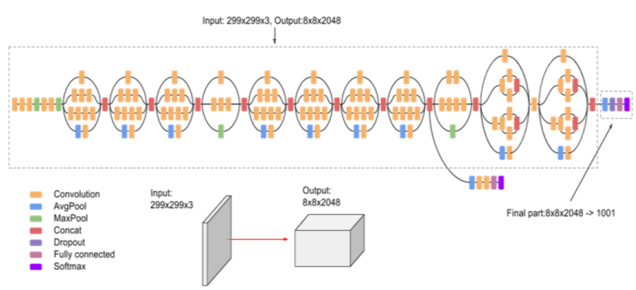

Fig.2.Inception-v3 [9]

Two prominent pre-trained models, namely [10] VGG16, and [11] InceptionV3, were selected for their remarkable performance in various computer vision tasks. Each of these models brings its unique architecture and capabilities to our study. In VGG16 we initiated our experiments with VGG16, a model renowned for its simplicity and effectiveness. By setting the weights of the pre-trained VGG16 model to [12]'imagenet,' we utilized the network's learned features while excluding its fully connected layers. This approach allowed us to incorporate the high-level abstractions captured by VGG16.Our second choice was InceptionV3, a model recognized for its utilization of multi-scale features through inception modules. The 'imagenet' weights were utilized, and as with the other models, we omitted the fully connected layers.

On top of each pre-trained model, we added custom deep layers that are tailored to our specific fundus images classification and retinal disease detection task. These layers included a flattening operation to transform the multi-dimensional feature maps into one- dimensional vectors. Subsequently, densely connected layers with ReLU activation functions were employed to capture complex relationships within the data. Dropout layers, were incorporated to reduce overfitting, with a dropout rate of 0.2. The last classification layer comprised three neurons with SoftMax activation, corresponding to the three classes of fundus images: Diabetic Macular Edema (DME), Diabetic Retinopathy (DR), and Healthy.

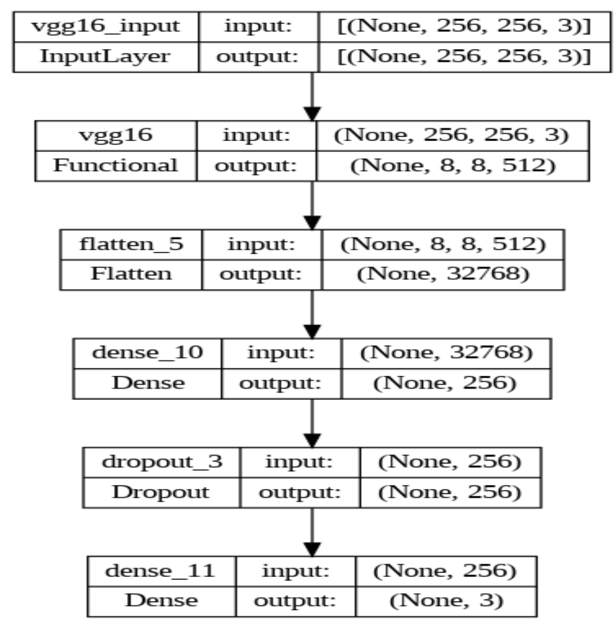

Fig.3. Various layers of PreTrained VGG16

To supplement and amplify the model's ability to generalize and adapt to variations in fundus images, we implemented data augmentation techniques during the training process. These techniques included random rotations, shearing, zooming, horizontal flips, and other transformations, thus augmenting our training dataset. The Adam optimizer was utilized to train the models, and a learning rate of 0.001 was used. We embedded early stopping mechanism with a patience of up to three epochs to prevent overfitting. The training dataset comprised both fundus and optical coherence tomography (OCT) images, contributing to a comprehensive and robust model.

An independent test set of fundus pictures was used to evaluate the models. To evaluate the effectiveness of each model, we calculated classification accuracy and created confusion matrices. In order to give thorough insights into the models' performance for each class (DME, DR, and healthy), classification reports were also created.

