Classification of Eye Diseases using Hybrid CNN-RNN Models

MSc Research Project
Data Analytics

Mayuresh Londhe
Student ID: X2013726

School of Computing
National College of Ireland

Supervisor: Dr. Christian Horn

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Mayuresh Londhe		
Student ID:	X20137265		
Programme:	Data Analytics	**Year:**	2020/2021
Module:	MSc. Research Project		
Supervisor:	Dr. Christian Horn		
Submission Due Date:	16/08/2021		
Project Title:	Classification of Eye Diseases using Hybrid CNN-RNN Models		
Word Count:	**7835**	**Page Count**	**20**

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: ……………………………………………………………………………………………………

Date: ……………………………………………………………………………………………………

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Classification of Eye Diseases using Hybrid CNN-RNN Models.

Mayuresh Londhe
X20137265

Abstract

Early diagnosis of eye diseases is essential to prevent irreversible vision loss. The traditional method used by ophthalmologists is to manually screen, images captured of back of an eye called as fundus images. Patient care is being harmed by an increase in the number of patients and a scarcity of qualified ophthalmologists. This study focuses on classifying fundus images into different types of eye illnesses: cataracts, glaucoma, and retinal diseases. The Convolution Neural Network-Recurrent Neural Networks (CNN-RNN) model has proven to be effective in a variety of disease classification challenges. The advantages of Transfer Learning and Long Short-Term Memory (LSTM) are combined in the proposed hybrid CNN-RNN models. Multiple different features are extracted from fundus images by Transfer Learning models, InceptionV3, InceptionResNetV2, and DenseNet169. The extracted features are classified using LSTM. The research is based on the Kaggle dataset, which comprises an imbalance number of images in each category. Therefore, augmentation approaches are used to balance the dataset. The performance of the models improved when trained and tested on increased and balanced data. The hybrid DenseNet169-LSTM model achieved the highest accuracy of 69.50% (87.40% specificity and 69.50% sensitivity).

Keywords: CNN-RNN, Multi-class classification, Transfer Learning, Long Short-Term Memory, Eye Diseases. K-fold Cross Validation.

1 Introduction

Sight, sound, smell, taste, and touch are the five primary senses used by the human body. Organs that help the human body achieve these senses are equally vital. These organs provide the ability to perceive the environment, out of this 80% of perception comes from the sense of Sight. Because of the importance of visual perception, it is vital to protect the eyes from diseases that could impair vision. The most commonly diagnosed diseases that cause serious vision impairment are cataracts, glaucoma, and retina diseases. According to a study by WHO (World Health Organization) common causes of blindness are 1) Cataract (47.8%) 2) Glaucoma (12.3%), and Retina Diseases (4.8%) (Acharya et al., 2006). These three types of diseases are responsible for more than 60% of the cases of blindness in humans.

Cataracts are a disease that is associated with the lens of the eye. The formation of the cloudy area in the lens is an indication of Cataract formation. Initially, it will be small in size which gradually increases over time resulting in loss or blurry vision depending on the scale of the Cataract size. Glaucoma is an eye condition affecting the Optic Nerve. This Nerve forms the connection between the eye and brain, damage to this nerve causes vision loss. Glaucoma must be detected early since visual loss occurring in the later stages is irreversible. The retina, which is a nerve layer at the back of the eye that is responsible for light sensing, is the photosensitive component of the eye. Vision loss occurs when this part of the eye is damaged. If retina diseases are not diagnosed early on, it may take longer to regain eyesight, which may necessitate repeated treatments. Figure 1, below shows the fundus images of a normal eye, an eye with cataracts, and those with glaucoma and retinal disease.

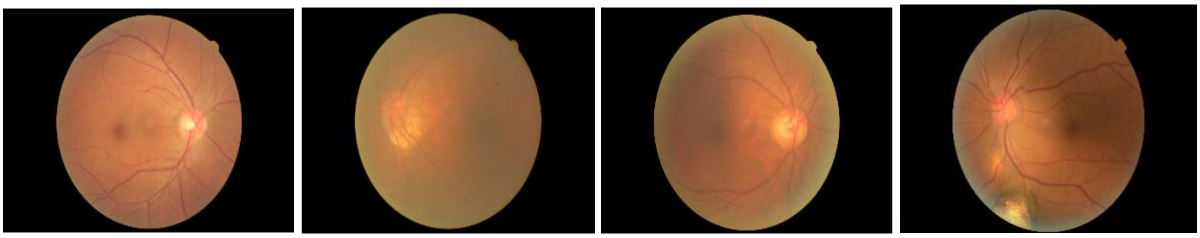

Figure 1: Normal Eye, Cataract, Glaucoma and Retina Disease Fundus Image from (L to R)

1.1 Research Motivation and Background

According to (Dong et al., 2017), the Republic of China has 5 million cataract cases, which is increasing at a rate of 1 million cases per year. The figures from one country represent the global magnitude of the numbers. Although one million cataract procedures are conducted free of charge each year, 1.5 million cases of cataract are reported annually among the indigent population (Acharya et al., 2006).

