Multi-label detection of ophthalmic disorders using InceptionResNetV2 on multiple datasets

Ali Ghadiri, Afrooz Sheikholeslami
Department of Electrical and Computer Engineering
Isfahan University of Technology
Isfahan, Iran
alighadiri@ec.iut.ac.ir, afrooz.sheikholeslami@ec.iut.ac.ir

Asiyeh Bahaloo*
VCteam
Tehran, Iran
asiyeh.bahaloo@gmail.com

* Corresponding author

***Abstract*—Recently, AI-based methods have been extensively used to help in the process of diagnosing eye diseases due to their prevalence. But since these methods can't be generalized well, they can't be used in the real world. In this paper, we compared the two fundamental approaches for improving the model's performance on the eye disease detection task. The idea is that, for real-world applications, using multiple datasets for robustness is more beneficial than enhancing the architecture just to increase the accuracy. To demonstrate this, we chose three state-of-the-art architectures as our baseline and changed them slightly so that the overfitting wouldn't happen. For the first approach, we change the classification head to XGB and SVM, and for the second approach, we combine the two datasets for the training stage. The results show that high-quality data with representative distribution can have a better effect than sophisticated architecture for real-world applications. This approach performed 3% better than the last state-of-the-art model. The implementation is available at https://gitlab.com/asiyeh.bahaloo/eye-disease**

Keywords—InceptionResNetV2, Multi-label classification, Eye disease, Fundus images

I. INTRODUCTION

Due to the increased interaction of human eyes with digital screens and the population's aging, there has been a dramatic increase in the prevalence of ophthalmic disorders worldwide over the past few years. According to recent studies, 2.2 billion people worldwide suffer from near or distance vision impairment, of which half of them may have been avoided by being addressed early [1]. Two prevalent eye impairments are glaucoma and cataracts. In patients suffering from glaucoma, the nerves that connect the eye to the brain become damaged. Cataracts happens when a cloudy shape develops in a patient's eyes, which leads to blurry vision and, eventually, vision loss. Like in any other healthcare field, Artificial Intelligence has offered effective solutions that can magnificently help specialists diagnose eye disease more accurately. Based on our research, deep learning (DL) models are utilized to identify eye problems both in the academic and industrial worlds. Moreover, the potential objectives of employing these models are divided into three major categories: disease diagnosis [2], disease severity classification [3], and segmentation for extracting the Regions of Interest (ROIs) [4]. While there are various techniques for capturing the retina surface in clinical diagnosis procedures, by convention, most ophthalmic databases contain either retinal fundus or optical coherence tomography (OCT) images. Fundus images are visual records that capture the retina's current ophthalmoscopic appearance in a patient. The central and peripheral retina, optic disc, and macula are the primary structures visible in a fundus photograph [5]. Our prime objective in this paper is to detect eight different eye diseases from fundus images due to the availability of adequate free-access fundus image datasets in the field of eye disease detection. Here, we have used the ODIR dataset as well as the cataracts dataset, which includes color fundus photographs of both left and right eyes. In this study, modern computer vision architectures like VGG16, Inception-V3, and InceptionResNet-V2 have been used to establish a baseline for future comparison. After setting a baseline for each architecture, we tried to assess the effects of the two approaches on the final performance of our models. In order to make the model more robust in the real world, we tried to combine multiple datasets. Subsequently, we attempt to improve the performance by combining deep learning feature extractor ability and machine learning classifiers. To report more reliable results, we provided various evaluation metrics for the experiments, such as kappa, f1-score, and AUC. The structure of the rest of this paper is as follows. Section 2 provides a summary of the dataset and the preprocessing performed on the data. Section 3 explores the relevant related works and discusses their approaches. Section 4 presents the description of the overall proposed framework. Section 5 provides the experimental studies which illustrate the performance of our baseline models and then compares the proposed methods' results with them. Finally, in Section 6, the main ideas are concluded.

