

Detection of Ocular Diseases using Ensemble of Deep Learning Models

Simi Sanya

Computer Science and Engineering dept.
G Narayananamma institute of technology and Science for
women, Hyderabad, India

M Seetha

Computer Science and Engineering dept.
G Narayananamma institute of technology and Science for
women, Hyderabad, India

Abstract— Deep learning models are very competent in image processing and four such models are used to detect diabetic retinopathy (DR), glaucoma and age-related macular degeneration (ARMD). DR, glaucoma and ARMD are the primary causes of vision impairment and blindness in India because indicators for these diseases often appear very late and sometimes cause long-lasting damage to the vision. An automated system put up by means of deep learning models is proposed for identification of these ocular diseases in their initial stages. The very little data acquired from Kaggle is amplified by applying data augmentation in the offline mode. Deep learning models like VGG16, DenseNet201 and Resnet50 are applied to the dataset obtained with various train-test split ratios. VGG16 provides an accuracy of 29.5%, DenseNet201 outcomes an accuracy of 44.5% and ResNet50 provides 93.9% accuracy for 70-30% split of the dataset. An ensemble model is built using VGG16, DenseNet201 and ResNet50 and trained using the same dataset. With both 60-40% and 70-30% train-test splits, the ensemble model provides an accuracy of 72.8% in both the cases. When the deep learning models and the ensemble model are compared with different inputs, it is observed that the ensemble model is more accurate while predicting the fundus image than VGG16, DenseNet201, ResNet50 as the image is processed thrice in the ensemble model.

Keywords-- Deep learning models; Ensemble learning; Data Augmentation; Image processing; Gradio;

I. INTRODUCTION

Visual sicknesses are triggered when there is an injury or damage in the structure of the eye, causing vision deficiency or permanent sightlessness. Diabetic retinopathy (DR) occurs when the patient is diabetic for over an extended period time. The blood vessels in the eye are impaired and the new blood vessels are not formed perfectly leading aneurysms, cotton wool spots etc.,

Glaucoma is initiated when the optic nerve of the eye is damaged due to high pressure build-up of the fluid known as ocular hypertension. Age-related macular degeneration (ARMD), as the name suggests is caused because of the degeneration of macula due to old age. All these diseases sooner or later result in blindness.

Diagnosing these diseases at early stages can prevent permanent vision loss, hence deep learning models are employed to perform detection of these diseases through the fundus images of the eye.

II. LITERATURE SURVEY

Glaucoma is a very common disease seen in the patients and various deep learning models are used or built for detection of glaucoma so far. The distinguishing patterns between glaucoma

images and healthy images are predicted using a CNN with six layers. A drop-out appliance is also useful in this case for attaining passable accuracy with the model [1]. Alternative technique for spotting glaucoma is done using CNN with feature extraction and then the extracted features are paralleled with different datasets. The arrangement of ResNet and Logistic regression works better for this case [2].

The identification of diabetic retinopathy (DR) through eye fundus images is usually performed by ophthalmologists by dilating the pupil by means of chemicals and then review for any delicate features, or injuries in the fundus image. This procedure is tiresome and a long one. A well-organized prediction system based on efficient CNN is engaged for forecasting if the fundus image has DR or not [3].

Age-related macular degeneration (ARMD) is a common threat to vision caused due to the deterioration of the macula due to old age. A network ensemble of six different neural networks is suggested for detecting the disease, with locally collected data, the model delivers an accuracy of 84.3% [4].

Fundus imaging is imperative for ocular and vision related medical diagnoses. The works propose that diabetic or optical diseases can also be anticipated from fundus imaging. In the case of eye diseases, fundus imaging can help early recognition of diseases, enabling defensive measures for imminent blindness or eye health risk [5]. The analyses of data are mostly studied by medical experts. Following the accomplishment of deep learning models in real world applications, they are applied for examining medical data [6].

It is challenging to attain larger datasets of fundus medical images. Hence, preprocessing techniques can be used to increase the current data. Data Augmentation is one of the popular techniques used for data preprocessing [7].