Diabetes, which leads to Diabetes Retinopathy, is one of the most common causes of retinal disorders. According to the World Health Organization's survey, there would be 435 million cases of diabetes worldwide by 2030 (Lyona et al., 2020). also report presented by International Diabetes Federation (IDF) states that, 415 million adults suffered diabetes worldwide and it is predicted to be reached at 645 million cases by the year 2040 (Nugroho et al., 2017). These growing numbers of diabetes patients would result in a greater number of patients suffering from retina diseases.

Due to the unreliability of traditional automatic detection approaches, eye disease screening is currently done manually by an ophthalmologist. As a result, patient care is delayed (Omar et al., 2017). Manual diagnostic procedures used by an ophthalmologist are error-prone due to human intervention (Zaheer et al., 2019). A rising number of common eye diseases is a cause for concern; however, these diseases often develop with minimal symptoms at first. As a result, accurate early identification of eye diseases becomes extremely difficult. Fundus image examination carried out by hand is laborious and time-consuming. There is a scarcity of skilled experienced ophthalmologists in rural areas. Therefore, the development of automatic eye disease detection is necessary, which would result in faster diagnosis and improved accuracy (He et al., 2021).

Various approaches for resolving the problem of disease identification have been presented by researchers. One type of research focused on identifying individual disease, while another focused at distinguishing different stages of individual disease based on severity. Recent studies have performed the multi-label and multi-class classification of eye diseases. The paper is based on multi-class classification of 4 classes. Researchers initially opted traditional machine learning algorithms such as Support Vector Machines (SVM), K-Nearest-Neighbours (KNN), K-Singular Value Decomposition (K- SVD) and Feature Vectors. Neural Networks and Deep Learning was considered to extend and improve the results of traditional machine learning algorithms. Although, Ensemble of multiple deep learning models (Jiang et al., 2019), Dense Correlation (DC) network (Li et al., 2020) models achieved promising results for classification. Accurate classification of images is still a challenging task with regards to classification of multiple eye diseases.

Therefore, this domain needs a solution that is robust and accurate when classifying multiple eye diseases. To achieve the goal, this paper presents a new approach for eye disease classification using the Hybrid CNN-RNN models. This would be integrating the advantages of CNN and RNN. Here CNN would act as a Feature Extractor for the Fundus Images and RNN would be used as a classifier.

1.2 Research Question

How well can the Hybrid CNN-RNN models classify the fundus images of the eye into 4 classes Normal, Cataract, Glaucoma, Retina Diseases?

1.3 Research Objectives

Objectives	Description	Metrics
1	A critical review of existing Eye Disease classification methodologies.	
2	Data and Image Pre-Processing, Data Augmentation.	
3	Implementation of CNN for Feature Extraction on balanced and Imbalanced dataset.	
4	Implementation of RNN for Eye Diseases classification on balanced and Imbalanced dataset.	
5	Comparison of CNN-RNN models on Balanced and Imbalanced Dataset.	Accuracy, Specificity, Sensitivity, Loss
6	Comparison of CNN-RNN model with models implemented for Multiclass Classification by other research.	Accuracy, Specificity, Sensitivity, Loss

Table 1: Research Objectives

1.4 Plan of Paper

Section 1 describes the Introduction and Motivation of the Research Topic; Section 2 explains the critical review of Related Work; Section 3 explains the Research Methodology. Section 4 describes Design Specification. Section 5 contains Implementation and Evaluation. Section 6 explains the comparison and results. Section 7 discusses the conclusion and future work.

2 Related Work

This section of the paper provides the critical reviews of paper related to Eye diseases classification. Each Subsection discusses different group of methodology or techniques with their results.