II. DATASET

ODIR is a public dataset that contains 7000 high-resolution fundus images of both left and right eyes [6]. The dataset includes an annotations file in which the labels of each patient's eyes are provided. Patients are labeled by specialists with eight different labels that identify the pathologies: N, D, G, C, A, H, M, and O, which represent Normal (Without any disease), Diabetes, Glaucoma, Cataracts, AMD (Degeneration of the macula), Hypertension, Myopia, and other diseases,

respectively. Each patient may be tagged with one or more labels indicating that they are suffering from various diseases concurrently. Out of these 7,000 images, 449 images were discarded from the dataset due to various reasons like poor quality, giving us a total of 6551 images. Furthermore, the distribution of the number of images in each class is highly unbalanced, as shown in Fig. 1. To create the train and validation sets, we randomly selected images from each of the eight classes in the dataset as well as those that had been labeled with multiple diseases to form a balanced validation set of 400 images. The remaining 6151 images were used as the training set. Furthermore, the cataract dataset is used in one of the proposed approaches. This dataset includes 600 fundus images categorized into four groups: Normal, Cataracts, Glaucoma, and Retina disease with 300, 100, 100, and 100 samples, respectively [7].

III. Related Works

The works carried out on ocular disease diagnosis are generally divided into two categories of conventional Machine Learning and Deep Learning techniques [8]. Researchers have effectively implemented conventional machine learning techniques in the past few years to classify and grade cataracts and associated severity [8]. Support Vector Machines (SVM) are used in [9][10] to classify the severity of cataracts using slit lamp images. Fudah et al. [11] created a system for the detection of cataracts using the K-nearest Neighbor (K-NN) algorithm. Furthermore, with the staggering rate of deep learning development, numerous DL algorithms have been applied to a broad range of eye disorders and diseases [12]. A fusion module comprising two weak classification models was proposed by researchers in [13] for the identification of one or more fundas disorders. Each weak classifier is made up of an EfficientNet that functions as a feature extractor, and the output is fed into a customized neural network designed as a multi-label classifier. The final detection result is then calculated by averaging the sigmoid probability of the two models. This research is very similar to ours as they utilized a number of outstanding feature extractors, whereas we did not use the combination of the output of two classifiers. Elsawy et al. [14] implemented a multi-disease deep learning diagnostic network (MDDN) of common corneal diseases encompassing dry eye syndrome (DES), Fuchs endothelial dystrophy (FED), and keratoconus (KCN) using anterior segment optical coherence tomography (AS-OCT) images. A hybrid InceptionV3 XGBoost model was developed by Ramaneswaran et al. [15] for the diagnosis of acute lymphoblastic leukemia (ALL). Through the experiments, they demonstrated that the performance would be boosted by applying XGBoost as the classification head. A novel two-step framework based on deep learning models was proposed in [16] for ocular disease detection to achieve high performance. In the initial phase, the CNN model classifies normal versus disease cases. The second step involves detecting subtypes of the disease. The outputs of the two models indicated above were then fused using a metamodel. In their deep learning model, Luo et al. [17] combined the focal loss and cross-entropy induced loss functions to create a novel mixture loss function that improved recognition performance. In [18], a deep learning algorithm capable of diagnosing diabetic retinopathy is developed. Their approach was to train the fine-tuned model using a cosine annealing strategy for decaying the learning rate with warm-up. For the purpose of identifying corneal abnormalities from slit-lamp images, Gu et al. [19] developed a novel hierarchical deep learning network comprised of a family of multi-task multi-label learning classifiers. Various studious have taken advantage of data-oriented approaches [20][21][22][23] for enhancement. In [20][21], left and right eye images of a patient are taken into account simultaneously. Two distinct approaches for using both eyes are implemented in [20]. They have both stacked feature maps obtained from applying left and right images individually and concatenated the two eye images in the first place as input to the model. Nonetheless, in [21], they utilized a backbone network for extracting features of the right and left eyes separately, and then the produced features will pass through a fusion module consisting of element-wise multiplication, sum, and concatenation. Another data-driven method for obtaining promising results is to combine several datasets [22][23]. The latter have used a combination of Messidor, Messidor-2, DRISHTI-GS, and Kaggle cataract datasets to classify five labels of Diabetic retinopathy (DR), Diabetic Macular Edema (DME), Glaucoma (GL), Cataracts (Ca) and Normal. Some scholars focused on segmentation tasks, including [24], who proposed ReLayNet, an encoder-decoder network that is utilized for the semantic segmentation of retinal layers and fluid masses in eye OCT scans. Also, Hue et al. [25] proposed a novel vessel segmentation method which is a mixture of CNN and fully connected conditional random fields (CRFs) making use of fundus images.