We wanted to use the information stored in pre-trained models while modifying them for our fundus picture classification assignment by adopting these transfer learning approaches. We were able to classify fundus pictures using our method accurately and effectively, which helped in the identification of diabetic eye problems.

Custom Multi-Scale Attention Network Architecture

In this study, we designed a unique [13] Multi-Scale Attention Network (MSAN) architecture specifically for the job of identifying pictures of healthy eye fundus, diabetic macular edema (DME), and diabetic retinopathy (DR). By capturing and integrating multi-scale information from the input photos, our unique MSAN allows the model to focus on important features and relationships at various granularities. Convolutional layers (Conv2D) are used at the outset of the MSAN to do feature extraction. The network can collect characteristics at various degrees of abstraction because to the different filter sizes of these convolutional layers. Max-pooling layers (MaxPooling2D) are applied after each convolutional layer to reduce spatial dimensions and create a hierarchical feature representation.

Fig.4.Advantages of using MSAN Approach [14]

A key component of our custom MSAN is the attention mechanism, which enables the network to dynamically weigh and emphasize certain features based on their importance. This attention mechanism is designed as follows: We apply a simple convolutional layer (Conv2D) to extract higher-level features. GlobalAveragePooling(GlobalAveragePooling2D) is used to compute the average value of each feature map. The attention mechanism includes two dense layers with ReLU activation functions. These layers compute attention scores and are followed by a SoftMax activation layer to normalize the scores. The attention scores are then used to weight the feature maps before multiplication. This step effectively directs the network's focus toward specific features.

Following the attention mechanism, a Flatten layer reshapes the output for compatibility with subsequent layers. The multiply layer combines the attention-weighted feature maps with the original features, giving more importance to the relevant features identified by the attention mechanism. The final layers of our custom MSAN consist of Global Average Pooling, followed by multiple dense layers with ReLU activation functions. These layers perform the task of classification, mapping the learned features to the three target classes: DME, DR, and healthy.

