Novel Approach for Automatic Grading of Hyperemia in Four Different Classes as per IER using VGG16

Sherry Verma[1], Latika Singh[2], Monica Chaudhary[3]
School of Engineering and Technology, Sushant University, Gurgaon. India[1,2]
Sushant School of Health Science, Sushant University, Gurgaon, India[3]

Article History: Received: 13 March 2020; Accepted: 05 August 2020; Published online: 28 August 2020

Abstract: Hyperemia is a condition associated with conjunctiva of the eye. The work in this paper focuses on evaluating the redness of bulbar conjunctiva of human eye. The condition can be a result of numerous factors like injury to eye, friction due to dust particles, improper or prolonged use of contact lenses and it can also be an indication of serious condition like glaucoma. It is evident that if this condition is not given timely treatment then it can result into irreversible eye damage. Several methods have been proposed in past to grade hyperemia but due to human intervention involved the results were subjective in nature. This work proposes the use of pre trained convolutional neural network like VGG16 (named after Visual Geometry Group) to analyze and evaluate the degree of redness. The different classes of redness are in accordance with the Institute for Eye Research (IER) earlier referred as Cornea and Contact Lens Research Unit (CCLRU) grading scale which is universally accepted. The level of redness can further aid in specific line of treatment depending upon the ocular disease related if any. A hybrid model is proposed to automatize the entire process to achieve the desired objective.

Keywords: Convolutional Neural Networks, Hyperemia, CCLRU, VGG16, Bulbar Redness

I. INTRODUCTION

Healthcare sector is amongst the numerous areas which are being benefitted by the advancements of technology. The situation of healthcare industry further got revolutionized in developed nations after digitization. With concepts like telemedicine, teleconsultation lot of time is saved, and timely and hassle-free patient care is being provided to the patient. On the other hand, the remote rural areas of developing nations are stills struggling to receive efficient and timely patient care. Inadequate number of health care workers, insufficient resources are amongst the several contributing factors for limited access to better health care facilities in these countries. The poorer sections of these areas are worst affected and in countries like India approximately 70% of the population resides in these areas. It has been reported in [1] that 39% of primary health care centers in Jharkhand, India doesn't even have adequate medical staff. Hence there are several health care related issues that need attention. One of them is eye care. In rural areas of developing nations like India, the ratio of optometrists to patient is very skewed. Improved medical facilities can be provided to people by integrating technology with health care. The work proposed in this paper proposes an automated approach to categorize hyperemia in the bulbar conjunctiva of the human eye into four different classes. Conjunctiva is a thin membrane covering the entire face of the eye as well as the inner area of eyelids. The membrane is further classified into three categories namely palpebral conjunctiva, bulbar conjunctiva, and fornix conjunctiva [2]. The proposed work focuses on the redness in the bulbar conjunctiva portion. The role of this membrane is to protect the eye from any external injury. It produces tears to wash off any dust particles that might irritate the eye. The damage to this sensitive membrane can be result of several pathological conditions. These conditions can vary from mild to severe. If not given timely treatment, it can result in irreversible eye damage. The work proposed focusses on a specific condition called hyperemia or injected conjunctiva which is a result of engorged blood vessels in the bulbar region of the eye.

Numerous works have been done in past decade to measure the degree of redness in this area of eye. Earlier work induced redness through solutions and compared the degree of redness with already established reference images

[3] [4]. These images were either captured through cameras or were drawn by artists [5] [6]. The drawback of this approach was the non-repeatable results. Different clinicians could not produce similar results for the same set of eyes. Further several new grading scales were proposed to grade redness like Efron scale [7] and Validated Bulbar Redness scale [8]. These scales had limitations like they were inefficient in quantitative evaluation and could not analyze the digital images efficiently. Due to the subjectivity involved in the earlier work, researchers gave way to objective approach of assessing hyperemia. In one such work image analysis was used to categorize redness in two categories [9]. The work incorporated the use of image detection and smoothing techniques to depict redness. The approach could not analyze the eye image of poor sharpness. Later in [10] the suggested work considered number of pixels representing blood vessels and the amount of red color present in the affected area. The work took very few parameters to conclude its results. Further in an automated approach effects of rigid contact lens on hyperemia were established by extracting the region of interest [11]. The work could not give enough statistical significance. Later with the advent of artificial intelligence, machine learning tools were used by researchers to quantify bulbar redness. In [12] the work suggested the selection of best frame from the video of eye images to evaluate redness. Several factors like blurriness, good contrast and lightness were considered to achieve this. The work though gave an automated approach but was not efficient enough for videos captured under varied illumination conditions. Later two methods of artificial neural network like radial basis function (RBF) and multi-layer perceptron (MLP) were used to convert nine extracted features to grading scales [13]. Further in [14] various classification and regression techniques were compared in analysis of hyperemia. The work concluded that though classification methods produce less errors, it is the regression techniques that accurately classify most of the eye images. The machine learning based approaches offered an automatic approach in quantifying redness but few of them did not even considered the entire affected region of conjunctiva. Moreover, the extraction of features was completely dependent on human intervention. In the present study concept of deep learning is used by incorporating the functionality of pre trained convolutional neural network (CNN) like VGG16 to grade the hyperemia in four different classes as per established CCRLU grading scale. The work minimizes human intervention in extracting features to analyze the input image. The remaining sections of paper are organized as follows. Section II briefly highlights the functionality of VGG16 along with basic functioning of CNN. Section III focusses on the methodology and the proposed hybrid model. Section IV and V discusses the outcomes and conclusion.