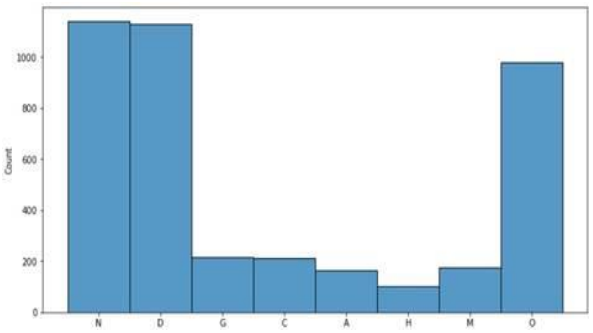

Fig. 1. The number of samples in each class of the dataset

IV. The Proposed Framework

A. Data preprocessing and augmentation

As mentioned, the ODIR dataset labeled each eye in text and assigned a label for both eyes. Here, we have used the preprocessed version of ODIR. In this version, each eye is labeled separately based on the physician's comment in the original dataset. More details of this conversion may be found here [26]. Given that the number of data records for each disease is highly unbalanced in the ODIR dataset, we decided to tackle this issue by augmenting the training set to have a fairly equal number of images for each class. Various treatments were applied to the photographs during the augmentation process, such as changing contrast and saturation and rotating the images. After applying the

augmentation, the final training set includes 25076 samples. In addition, we performed some preprocessing on the images we fed to our model. Our preprocessing pipeline is as follows: First, we have removed the dark paddings of the images as these dark padding doesn't provide any useful information for the model. Then the photos were resized to have identical shapes of (224*224*3).

B. Base Models

In this section, three well-known and popular convolutional neural networks (CNN), including VGG16 [27], InceptionV3 [28], and InceptionResNetV2 [29], are selected and represented. InceptionResNetV2 is a combination of the two state-of-the-art CNNs, Inception and ResNet, which is an Inception-style network that takes advantage of residual connections instead of filter concatenation. In this network, batch-normalization is only used on top of traditional layers but not on top of summations.

We improved the CNNs that were already trained on the Imagenet dataset by loading the Imagenet weights and then training all of the layers. In order to get even superior results, we made some structural changes to the architecture. Firstly, we added a dropout layer following every convolutional layer in the three architectures aiming to prevent over-fitting. In addition, weight decay or L_2 Regularization was added. Weight decay is a regularization method used in deep learning that applies a penalty term to a neural network's cost function. This causes the weights to shrink during backpropagation and protects the network from overfitting and the exploding gradient problem. L2 regularization is applied by adding the squared sum of the weights to the error term, and it is multiplied by a manually chosen hyperparameter called lambda. The equation below shows the loss function when the L_2 regularization term is added:

$$Loss = Error(y, \hat{y}) + \lambda \sum_{i=1}^{N} w_i^2 \quad (1)$$

Where N indicates the number of samples, w is model weights, y is the true labels, and yhat is the predicted output.

C. Multi-label Classification

As we are dealing with a multi-label classification problem, we cannot use the traditional loss function and softmax activation function. Unlike the softmax function that forces the output probabilities to have a sum of one, the output produced by a sigmoid is independent and does not produce a probability vector and thus is suited for multi-label problems. In addition, the cross-entropy loss can be employed as a loss function for a multi-label problem by computing binary cross-entropy for each class separately and summing up the losses, which is shown below:

$$Loss = \sum_x -\big(p(x).\log q(x) + \big(1 - q(x)\big).\log\big(1 - q(x)\big)\big) \quad (2)$$

Where q(x) is the predicted probability of class x, and p(x) is the label of class x in targets.