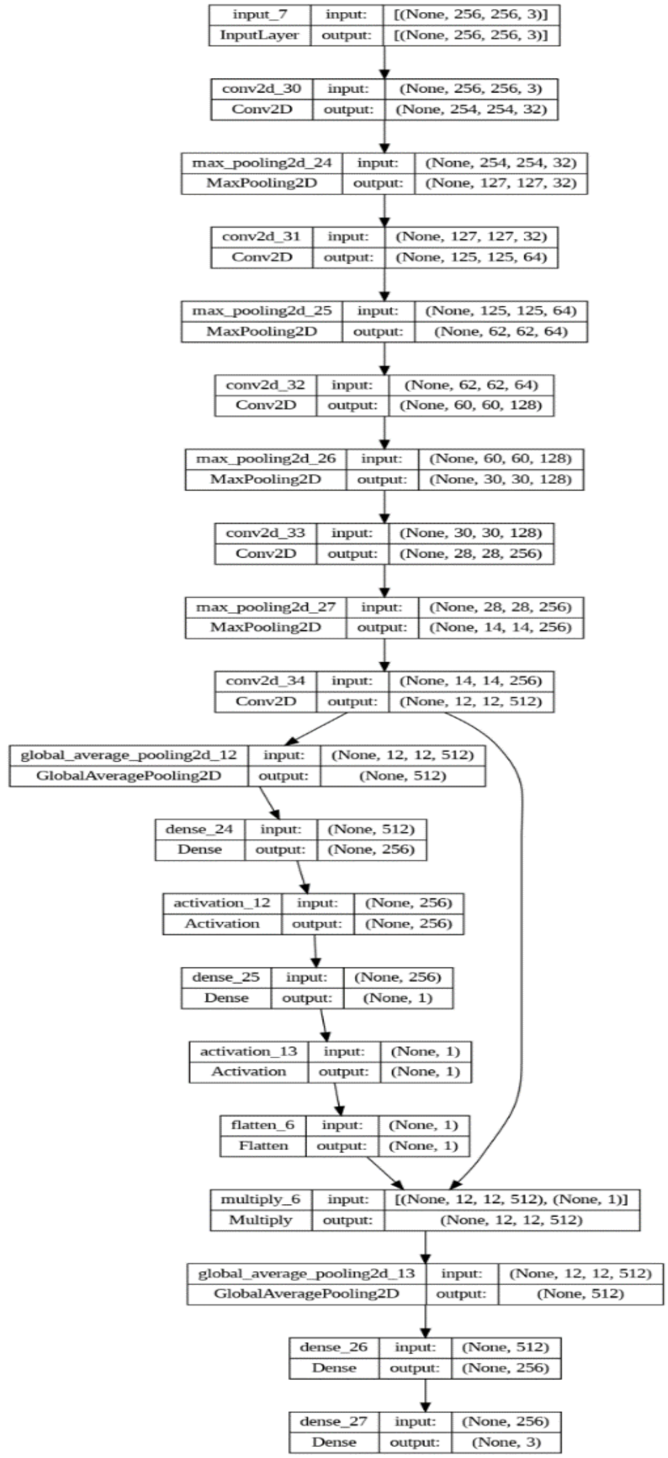

Fig.5.Various Layers of Custom-MSAN

The Adam optimizer is used to create the customized MSAN, with a categorical cross-entropy loss and a learning rate of 0.001. As the main metric for measuring the model's performance, accuracy is used. By customizing the architecture in this manner, our MSAN is designed to effectively capture multi-scale features and relationships within eye fundus images, enabling accurate classification of DME, DR, and healthy cases.

The attention mechanism plays a determining role in enhancing the model's ability to focus on critical image regions and contributes to its overall performance. We Decided to go with the stellar performing Multi- Scale Attention network mechanism instead of other counterparts in attention networks as it was more apt to the dataset used. MSAN can perceive the retinal images in multiple scales which benefits in detection of both large, significant changes in the retina as well as minor and intricate changes in the fundus images.

III.RESULTS AND DISCUSSIONS

Here, we present the results and discussions stemming from our proposed methodologies. To rigorously assess our models, we partitioned the FUNDUS dataset into an 80% training set and a 20% testing set. Leveraging the computational power of Google Colab equipped with a T4 GPU, we conducted our experiments. Prior to model training, we meticulously preprocessed the images, incorporating a range of image augmentation techniques. Additionally, we standardized the image dimensions to 256x256x3 pixels.

TABLE I. HYPER PARAMETERS OF VGG16 AND CUSTOM-MSAN

Parameter	**Value**
Optimizer	Adam & SGDM
Batch Size	32
Device Type	T4 GPU
Learning Rate	0.1/0.01/0.001
Epoch	20

TABLE II. HYPER PARAMETERS OF VGG16 AND CUSTOM-MSAN

TL Model	Epoch	Learning Rate	Adam-Accuracy	SGDM-Accuracy
custom-msan	20	0.001	0.9100	0.8810
custom-msan	10	0.001	0.9096	0.8971
custom-msan	5	0.001	0.8838	0.8939
custom-msan	20	0.01	0.8778	0.7010
custom-msan	10	0.01	0.8875	0.7010
custom-msan	5	0.01	0.7363	0.7010

TL Model	Epoch	Learning Rate	Adam-Accuracy	SGDM-Accuracy
VGG16	5	0.1	0.8715	0.6771
VGG16	10	0.1	0.7847	0.6771
VGG16	5	0.01	0.9167	0.8819
VGG16	10	0.01	0.9028	0.8611
VGG16	5	0.001	0.9167	0.8924
VGG16	10	0.001	0.8854	0.8924
VGG16	20	0.1	0.6562	0.6771
VGG16	20	0.01	0.8333	0.8924
VGG16	20	0.001	0.9340	0.8785
Inception-V3	10	0.1	0.8889	0.8958
Inception-V3	10	0.01	0.8924	0.9062
Inception-V3	10	0.001	0.9132	0.9028
Inception-V3	20	0.1	0.8785	0.6771
Inception-V3	20	0.01	0.9167	0.9132
Inception-V3	20	0.001	0.9132	0.8924
Inception-V3	5	0.1	0.8889	0.8819
Inception-V3	5	0.01	0.9028	0.8993
Inception-V3	5	0.001	0.9167	0.8819