Ensemble learning is the procedure of merging base learners to acquire a combined prediction from one meta learner. There are two types of ensemble learning techniques, they are: homogeneous, heterogeneous learning [8]. Automated systems for spotting DR, glaucoma and ARMD are not in actuality because these diseases are only prominent in India and the datasets are not easily available for them. However, some computerized systems exist for these diseases individually. An automated system is suggested using ensemble of VGG16, DenseNet201, ResNet50 models for carrying out the classification of DR, glaucoma and ARMD based on a dataset of fundus images belonging to all the three diseases.

III. METHODOLOGY

Deep learning models have formerly been used for spotting ocular diseases, but Diabetic retinopathy, glaucoma and ARMD are never considered at once. Hence this model is planned to perform make a diagnosis of the above ocular diseases. The methodology of the model is shown in the figure 1.

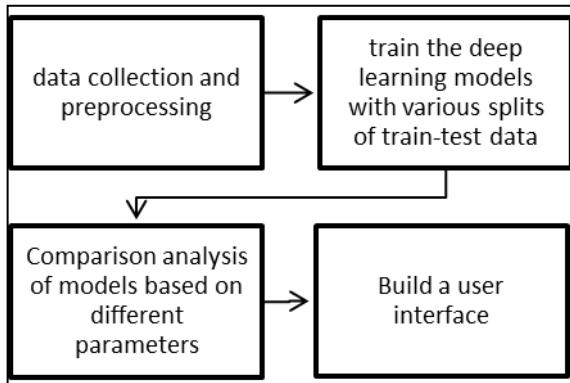


Fig. 1. Methodology of the automated system

The first step is to accumulate the data from numerous sources like kaggle [9] [10] and preprocess the data using different augmentation techniques and applying the augmentation in different modes on the data, before training the deep learning models. The next step contains training the data using different deep learning models like VGG16, DenseNet201, ResNet50 and an ensemble of all these three models.

Ensembling is the process of attaining a final prediction from the predictions of the individual models. Training the models using different splits of data i.e., 60-40% and 70-30% split regulates how the models carry out with different volumes of training data. A user interface is used to compare the models once the training is completed against different types of inputs to check the performance of the models.

The comparative analysis of the models is completed based on various hyper parameters like activation function, learning rate, convolution stride and a best performing model is chosen. The best model is then compared with the ensemble model based on precision which is explained in detail in chapter 4.

A user interface is the most vital part of any computerized system, so an interface is made for the forecasting the ocular disease based on the input fundus image and provide an estimate of the ocular disease. This automated system is built with the intent of providing a faster and accurate provision in detecting eye diseases. The application of the deep learning models is done in python programming language, using Jupyter notebook. Further chapters give the extensive process of implementation and results of the models used.

IV. IMPLEMENTATION

A. Dataset collection and Preprocessing

The dataset for diabetic retinopathy, glaucoma and ARMD is not freely available, so a customized dataset is put together comprising of fundus images related to all three diseases. A total of 1107 images are collected relating to four categories i.e., DR, Glaucoma, ARMD, Healthy fundus images as shown

in the table 1 below, from numerous sources and stored in a single directory. For diabetic retinopathy the “diabetic retinopathy detection -224*224-2019 data” [9] dataset is used, while for glaucoma, age-related macular degeneration and healthy “Ocular disease detection” [10] dataset is used for collecting the fundus images.

The data composed using kaggle is still not sufficient for the deep learning models hence data augmentation (DA) techniques are employed to increase the quantity of the data used for training the deep learning models. Data augmentation is the method of increasing the amount of data by faintly adding alterations to the current data or synthetically generating new data.

Data augmentation can be practically applied in three modes:

- Online: On the fly or real time data augmentation
- Offline: Data is stored on the disc after applying augmentation.
- Combination of online mode and offline mode.

Techniques like flipping, contrast, zooming, brightness, rotation, shear range etc., are applied to the existing data according to the third mode i.e., both offline and online data augmentation. The table 1 shows the quantity of images before and after augmentation is used.