D. Classification Head

In this section, a novel architecture is proposed in which we have employed both XGboost and SVM classifiers instead of dense layers as a classification head. XGBoost [30] is a proven tree boosting algorithm. The base models introduced in the previous section provide input features for Xgboost and SVM classifiers. Since a real fundas image is more likely to consist of several diseases, our task is a multi-label, multi-class classification. In addition, XGBoost and SVM classifiers are not able to handle multi-label classification inherently, and thus we need to take advantage of a method in order to extend them to support multi-label classification as described below. The whole process is shown in Fig. 2

1) Binary Relevance Method

This strategy trains L classifiers C_1, ..., C_L, and each classifier C_i is responsible for predicting the label $l_i \in L$ [31]. This method ignores existing label correlations. To clarify, Binary Relevance Method converts multilabel classification problems into simple binary classifications for each label on which the binary learner is applied.

2) Chain Classifiers Method

This method utilizes L binary classifiers resembling Binary Relevance strategy [31]. However, classifiers are connected in a chain where each classifier Ci is augmented by previous classifiers C_1, …, and C_{i-1} predictions. This is the reason why chain classifiers are capable of modeling class label correlations and overcoming the problem of label independence.

E. Dataset Combination

The second approach we adopted in this paper was to examine the effect of adding more data to our training set rather than working on enhancing the model architecture. To accomplish this, we decided to utilize the Cataract dataset and enrich the training data. Our prime goal was to diversify each of the batches during the training process. To do so, according to the size of the augmented ODIR training dataset and the Cataract dataset, we combined 31 images of the augmented ODIR training dataset and one image of the Cataract dataset to form each of the batches we train our model on. The identical preprocessing that was used for the ODIR dataset was performed on the Cataract dataset as well. To be consistent with our dataset labels, we have discarded the retina disease class and just merged the 500 images of other classes to the ODIR dataset images to form our batches.

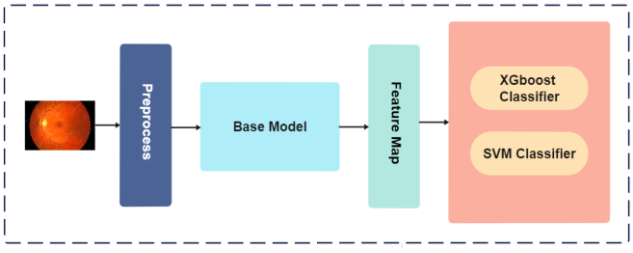

Fig. 2. The architecture of the proposed hybrid model with XGBoost and SVM classifier

V. EXPERIMENTS

In this section, we first review the metrics which are used in our experiments, then we elaborate on the configurations we used to conduct the experiments as well as the hyper-parameters of each model, and finally, we provide and discuss the outcomes of our models.

A. Metrics

Our experimental results use various evaluation metrics to assess each experiment from different aspects. The number All the defined metrics and their formula are listed below:

$$F1\_score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (1)$$

$$Kappa = \frac{p_0 - p_e}{1 - Pe} \quad (2)$$

$$Final\_score = \frac{Kappa + F1\_score + AUC}{3} \quad (3)$$

To be more precise, the averaging method that is used in our metrics is "micro," in which we calculate metrics globally by counting the total number of true positive, true negative, false positive, and false negative cases. This strategy was chosen since it doesn't handle each class differently and seems more accurate. In this method, it is more likely that most of the classes be predicted more precisely as the metrics rise rather than one class being predicted accurately while the others are misclassified. Kappa is a metric for consistency evaluation. AUC refers to the area under the receiver operating characteristic (ROC) curve.

B. Configuration

The shared hyper-parameter setup of all models is illustrated in TABLE I. After carrying out numerous trials for each model, the resulting most optimum hyper-parameters for our base models are shown in TABLE II. and below it, other changed hyper-parameters of our proposed methods are depicted. To be more precise, "#Given features" represents the number of features extracted by the CNN backbone we are feeding to the classifier that depends on which dense layers we are getting the features from. Among the possible ways for enabling the XGBoost and SVM algorithms to handle multi-label classification, the Binary Relevance method is chosen due to its better or identical performance compared to the Chain Classifiers method, according to our experiments. The kernel function we used in the SVM classifier is polynomial, and the parameter C is set to 1. All experiments for this project are done on an Azure server with an NVIDIA Tesla K80 graphics card and 12 GB of memory.