TABLE III. ACCURACY RESULTS VIA HYPERPARAMETER TUNING FOR PRETRAINED VGG16

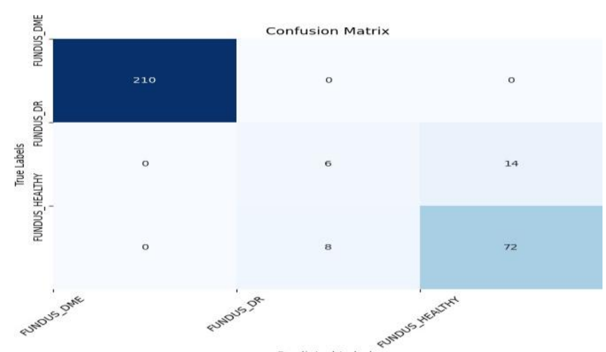

Fig.6.VGG16 model's Confusion Matrix

Subsequently, these meticulously prepared images served as input for both the pretrained VGG16 model and our custom Multi-Scale Attention Network (MSAN) model during the training process. provides an overview of the default training parameters employed in our proposed model. As shown in Tables II and III, we specifically adjusted important variables including the optimizer, learning rate, and number of epochs to improve accuracy.

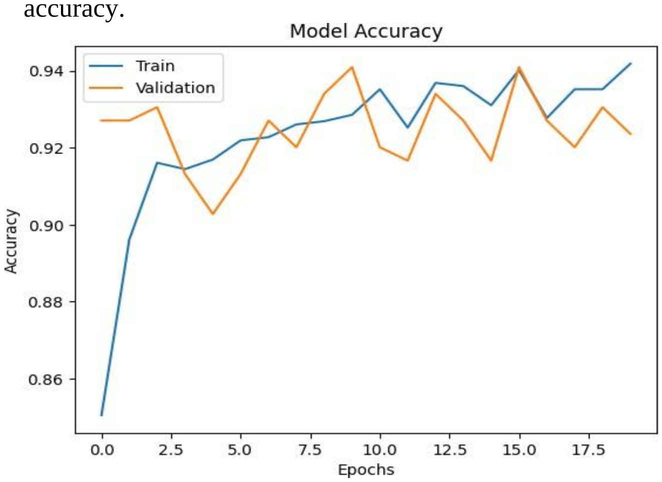

Fig.7.VGG16 Training History

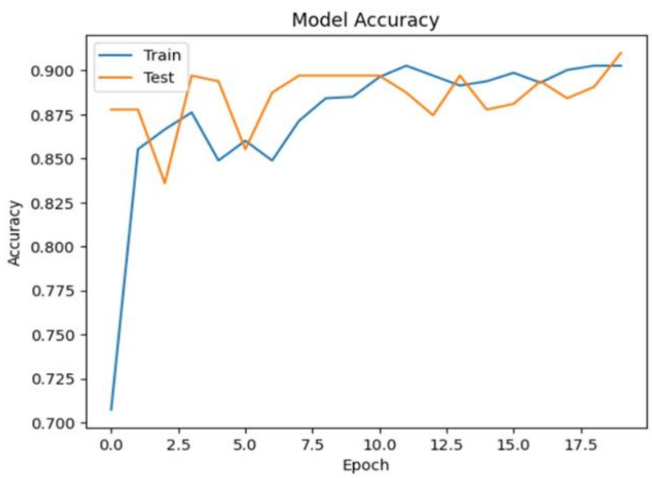

Fig.8.Custom-MSAN Training History

Specifically, using the Adam optimizer with a learning rate of 0.001 during 20 training epochs, our suggested VGG16 model optimizer yielded improved results compared to the stochastic gradient descent optimizer attained an impressive accuracy of 93.40%. Similar to other models, our Custom-MSAN model performed well, reaching accuracy of 91% with the Adam optimizer set up, a learning rate of 0.001, and a training time of 20 epochs and this is very obvious via the Figuers [7] and [8].