II. VGG16

Artificial intelligence (AI) is a field of science that works in bridging the gap between humans and machines. The aim of this field is to make machines think, act, and perceive like humans. One of the sub field of AI known as deep learning (DL) deals in networks whose functionality is based on the how the neurons are structured in human brain. Such networks are known as neural networks. CNN is a special class of neural networks which deals in image recognition and classification. There are numerous advantages of these networks over traditional classification algorithms. High accuracy rate, computational efficiency, automatic feature extraction, and flexibility in learning complex features from input image makes this network the most sought-after choice for image classification. This network comprises of various layers namely convolutional layer, pooling layer, and fully connected layer. All these layers work together in compressing the input image in such a from which is easier to analyze and process without losing the prominent features thereby aiding in better classification of the image. The role of convolutional layer is to learn the prominent features of the input image. The features are extracted using filters. The output of this layer is called as convolved feature. The initial layers extract the low-level features and as more and more layers are added and the input of the previous layers are fed to next layers, the network can extract more complex features. The pooling layer aids in reducing the computational power by reducing the dimensionality of the convolved feature. After the understanding of the features the output is flattened and fed to fully connected layer. After a series of forward and backward propagation the model can classify the input image correctly. There are several architectures of CNN available which are already trained on huge data set. Some of them are LeNet, AlexNet, GoogleNet, and VGG16 etc. the proposed work has incorporated VGG16. It is a pre trained CNN which has been trained on ImageNet dataset. This dataset consists 14 million images that can be classified into 1000 categories [15]. Figure 1 shows the architecture of VGG16.

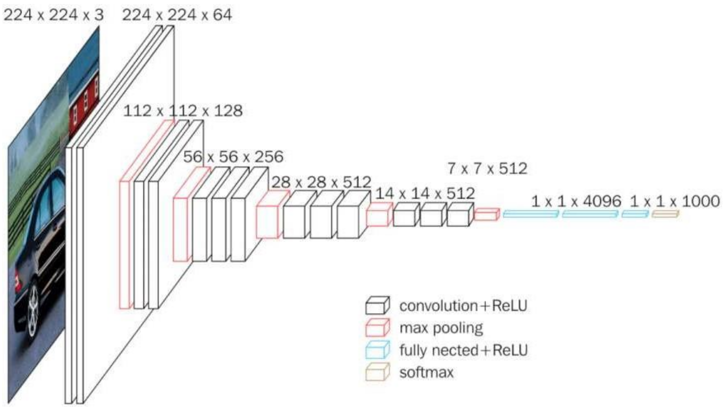


Fig.1. Architecture of VGG16 [16]

Transfer learning, feature extraction, and classification are the three reasons for choosing a pretrained CNN over a custom-made CNN. These networks can be applied directly for any classification problem. Also, they can be used for extracting the feature by using activation as the features. Last but not the least their layers can be used and fine-tuned on a new data set. Because these networks are already trained on huge dataset, they require less training and efforts in building a new model architecture. VGG16 has thirteen convolutional layers and three fully connected layers and an input of 224*224*3 dimensions was fed to this model. Figure 2 shows the stacking of layers of this network.

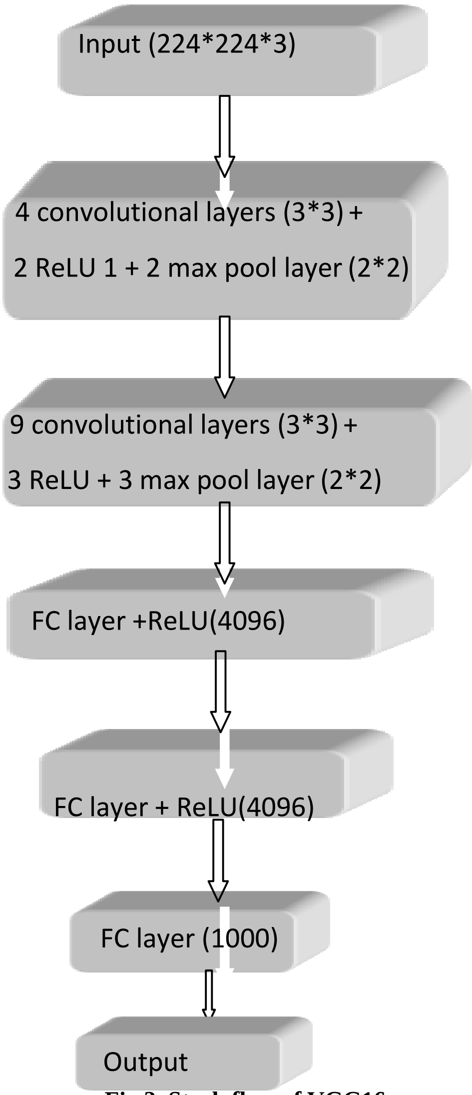


Fig.2. Stack flow of VGG16

A batch size of 256 with 74 epochs was used to train this model Some of the advantages of VGG16 are as follows.

- It helps in minimizing the time taken to train a model
- It offers scaled up performance of the model
- It is an efficient architecture to use as it encompasses huge number of parameters
- It is best suited when the data set is not very large.

VGG16 achieved an accuracy of 92.7% while training on ImageNet with 1000 classes [15].