A total of 3225 augmented images are acquired after offline data augmentation and then they are pooled with 1107 original images resulting in a final dataset of 4332 images divided into four categories as shown in table 1. The newly generated dataset is then used to train VGG16, DenseNet201, ResNet50 models where the same techniques are again applied in the online mode thus making the models more robust.

Table I. Dataset Information

| Category of the disease | Original dataset | Augmented images | Total images |
|----------------------------------|------------------|------------------|--------------|
| Diabetic retinopathy | 206 | 801 | 1007 |
| Glaucoma | 352 | 816 | 1168 |
| Age-related macular degeneration | 206 | 810 | 1016 |
| Healthy | 342 | 798 | 1141 |

B. Training the models

Deep learning models train differently for different amounts of data. So, the total data is split into 60% training, 40% testing and 70% training, 30% testing data thus resulting in two different scenarios. Some models require more training data, whereas some require more testing data. The 60-40 and 70-30 splits are shown in the table 2 below.

The more amount of data the model uses to train, the higher the accuracy is obtained, but for making comparisons between models trained on various training and testing data, to observe the performance of the models.

Table II. Train-Test Data split

| Disease | Total images | 60-40% ratio | 70-30% ratio |
|----------|--------------|--------------------------------|--------------------------------|
| DR | 1007 | 60% = 0-604 40% = 605- 1006 | 70% = 0-704 30% = 705- 1006 |
| Glaucoma | 1016 | 60% = 0-609 40% = 610-1015 | 70% = 0-711 30% = 712- 1015 |
| ARMD | 1168 | 60% = 0-700 40% = 701-1167 | 70% = 0-817 30% = 818- 1167 |
| Healthy | 1141 | 60% = 0-684 40% = 685-1140 | 70% = 0-798 30% = 799- 1140 |

The models work differently for the splits and the results vary exponentially in the case of VGG16 and DenseNet201 models. After the preprocessing of the data is completed, the most popular deep learning models, VGG16, DenseNet201, ResNet50 are used to train the data for performing the classification. The results obtained with various splits of the data are detailed in the next chapter.

- VGG16:** VGG16 is an acronym for visual geometric group, it represents a convolutional neural network which is 16 layers deep. It has approximately 138 million parameters and can classify objects into 1000 categories like keyboard, mouse, pencil, animals etc., It consists of an arrangement of max and pool layers consistently followed by two fully connected layers and a softmax activation function. It can be considered as the most used deep learning model for image processing.
- DenseNet201:** DenseNet201 is an acronym for densely connected convolutional neural networks and it is one of the new discoveries in neural networks for visual object recognition, A DenseNet is composed of dense blocks and each layer in the network receives a collective knowledge from all the preceding layers. The outputs from all the layers are concatenated and then sent further to the next layers. This model of DenseNet has 201 layers of convolutional neural network
- ResNet50:** ResNet50 is an acronym for residual convolutional neural networks that is 50 layers deep, A ResNet is composed of residual blocks where the output of the final layer in the network is calculated by combining the outputs of previous second and third layers. The architecture of the deep learning model consists of two main blocks. The first block is the convolution block and the second is called an identity block.
- Ensemble Model:** Ensemble learning is the process of combining different base learners to obtain a combined meta learner for making a prediction based on the input image. In order to perform ensembling the three deep learning models mentioned above are used to obtain a final model as shown in Fig 3.1. In this case the base learners are the deep learning models and the resultant model is called a meta learner. There are two types of ensemble learning, they are homogenous and

heterogeneous. In this case heterogeneous ensemble learning is employed to combine VGG16, DenseNet201, ResNet50 and one final model is built.