Figures [10], [11] show the confusion matrix, critical model parameters such as accuracy, recall, F1-score, and support score, as well as fundamental training plots. These metrics provide comprehensive analysis of the effectiveness of our planned effort.The confusion matrix, shown in Figures [6], [9], in particular, is crucial for evaluating the models. It offers a thorough analysis of DME, DR, and healthy categories for eye disease issues.

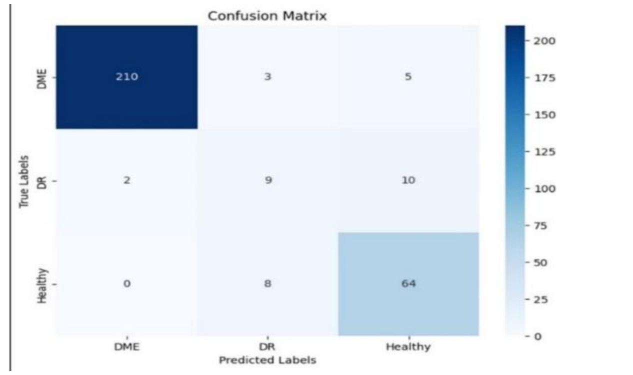

Fig.9.Custom-MSAN confusion matrix

We got respectable results in the VGG16 model, properly classifying 210 occurrences as DME, 6 as DR, and 72 as healthy. Similar results were seen in the Custom-MSAN model, where 210 instances were accurately identified as DME, 9 as DR, and 64 as healthy. These outcomes highlight the reliability and efficiency of our suggested algorithms in correctly classifying eye disease situations.

	precision	recall	f1-score	support
FUNDUS_DME	1.00	1.00	1.00	210
FUNDUS_DR	0.43	0.30	0.35	20
FUNDUS_HEALTHY	0.84	0.90	0.87	80
accuracy			0.93	310
macro avg	0.76	0.73	0.74	310
weighted avg	0.92	0.93	0.92	310

Fig.10.Metrics of VGG16 Model

	precision	recall	f1-score	support
DME	0.99	0.96	0.98	218
DR	0.45	0.43	0.44	21
Healthy	0.81	0.89	0.85	72
accuracy			0.91	311
macro avg	0.75	0.76	0.75	311
weighted avg	0.91	0.91	0.91	311

Fig.11. Metrics of Custom-MSAN Model

IV. CONCLUSIONS

A new retinal illness prediction system using transfer learning and attention network techniques is presented in this research paper. The pre-processing of the dataset and training of retinal pictures using the pre-trained model VGG16 and Custom-MSAN models for the categorization of retinal illness into three groups make up the proposed work. The experimental findings show that the categorization of retinal diseases is highly accurate (93.40% and 91%). To illustrate the recommended method's supremacy, its performance is then evaluated against that of other previously trained models and other approaches. In order to prevent vision loss and to urge prompt diagnosis, it was advised that the proposed technology be employed for real-time detection of retinal disorders.


Fig.12a


Fig.12.b

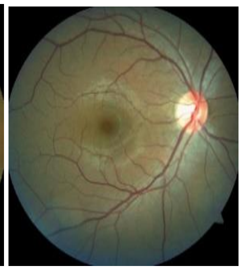
Fig.12.c

ACKNOWLEDGMENT:

The authors are extremely exuberant and elated towards the unbounding support provided by the Thiagarajar College of Engineering (TCE) for reinforcing and buttressing us in getting this quality research work done. Special thanks to Hughes Cano, J. A., Olivares Pinto, U. & Thébault, S. C. “Dataset of Eye Fundus and OCT Images for the study of Diabetic Macular Edema and Diabetic Retinopathy” for providing us with the dataset.