III. INSTITUTE FOR EYE RESEARCH GRADING SCALE

The degree of redness of bulbar conjunctiva can be categorized into different classes depending upon the severity level [17]. After the condition of an eye is identified as healthy or unhealthy, the next step is to ascertain whether the unhealthy eye needs immediate treatment or not. There are several standards that are being adopted by clinicians to classify the health of an eye into various classes. There are many benefits of using grading scales as a reference like it aids in accurate and repeatable assessment of various eye conditions. Also, it acts as a standard that can be used to minimize the variability in observations. One of the commonly used standards is IER earlier known as CCLRU [18]. It is one of the grading scales which are used to access the acuteness of various pathological conditions including the ones related to contact lens users. This grading scale encompasses the concept of photographs to evaluate the different range of conditions.

This grading scale classifies the health of an eye into four different categories ranging from 0 to 3. The classes represent the severity of the hyperemia in the eye. The first class represents absence of hyperemia whereas the category 3 indicates severe degree of bulbar redness. This stage is an indication of immediate medical attention failing which might result into vision. Figure 3 shows the different classes of redness being assessed by this scale.

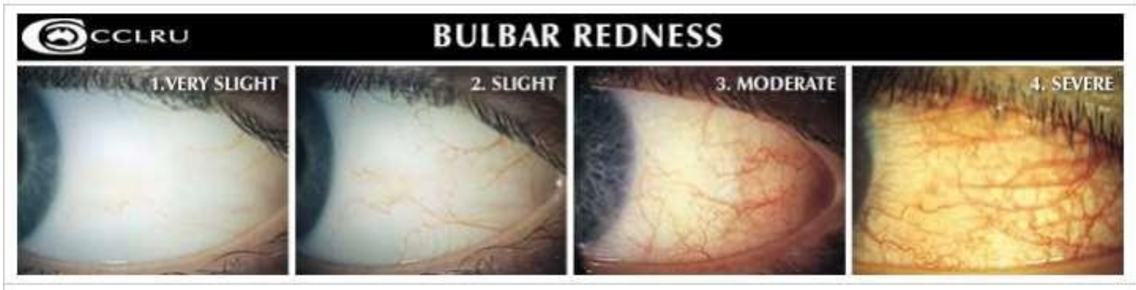


Fig.3. The degree of redness according to CCLRU [19]

The work presented in this paper focusses on grading bulbar redness in four different classes as per established CCLRU grading scale. Th entire process is automatized using combination of pre trained CNN and custom-made CNN.

IV. METHODOLOGY

A. Data Collection

The images were captured using slit lamp (HUVITZ HS 5000). The lamp also had a Canon camera attached to it. These images were taken by Sushant Vision Care Centre, under School of Health Sciences, Sushant University, Gurugram, Haryana, India. The entire set of images was divided into four classes as per the standard IER/CCLRU grading system. Each class consisted of total 42 images each. Table 1 shows the four classes.

Table I: Four classes as per IER scale

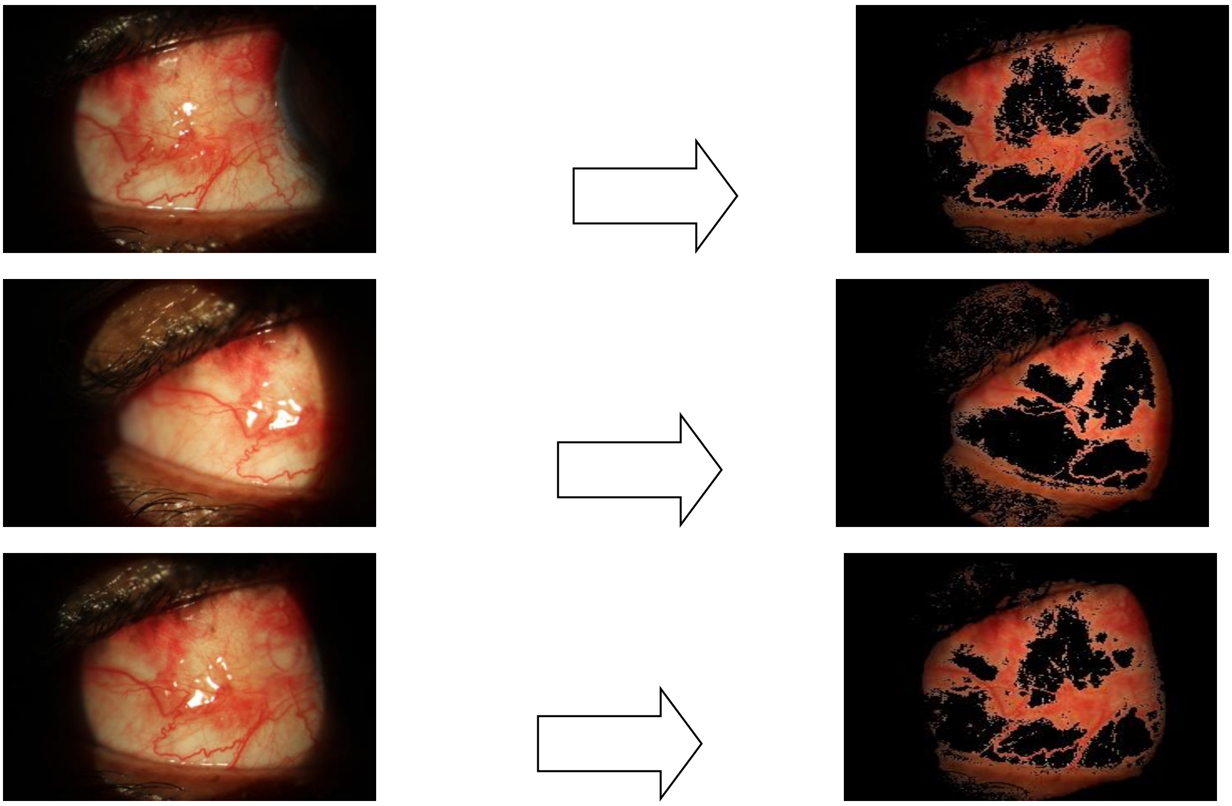
Class	Interpretation
Class 0	The class 0 represented the least affected eye, almost near to healthy eye.
Class 1	Slight
Class 2	Moderate
Class 3	Severe

B. Algorithm for Extraction of Color Gradient from Eye

The color gradient of the eye images was extracted using OpenCV library of python. The extraction was done before feeding the input images to the hybrid model. The aim was to bring uniformity in the dataset by extracting the red color from the eye images. Following algorithm was implemented for extraction of color.