C. Results

When it comes to comparing the outcomes of different approaches, we report the metrics of the epoch with the lowest val_loss in TABLE III. However, in some cases, our models could reach a higher val_final_score in the following epochs. Thus, when we are comparing our base models with state-of-the-art works, we consider the overall best val_final_score since other papers didn't report the metrics for the epoch having the lowest val_loss.

TABLE I. SHARED HYPER-PARAMETER CONFIGURATION

Configuration	***Value***
Loss Function	Binary cross-entropy
Optimizer	SGD
Optimizer decay rate	0.95
Optimizer momentum rate	0.9
Nesterov	True
Batch Size	32
EarlyStopping	Monitor='val_loss', Patience=8

TABLE II. SPECIFIC HYPER-PARAMETER CONFIGURATION

Configuration	**Base Models**		
	VGG16	***InceptionV3***	***InceptionResNetV2***
LR	0.001	0.01	0.01
Dropout	0.25	0.25	0.3
WeightDecay	0.2	-	0.5
BatchNormalization	No	Yes	Yes
#Parameters	134,293,320	23,874,728	55,858,280
	With Mix dataset		
LR	0.001	0.005	0.01
	With XGBoost & SVM		
#Given features	4096	1024	1536

As we can see in TABLE III. , the InceptionResnetV2 base model outperformed the other base model, with a val_final_score of 66.3%. Moreover, the highest achieved val_final_score for this method was 70% which is 3% better than the reported score in state-of-the-art attempts [13]. Furthermore, comparing our results with [20], we can see higher val_accuracy than what is reported for both of their examined approaches.

Our first approach for enhancing the model's performance was to add a new dataset to our training batches. This approach results are reported with the keyword "Mix dataset" in TABLE III. All models trained on the Mix dataset showed between 0.4% and 2.1% higher val_final_score compared with the base models. Again, the InceptionResnetV2 model outperformed the other models, with a val_final_score of 67.5%.

Comparing the results of models with XGBoost as the classifier with our base models, while the val_final_score for both InceptionV3 and InceptionResnetV2 models witness an increase, for VGG16 architecture, there is a marginal decrease.

Looking at the results of the models with SVM as its classifier, while the VGG16 base model showed far better performance compared with its SVM version, the InceptionV3

TABLE III. MODELS' PERFORMANCES

	elapsed time	*loss*	*Val Loss*	*Final score*	*Val_Final Score*
VGG16	18.5 h	51.23	51.33	0.768	0.654
InceptionV3	16.8 h	0.231	0.291	0.697	0.635
InceptionResnetV2	26 h	0.178	0.282	0.735	0.663
VGG16 + Mix dataset	22.1 h	51.47	51.63	0.879	**0.658**
InceptionV3+ Mix dataset	32 h	0.245	0.288	0.675	**0.656**
InceptionResnetV2 + Mix dataset	32 h	0.22	0.279	0.72	**0.675**
VGG16 + XGBoost	30 m	0.64	0.689	0.996	0.65
InceptionV3 + XGBoost	32 m	0.642	0.692	0.985	0.659
InceptionResnetV2 + XGBoost	33 m	0.64	0.688	1.00	0.677
VGG16 + SVM	9.6 h	0.685	0.686	0.583	0.585
InceptionV3 + SVM	2.4 h	0.677	0.682	0.700	0.640
InceptionResnetV2 + SVM	2.6 h	0.667	0.682	0.793	0.658

architecture with SVM could slightly outperform its base model. Also, the InceptionResnetV2 model's performance didn't improve with changing the classifier to SVM.

According to outcomes, it seems the models trained on the Mix dataset are more stable than the other approaches since the difference between their training and validation metrics is lower.

We contrast the patterns in the val_final_score between our base models and models trained on the Mix dataset in Fig. 3. We can observe that in both models, the one that was trained on the Mix dataset outperforms the base model, which is similar to the results presented in TABLE III.

The same information is shown for the InceptionResnetV2 model in Fig. 4. Although the base model reached a higher val_final_score in total, because of its fluctuations, it seems the predicted results by the Mix dataset version are more reliable and robust.