REFERENCES

[1]. Hughes-Cano, JA & Quiroz-Mercado, H & Hernández-Zimbrón, LF & García-Franco, R & Mijangos, JF & López, Ellery & Garcia Roa, Marlon & Lansingh, Van & Olivares-Pinto, Ulises & Thebault, Stephanie. (2023). Improved predictive diagnosis of diabetic macular edema based on hybrid models: an observational study. 10.1101/2023.04.05.23288182.

[2]. A. Bajwa, N. Nosheen, K. I. Talpur, and S. Akram, “A Prospective Study on Diabetic Retinopathy Detection Based on Modify Convolutional Neural Network Using Fundus Images at Sindh Institute of Ophthalmology & Visual Sciences,” Diagnostics, vol. 13, no. 3, p. 393, Jan. 2023, doi: 10.3390/diagnostics13030393.

[3]. Oh, K., Kang, H.M., Leem, D. et al. Early detection of diabetic retinopathy based on deep learning and ultra-wide-field fundus images. Sci Rep 11, 1897 (2021). https://doi.org/10.1038/s41598-021-81539-3

[4]. Nagasawa, T., Tabuchi, H., Masumoto, H. et al. Accuracy of ultrawide-field fundus ophthalmoscopy-assisted deep learning for detecting treatment-naïve proliferative diabetic retinopathy. Int Ophthalmol 39, 2153–2159 (2019). https://doi.org/10.1007/s10792-019-01074-z

[5]. Hughes Cano, J. A., Olivares Pinto, U. & Thébault, S. C. “Dataset of Eye Fundus and OCT Images for the study of Diabetic Macular Edema and Diabetic Retinopathy” “https://github.com/Traslational-Visual-Health-Laboratory/OCT-AND-EYE-FUNDUS-DATASET”

[6].C, Dharani; S, Charulatha; Suguna, Mariappan “Transfer Learning based Diabetic Retinopathy Classification” DOI: http://dx.doi.org/10.12785/ijcds/14011

[7]. Aditi Arora, Shivam Gupta, Shivani Singh & Jaya Dubey Eye Disease Detection Using Transfer Learning on VGG16 https://doi.org/10.1007/978-981-19-1142-2_42

[8] S. Sarah, V. Singh, M. K. Gourisaria, and P. K. Singh,“Retinal Disease Detection using CNN through Optical Coherence Tomography Images,” in 2021 5th International Conference on Information Systems and Computer Networks (ISCON), 2021, pp. 1–7, doi: 10.1109/ISCON52037.2021.9702480.

[9]. Li F, Liu Z, Chen H, Jiang M, Zhang X, Wu Z. Automatic Detection of Diabetic Retinopathy in Retinal Fundus Photographs Based on Deep Learning Algorithm. Transl Vis Sci Technol. 2019 Nov 12;8(6):4. doi: 10.1167/tvst.8.6.4. PMID: 31737428; PMCID: PMC6855298.

.

[10]. Chen, Wei & Zhong, Funing & Tan,. (2019). “Multiple-Oriented and Small Object Detection with Convolutional Neural Networks for Aerial Image. Remote Sensing.” doi: 11. 2176. 10.3390/rs11182176.

[11]. Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jonathon Shlens, Zbigniew Wojna “Rethinking the Inception Architecture for Computer Vision”, doi: https://doi.org/10.48550/arXiv.1512.00567

[12]. J. Deng, W. Dong, R. Socher, L. -J. Li, Kai Li and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 2009, pp. 248-255, doi: 10.1109/CVPR.2009.5206848.

[13]. Al-Antary, Mohammad & Arafa, Yasmine. (2021). Multi-Scale Attention Network for Diabetic Retinopathy Classification. IEEE Access. PP. 1-1. 10.1109/ACCESS.2021.3070685.

[14]. He X, Deng Y, Fang L, Peng Q. Multi-Modal Retinal Image Classification with Modality-Specific Attention Network. IEEE Trans Med Imaging. 2021 Jun;40(6):1591-1602. doi: 10.1109/TMI.2021.3059956. Epub 2021 Jun 1. PMID: 33625