- Converting image from RGB (red, green, and blue) format to HSV (hue, saturation, and value) format.
- Upper mask and lower mask boundaries were defined.
- All pixels were set to zero except those which were like the ones which were obtained after joining the masks.

Figure 4 shows the output of the mask applied on red eye images.



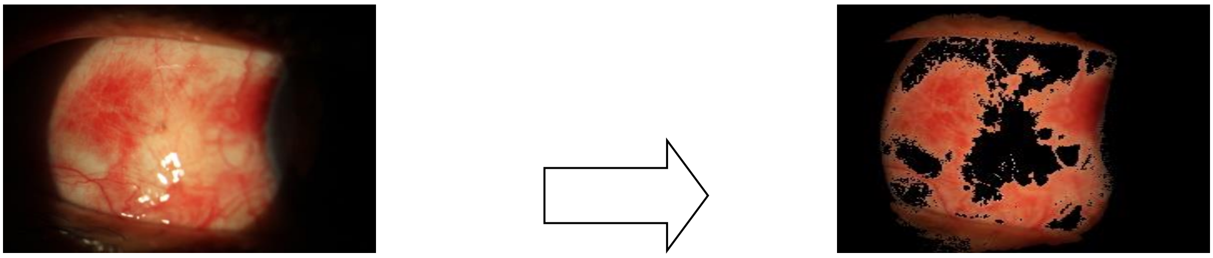


Fig.4. Output of mask application on red eyes

Figure 5 shows the output of applying mask on white eyes.