In Fig. 5, InceptionV3 and InceptionResnetV2 models' validation loss trends during the training process are depicted. For the InceptionV3 model, it is clear that using the MIX dataset for training the model lowered the validation loss in general. Turning to the InceptionResnetV2, while the validation loss of the Mix dataset version is slightly higher than the base model due to the unsteady changes in validation loss figures in this model, again, we can conclude that the outcomes of the model trained on the Mix dataset are more reliable.

VI. CONCLUSION

In this article, we presented an automatic data-driven method for diagnosing eye diseases and demonstrated that this method has better accuracy and stability than architecture-based methods for real-world usage. Therefore, in real medical applications, it is preferable to use simpler models trained on more high-quality data. Furthermore, this research shows that

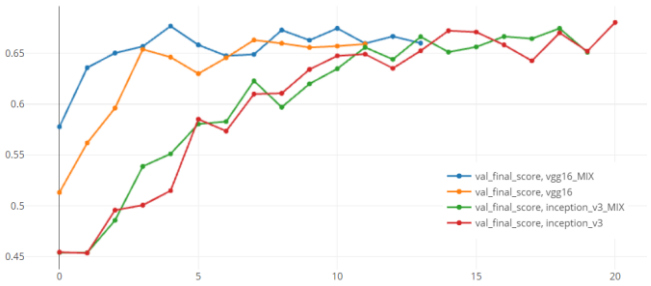

Fig. 3. Comparison of the val_final_score for VGG16 and InceptionV3 models

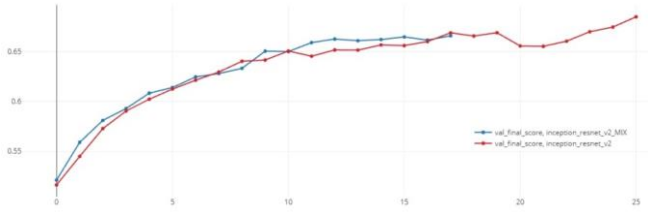

Fig. 4. Comparison of val_final_score for InceptionResnetV2 model

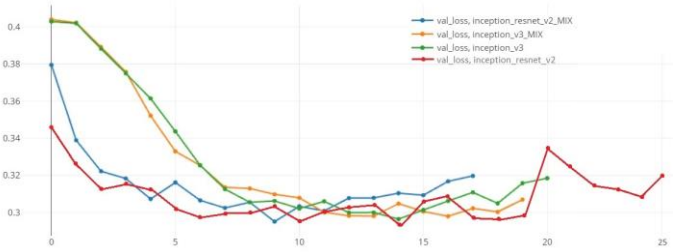

Fig. 5. Comparison of val_loss for InceptionV3 and InceptionResnetV2 models

relying on the best metrics can be misleading if the training process is not examined.

ACKNOWLEDGMENT

This research was partially supported by the KeyLead Health company. We thank Amin Jamshidi for his assistance with this paper.

REFERENCES

[1] Steinmetz, Jaimie D., et al. "Causes of blindness and vision impairment in 2020 and trends over 30 years, and prevalence of avoidable blindness in relation to VISION 2020: the Right to Sight: an analysis for the Global Burden of Disease Study." The Lancet Global Health 9.2 (2021): e144-e160.

[2] Sarki, Rubina, et al. "Automated detection of mild and multi-class diabetic eye diseases using deep learning." Health Information Science and Systems 8.1 (2020): 1-9.

[3] Grassmann, Felix, et al. "A deep learning algorithm for prediction of age-related eye disease study severity scale for age-related macular degeneration from color fundus photography." Ophthalmology 125.9 (2018): 1410-1420.

[4] Budai, Attila, et al. "Robust vessel segmentation in fundus images." International journal of biomedical imaging 2013 (2013).

[5] fundus photography overview - ophthalmic photographers' society URL: https://www.opsweb.org/page/fundusphotography

[6] Odir-2019-grand challenge URL: https://odir2019.grand-challenge.org/

[7] Cataract dataset URL: https://www.kaggle.com/datasets/jr2ngb/cataractdataset

[8] Zhang, Xiao-Qing, et al. "Machine Learning for Cataract Classification/Grading on Ophthalmic Imaging Modalities: A Survey." Machine Intelligence Research 19.3 (2022): 184-208.