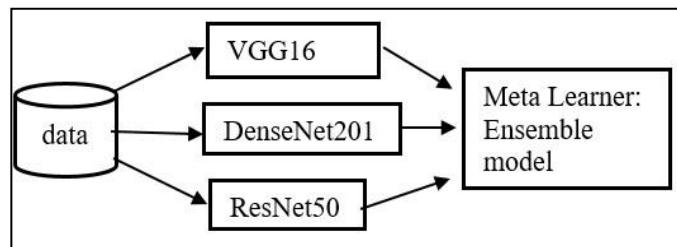


Fig 2. Working of ensemble model

V. RESULTS AND DISCUSSIONS

The following Figure 3 displays deep learning models, number of epochs used to perform the training process, Activation Function, Accuracy obtained using 60-40% train-test split and accuracy with 70-30% train-test split of the data. Based on the accuracies obtained, it is concluded that the ResNet50 and Ensemble model are best trained with the current data. The deep learning models are compared based on different parameters like learning rate, activation function, loss function.

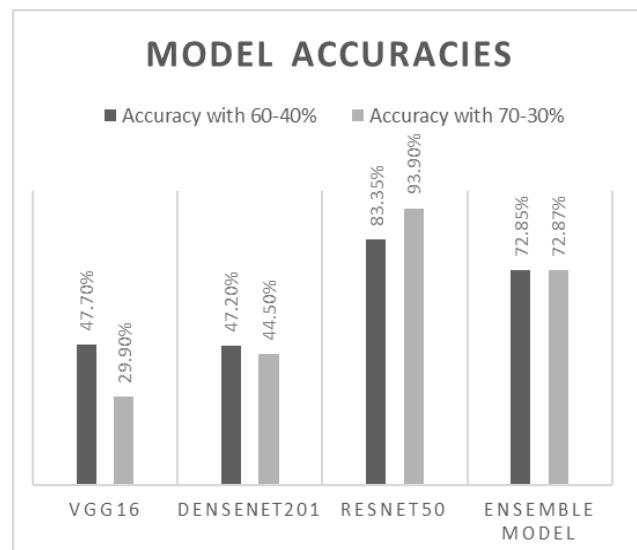


Fig 3. Model Accuracies

Table 3 in this chapter gives the information about the hyper parameters used in all the models, learning rate, convolution stride and the activation function used. These hyper parameters are not necessarily same for all the deep learning models, so the comparison is done, these parameters are not used to compare the ensemble model with the individual deep learning models because, the ensemble model has the advantage of three models and the comparison is not reasonable in this case.

Table III. Comparing Hyper parameters of models

| Models | Learning rate | Activation Function | Convolution Stride |
|--------------|----------------|---------------------|--------------------|
| VGG16 | 2e-5 ≈ 0.00002 | Sigmoid | 1 |
| DenseNet 201 | 0.1 | Softmax | 2 |
| ResNet50 | 2e-5 ≈ 0.00002 | Sigmoid | 2 |

ResNet50 and Ensemble models are used for the user interface and compared based on precision. Precision can be calculated as shown in the equation (1).

$$\text{Precision} = \frac{\text{True positive}}{\text{true positive} + \text{false positive}} \quad (1)$$

The Ensemble model and the ResNet50 are compared based on precision individually. For 15 different inputs, the models provide the accuracies as shown in table 4.

Table IV Comparison of ResNet50 and Ensemble Models Based on Precision

| Model | Precision |
|----------------|-----------|
| ResNet50 | 0.33 |
| Ensemble Model | 0.66 |

The table 4 shows that, even though with less accuracy, the Ensemble model is more robust than the ResNet50 in terms of precision, i.e., The Ensemble model predicts the input fundus image more accurately than the ResNet50 model.

VI. USER INTERFACE

The user interface for the models is built using a python module called gradio, which has an option of customizing the input fundus image. The figure 4 shows an example of the user interface where ensemble model is used for performing the classification of the input fundus image.

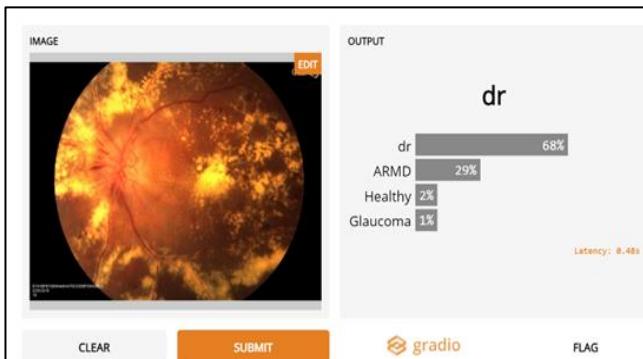


Fig 4. User interface built using the Ensemble model identifies the image correctly as DR.