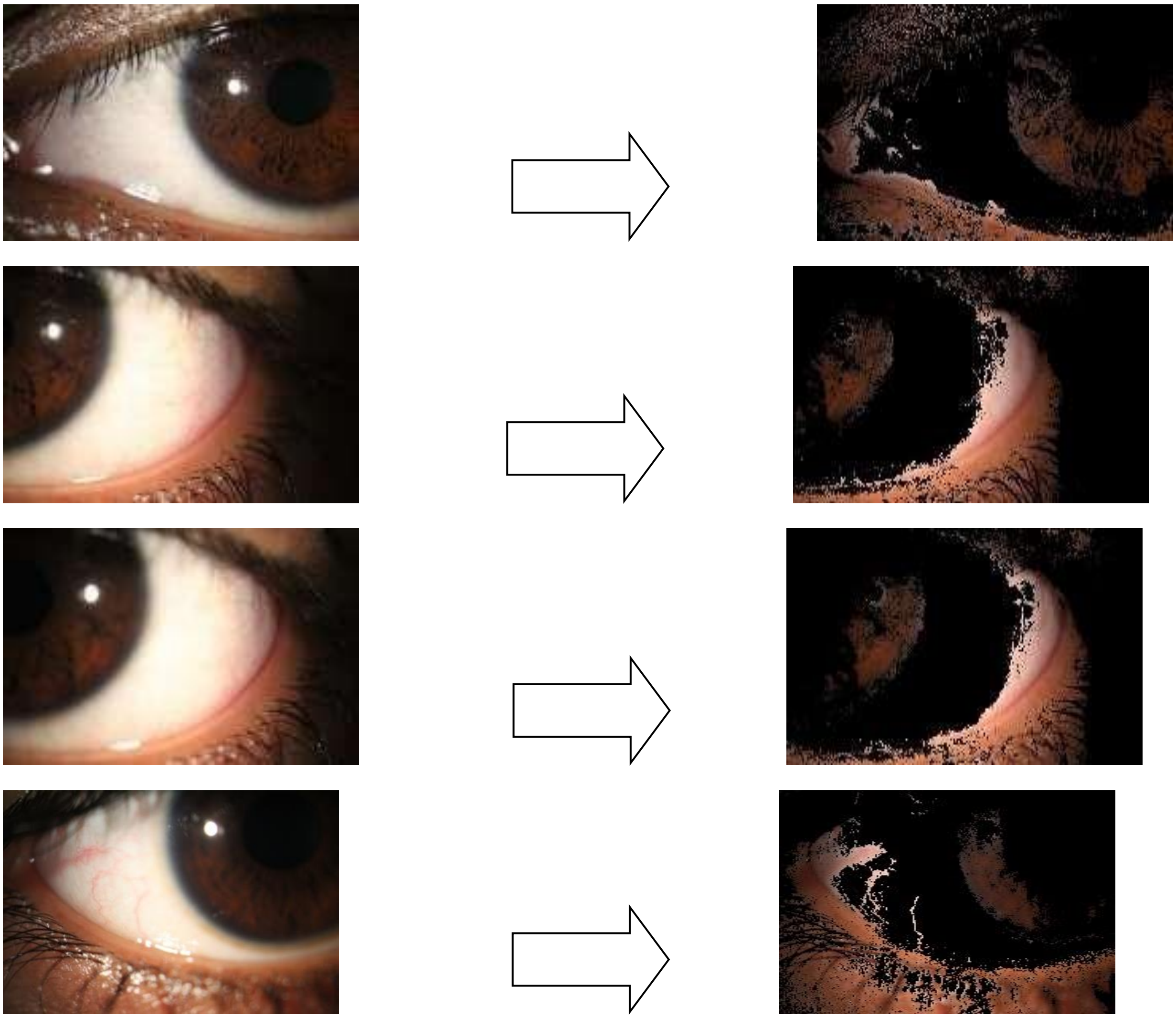


Fig.5. Output of mask application on white eyes

C. Proposed Hybrid Model

The hybrid model consists of three models where the extracted red color images of both red and white color were fed to the first model which is a pre trained VGG16. The fully connected layers of this pre trained network are removed and the parameters are fine-tuned so that it can correctly classify the health of eye. The output of the first model is either healthy or unhealthy eye referred to as benign and baleful. If the output of the model 1 is benign then it is fed further to model 2 which classifies the input into either very slight degree of redness or slight redness. On the other hand, if the output of first model is baleful then it is fed to model 3 to further quantify the degree of bulbar redness. The model classifies input as either moderate or severe. Model 2 and model 3 are custom made CNN which were trained with the input dataset from scratch.

The concept of transfer learning is used in this proposed model. The aim is to have improved learning rate. Figure 6 shows the performance curve of a model with transfer learning and without transfer learning. It is clear from the training curve that the model when trained with transfer learning concept showcases better learning rate thereby giving better accuracy. On the hand the model trained from scratch without incorporating transfer learning has lower learning rate and hence low accuracy.

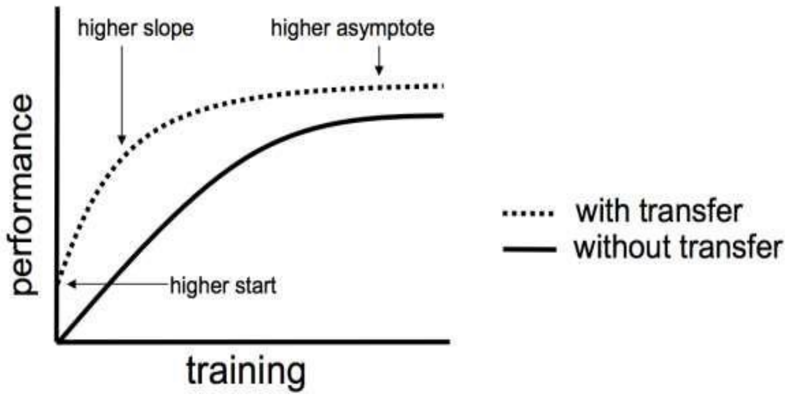


Fig.6. Graph with transfer learning vs without transfer learning [20]

Figure 7 shows the proposed hybrid model comprising of model 1, model 2, and model 3.

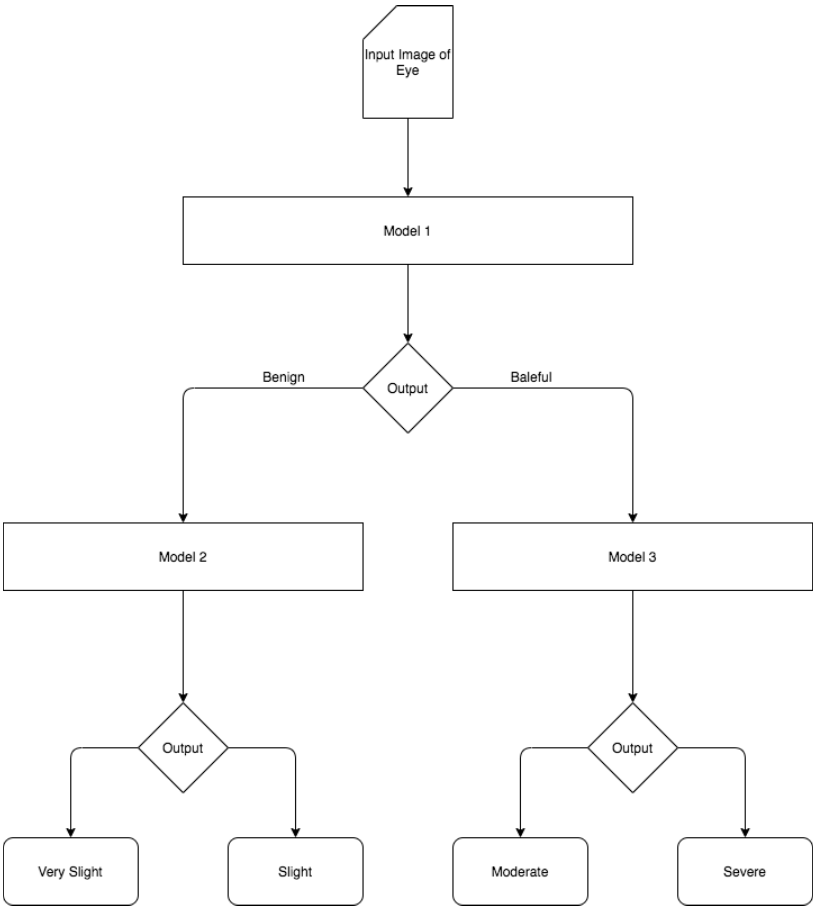


Fig.7. Hybrid Model for classifying health of eye in four categories

D. Hyperparameters of the Hybrid Model

- The first model was the VGG16 which had sixteen weighted layers with 138 million parameters. It was originally constructed to detect objects, so it was fine tuned to achieve the desired objective of proposed work.
- The output layer of VGG16 was removed and two fully connected layers with of 256 and 128 nodes were added.
- Model 2 and model 3 shared similar architecture. Two convolutional layers with twenty nodes and fifty nodes each were used. Kernel of 5*5 and a stride of 1*1 was used along with Rectified linear unit activation function.
- To reduce the dimensionality max pool of size 2*2 with stride of 2*2 was used.
- Lastly fully connected layer of 500 nodes along with softmax function at the output layer was used

The model was trained for forty epochs with batch size of sixteen.