[9] Jiang, Jiewei, et al. "Automatic diagnosis of imbalanced ophthalmic images using a cost-sensitive deep convolutional neural network." Biomedical engineering online 16.1 (2017): 1-20.

[10] Wang, Liming, et al. "Comparative analysis of image classification methods for automatic diagnosis of ophthalmic images." Scientific reports 7.1 (2017): 1-11.

[11] Fuadah, Y. Nur, A. Wahyu Setiawan, and T. L. R. Mengko. "Performing high accuracy of the system for cataract detection using statistical texture analysis and K-Nearest Neighbor." 2015 International Seminar on Intelligent Technology and Its Applications (ISITIA). IEEE, 2015.

[12] Kamal, Md Muntasir, et al. "A Comprehensive Review on the Diabetic Retinopathy, Glaucoma and Strabismus Detection Techniques Based on Machine Learning and Deep Learning."

[13] Wang, Jing, et al. "Multi-label classification of fundus images with efficientnet." IEEE Access 8 (2020): 212499-212508.

[14] Elsawy, Amr et al. “Multidisease Deep Learning Neural Network for the Diagnosis of Corneal Diseases.” American journal of ophthalmology vol. 226 (2021): 252-261. doi:10.1016/j.ajo.2021.01.018

[15] Ramaneswaran, S., et al. "Hybrid inception v3 XGBoost model for acute lymphoblastic leukemia classification." Computational and Mathematical Methods in Medicine 2021 (2021).

[16] An, Guangzhou, et al. "Hierarchical deep learning models using transfer learning for disease detection and classification based on small number of medical images." Scientific reports 11.1 (2021): 1-9.

[17] Luo, Xiong, et al. "Ophthalmic disease detection via deep learning with a novel mixture loss function." IEEE Journal of Biomedical and Health Informatics 25.9 (2021): 3332-3339.

[18] Chetoui, Mohamed, and Moulay A. Akhloufi. "Explainable end-to-end deep learning for diabetic retinopathy detection across multiple datasets." Journal of Medical Imaging 7.4 (2020): 044503.

[19] Gu, Hao, et al. "Deep learning for identifying corneal diseases from ocular surface slit-lamp photographs." Scientific reports 10.1 (2020): 1-11.

[20] Gour, Neha, and Pritee Khanna. "Multi-class multi-label ophthalmological disease detection using transfer learning based convolutional neural network." Biomedical Signal Processing and Control 66 (2021): 102329.

[21] Li, Ning, et al. "A benchmark of ocular disease intelligent recognition: One shot for multi-disease detection." International Symposium on Benchmarking, Measuring and Optimization. Springer, Cham, 2020.

[22] Gargeya, Rishab, and Theodore Leng. "Automated identification of diabetic retinopathy using deep learning." Ophthalmology 124.7 (2017): 962-969.

[23] Li, Ning, et al. "A benchmark of ocular disease intelligent recognition: One shot for multi-disease detection." International Symposium on Benchmarking, Measuring and Optimization. Springer, Cham, 2020.

[24] Roy, Abhijit Guha, et al. "ReLayNet: retinal layer and fluid segmentation of macular optical coherence tomography using fully convolutional networks." Biomedical optics express 8.8 (2017): 3627-3642.

[25] Hu, Kai, et al. "Retinal vessel segmentation of color fundus images using multiscale convolutional neural network with an improved cross-entropy loss function." Neurocomputing 309 (2018): 179-191.

[26] Jordi, C. C., N. D. R. Joan Manuel, and V. R. Carles. "Ocular disease intelligent recognition through deep learning architectures." Universitat Oberta de Catalunya: Barcelona, Spain (2019).

[27] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556 (2014).

[28] Szegedy, Christian, et al. "Going deeper with convolutions." Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.

[29] Szegedy, Christian, et al. "Inception-v4, inception-resnet and the impact of residual connections on learning." Thirty-first AAAI conference on artificial intelligence. 2017.

[30] Chen, T., and C. Guestrin. "XGBoost In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining-KDD ‘16." (2016).

[31] Read, Jesse, et al. "Classifier chains for multi-label classification." Machine learning 85.3 (2011): 333-359.