The user interface provides many options of customization and also multiple models can be used for making comparisons. The figure shows that the model identifies the image as diabetic retinopathy with 68% accuracy.

VII. CONCLUSIONS

Ocular diseases often show painless and late symptoms and are not diagnosed till the advanced stage. Three deep learning models like VGG16, DenseNet201, ResNet50 are used for building an automated system, that can be used as a support for diagnosing diabetic retinopathy, glaucoma, age-related macular

degeneration. The deep learning models trained on the customized data are then compared based on hyper parameters like learning rate, epochs, convolutional stride, activation function etc. It is evident that the ResNet50 model is strongest among all three models because it has the maximum accuracy when the models are tested in the user interface.

ResNet50 and ensemble model are equated with respect to precision by calculating true and false predictions for different types of input. Ensemble model gets a precision value of 0.66 while ResNet50 results a precision of 0.33. Based on the precision obtained, it is determined that, ensemble model outperforms ResNet50. The work can be extended further by collecting more data from local hospitals all over India and build a large enough dataset for the models.

ACKNOWLEDGMENT

A special appreciation for the constant support and the seed fund provided by “GNITS”. This work is deemed purely as research on how artificial intelligence, deep learning is incorporated for diagnosing ocular diseases and the performance of the models employed for diagnosing diabetic retinopathy, glaucoma and age-related macular degeneration. Professional personnel need to be consulted for diagnosis of these diseases or for the purpose of treatment

REFERENCES

- [1] Risha Shetty, Sakshi Jain, Sejal D'mello, Aniket Patil “Glaucoma Detection using Convolutional Neural Network” International Research Journal of Engineering and Technology, 7(01): 2395-0056, 2020.
- [2] Roberto Matheus Pinheiro , Alan Carlos de Moura Lima, Lucas Bezerra Maia, Pereira, Geraldo Braz Junior, Jo ´ao Dallyson Sousa de Almeida, Anselmo Cardoso de Paiva “Glaucoma Diagnosis over Eye Fundus Image through Deep Features” in IEEE, 65080-805, 2018.
- [3] Cyril Leung, Jiaxi Gao, Chunyan Miao “Diabetic Retinopathy Classification Using an Efficient Convolutional Neural Network”, in IEEE international conference on agents, 978-1-7281-4026-1, 2019.
- [4] Caroline Brandl, Felix Grassmann, Judith Mengelkamp, Sebastian Harsch, Martina E. Zimmermann, Birgit Linkohr, Annette Peters, Iris M. Heid, Christoph Palm, Bernhard H.F. Weber “A Deep Learning Algorithm for Prediction of Age-Related Eye Disease Study Severity Scale for Age-Related Macular Degeneration from Color Fundus Photography”, in American academy of ophthalmology, 0161-6420/18, 2018.
- [5] https://www.onhealth.com/content/1/common_eye_conditions, visited: 25-04-2021
- [6] Muhammad Imran Razzak, Saeeda Naz and Ahmad Zaib “Deep Learning for Medical Image Processing: Overview, Challenges and the Future”, In book: Classification in BioApps (pp.323-350) , 2018.
- [7] https://en.wikipedia.org/wiki/Data_augmentation, visited: 22-05-2021.
- [8] <https://www.analyticsvidhya.com/blog/2021/06/ensemble-deep-learning-an-ensemble-of-deep-learning-models/>, visited: 23-05-2021
- [9] <https://www.kaggle.com/sovitrath/diabetic-retinopathy-224x224-2019-data>, visited: 23-05-2021
- [10] <https://www.kaggle.com/andrewmvd/ocular-disease-recognition-odir5k>, visited: 22-04-2021.