V. RESULTS

Figure 8 shows the output of model-1 and model-2 in classifying the input image as benign with 100% accuracy and then as very slight with accuracy of 99.40%.

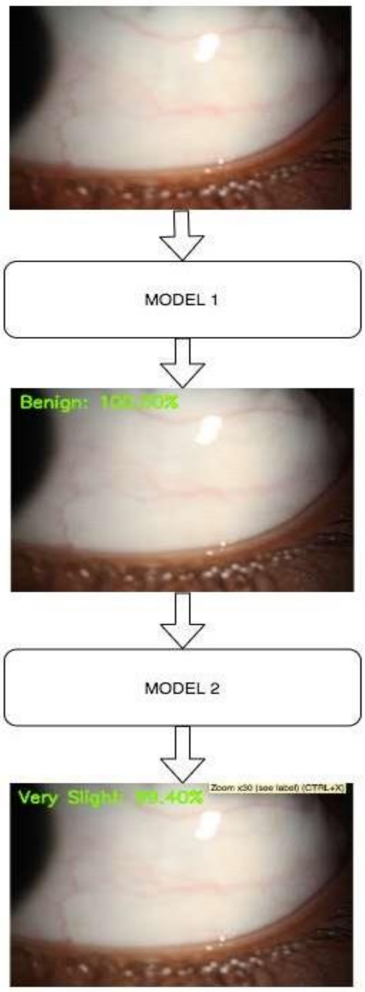


Fig.1. Output of very slight eyes

Figure 9 depicts the classification result of model-1 into benign class with an accuracy of 98.92%. Further the result of CNN based model-2 is shown. Model-2 after taking output of model-1 as input classifies the image as slight degree of redness with a satisfactory accuracy of 100%.

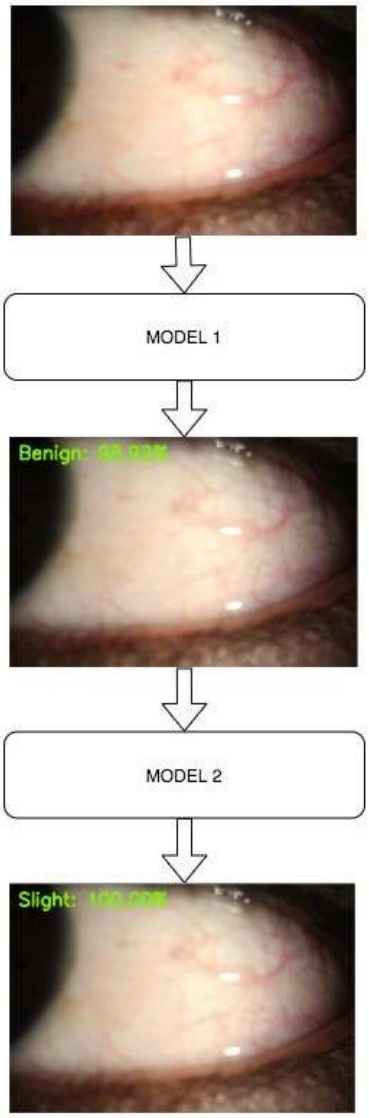


Fig.2. Output of slight eye

Next output shows the performance of proposed model in evaluating the hyperemia classes into baleful. Figure 10 shows the performance of proposed VGG16- based model-1 in categorizing the input image as baleful with an accuracy of 96.14%. Model-3 also showed good results in classifying the image as moderate. The accuracy that this model achieved is 99.99%.

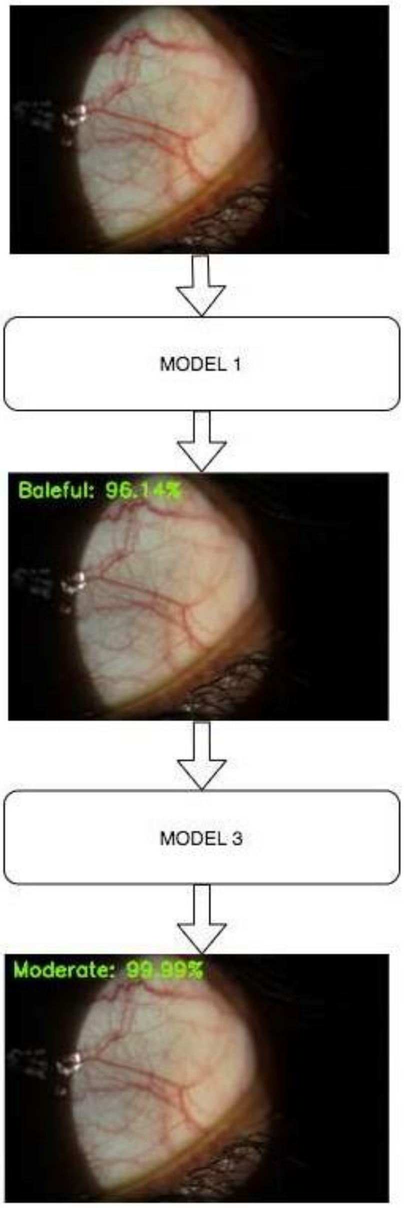


Fig.3. Output of moderate class

Last output as depicted in Figure 11 is regarding the classification of the baleful image as severe. Here the model-1 has identified the input image as baleful with an impressive accuracy of 99.99%. Model-3 on the other hand has also showcased a satisfactory performance with an accuracy level of 99.99% in classifying the input image (which is the output of model-1) as severe. The results of all the three models (model -1, model- 2 and model-3) are quite satisfactory and the proposed work has successfully achieved the objective outlined.

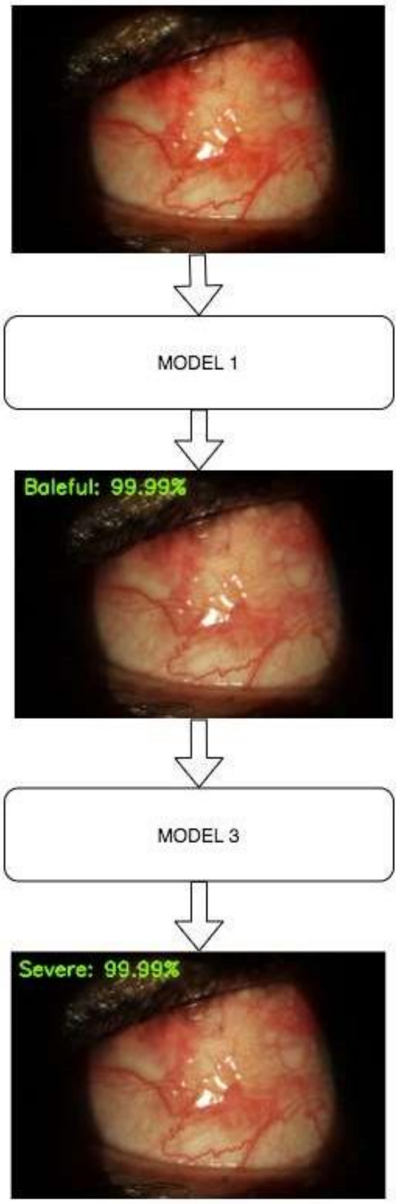


Fig.4. Output of severe eye

The average accuracy of the models proposed in this study is very adequate. The model-1 depicts the average accuracy of 99% in classifying input images as benign and 97% in classifying them as baleful. Further the model-2 showcases an average accuracy of 99% in slight as well as very slight category both. This is quite a satisfactory performance. The model-3 on the other hand exhibits good performance by classifying the images as moderate and severe with an accuracy of 99%. Figure 12 shows the average accuracy of model -1.

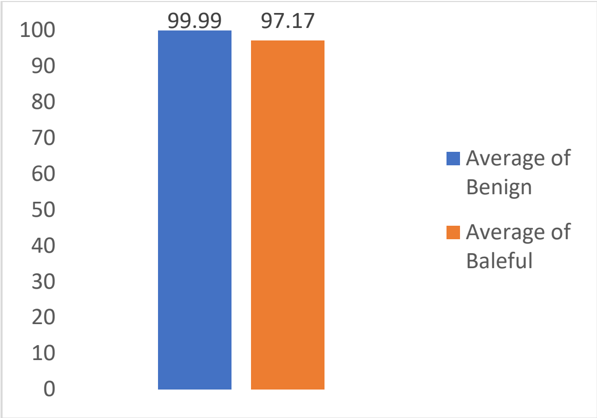


Fig.5. Average accuracy of proposed model 1

Figure 13 and Figure 14 show the accuracy of model2 and model3.

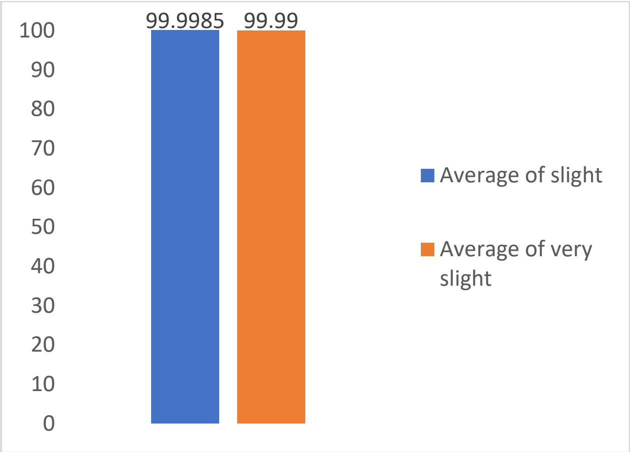


Fig.6. Average accuracy of model 2

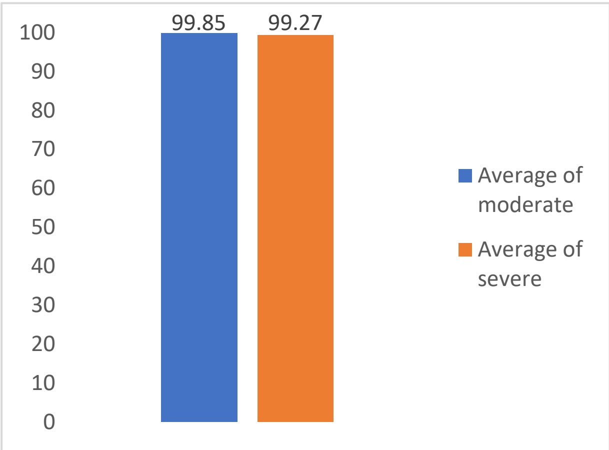


Fig.7. Average accuracy of model 3

VI. CONCLUSION

The purpose of this study was to automatize the process of grading hyperemia based on its severity as per the standard grading scale. The work successfully achieved the results by combing the features of pre trained CNN along with hand crafted CNN. The results clearly indicate the excellent level of accuracy approximately 98% achieved by each of the three models. This work involves minimum human intervention and subjectivity in feature extraction, thereby bridging the gap in the population and number of eye care professionals available especially in rural areas. Further work can be done by considering large dataset from varied eye care institution to obtain more general results.

REFERNCES

[1] D. Srinivasa and Y. Siddegowda, "Rural health care towards a healthy rural India: A social work response", Asian Journal of Development Matters, vol. 12, no. 1, pp. 68- 74, 2018.

[2] M. Doughty, Comparative anatomy and physiology of the cornea and conjunctiva. Ocular surface. LA: CRC Press, 2012, pp. 32-78.

[3] M. Dundas, A. Walker and R. Woods, "Clinical grading of corneal staining of non- contact lens wearers", Ophthalmic and Physiological Optics, vol. 21, no. 1, pp. 30-35, 2001. Available: 10.1016/s0275-5408(00)00011-9.

[4] I. Bailey, M. Bullimore, T. Raasch and H. Taylor, "Clinical grading and the effects of scaling", Invest Ophthalmol Vis Sci, vol. 32, no. 2, pp. 422-432, 1991.

[5] J. Twelker and I. Bailey, "Grading conjunctival hyperemia using a photography- based method", Investigative Ophthalmology and Visual Science, vol. 41, no. 4, pp. S927-S927, 2000.

[6] N. Efron, P. Morgan and R. Jagpal, "The combined influence of knowledge, training and experience when grading contact lens complications", Ophthalmic and Physiological Optics, vol. 23, no. 1, pp. 79-85, 2003. Available: 10.1046/j.1475- 1313.2003.00091.x.

[7] N. Efron, N. Brennan, J. More and K. Rieper, "Temperature of the hyperemic bulbar conjunctiva", Current Eye Research, vol. 7, no. 6, pp. 615-618, 1988. Available: 10.3109/02713688809031818.

[8] M. Schulze, N. Hutchings and T. Simpson, "The Conversion of Bulbar Redness Grades Using Psychophysical Scaling", *Optometry and Vision Science*, vol. 87, no. 3, pp. 159-167, 2010. Available: 10.1097/opx.0b013e3181ce07f1.

[9] J. Villumsen, J. Ringquist and A. Alm, "Image analysis of conjunctival hyperemia", *Acta Ophthalmologica*, vol. 69, no. 4, pp. 536-539, 2009. Available: 10.1111/j.1755-3768.1991.tb02036.x.

[10] F. Willingham, K. Cohen, J. Coggins, N. Tripoli, J. Ogle and G. Goldstein, "Automatic quantitative measurement of ocular hyperemia", *Current Eye Research*, vol. 14, no. 12, pp. 1101-1108, 1995. Available: 10.3109/02713689508995816.

[11] Owen CG, Fitzke FW, Woodward EG. A new computer assisted objective method for quantifying vascular changes of the bulbar conjunctivae. Ophthalmic Physiol Opt. 1996;16:430–437.

[12] L. Sánchez-Brea, N. Barreira-Rodríguez, A. Mosquera-González, C. García-Resúa and E. Yebra-Pimentel, "Automatic Selection of Video Frames for Hyperemia Grading", in *International Conference on Computer Aided Systems Theory*, 2015, pp. 479-486.

[13] L. Sánchez, N. Barreira, H. Pena-Verdeal and E. Yebra-Pimentel, "A Novel Framework for Hyperemia Grading Based on Artificial Neural Networks", in *International Work-Conference on Artificial Neural Networks*, 2015, pp. 263-275.

[14] R. Burgess, "Research shows computers make better medical decisions than doctors", *TechSpot*, 2013. [Online]. Available: https://www.techspot.com/ news/51626-research-shows-computers-make-better-medical-decisions-than-doctors.html.

[15] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition", in *ICLR*, 2015.

[16] "VGG16 - Convolutional Network for Classification and Detection", *Neurohive.io*, 2018. [Online]. Available: https://neurohive.io/en/popular-networks/vgg16/.

[17] P. Murphy, J. Lau, M. Sim and R. Woods, "How red is a white eye? Clinical grading of normal conjunctival hyperaemia", Eye, vol. 21, no. 5, pp. 633-638, 2006. Available: 10.1038/sj.eye.6702295.

[18] Q. Sun, L. Deng, J. Liu, H. Huang, J. Yuan and X. Tang, "Patch-Based Deep Convolutional Neural Network for Corneal Ulcer Area Segmentation", *Fetal, Infant and Ophthalmic Medical Image Analysis*, pp. 101-108, 2017.
[19] IER. IER grading scales. Institute for Eye Research: Sydney, Australia; 2007
[20] J. Brownlee, "A Gentle Introduction to Transfer Learning for Deep Learning", *Machine Learning Mastery*, 2017. [Online]. Available: https://machinelearningmastery.com/transfer-learning-for-deep-learning.