2.1 Classification of Eye Disease using Machine Learning

Qiao et al. (2017) presented a model for cataract and non-cataract image classification that combines the advantages of SVM and Genetic Algorithm. Fundus images were divided into 16 blocks by applying segmentation. Colour features of the images were extracted using Histogram Equalization, which captures the grey levels and dynamic range for each pixel. Gray Level Co-occurrence Matrix (GLMC) is used to perform the extraction of texture features. An angular second moment, contrast, entropy, correlation, and inverse difference moment are computed by GLMCs at an angles of 0,45, 90, and 135 degrees. The third feature extractor employed was Haar wavelet to translate the image into the frequency domain. A genetic algorithm is used to apply weighing to the retrieved features. The Support vector machine receives the output of the Feature weighting as input for classification. The research achieves 95.33 % accuracy for Normal and Abnormal classes and 87.52% for Normal, Mild, Moderate, and Severe classes. The inclusion of a genetic algorithm for weighing increases the time required to compute and classify data.

Omar et al. (2017) proposed a method for classifying diabetic retinopathy. Vessels, Haemorrhage, Blood Vessels, Exudates, and Optic disc are among the features derived from the fundus images. The intensity, size and presence of these features are used to classify the images into four categories of mild Non-Proliferative Diabetic Retinopathy (NPDR), moderate NPDR, severe NPDR and Proliferative Diabetic Retinopathy (PDR). To train the model, 49 images were used, while 89 images were used in the testing phase. From the training to testing phase, algorithm improved the accuracy rate from 86% to 98% respectively. Positives from this research are that there is no overfitting of the data. The model was built and tested with a small number of images; however, the algorithm would be inefficient in extracting features when the images are not focused, which is an aspect where it could be improved.

Researchers have separately employed Segmented and Non-Segmented fundus image features, but Purandare and Noronha (2016) implemented a model that extracts both types of features from an image. Contrast, Correlation, Homogeneity, Energy, and Entropy are non-segmented characteristics of the images, whereas Exudates, Blood vessels, and Optic Disc are segmented features. The output of the feature extraction is fed into a Support Vector Machine with 10-fold cross-validation with three independent kernels: Polynomial, Radial Bias Function, and Linear. The model acquired a significant specificity of 96.00% but failed to achieve a similar result for the sensitivity of 78.00%, which is a point of concern because the model is unable to correctly classify images of Diabetic Retinopathy class.

KNN is a machine learning algorithm that can be applied to classification problems. The model trains itself based on the supplied value of nearest neighbours. Pandey, Patre and Minj (2020) have implemented KNN for classifying diseases pertaining to Glaucoma. Feature extraction, image segmentation using global thresholding, and Region of Interest extraction are the initial phases in the model. The radius of the optic cup and the radius of the disc are identified and utilised to train the model.

2.2 Classification of Eye Disease using Transfer Learning

The first section of the related work is comprised of a review of the classification models which has implemented traditional machine learning algorithms. This section contains the research work which has considered the approach of Transfer Learning for the classification of eye diseases.

He et al. (2020) proposed a methodology that combines feature extraction, bilateral and unilateral weighing, and fusion. The method comprises three distinct modules for each task, with ResNet being used in the Feature Extractor Module. Its output was passed on to the Feature Fusion Module, which uses three different forms of attention weights: unilateral, bilateral, and both. Weights are also added, concatenated, and multiplied to create a fusion of the weights. The classification of these feature vectors is performed by the final module. The dataset contains images for eight labels, including patients with multiple eye diseases. The model is using 1,167 images and 3-fold cross-validation for multi-label classification. For model evaluation, Kappa, AUC, and F1 Score are used. The accuracy of the model with Res-Net101 as a feature extractor is 93.40 %. Although the model has attained substantial accuracy, a model of this complexity has the disadvantage of being less interpretable.

CNN's are considered to be the most effective in extracting features from image data. The application of CNN for feature extraction was considered by Wang et al. (2020). Efficient-NetB3 was used as a feature extractor in the proposed model. The classification was performed with a custom-built neural network block that contains two convolution layers, seven Memristive Binary Convolution (MBCNN) layers with average pooling, dropout, dense, and sigmoid layer. This approach has produced excellent results, but there are a few issues that might compromise its efficacy. To begin with, the data comprises images from eight different labels, however, they are all imbalanced. The architecture of a neural network is such that it learns features on its own, thus interpreting the features learnt becomes difficult.

Bulut et al. (2020) presented a methodology for disease classification that include 21 different diseases. The method implemented here is Xception with multiple hyper-parameters tuning that includes Global Average Pooling, Dropout layers, batch size of 128 and learning rate of 0.001. Researchers made the appropriate use of serialization as the dataset contains a large number of images the size of the dataset would be higher therefore images were loaded in chunks of 100-200 MB. Although the dataset comprises 9565 images, it is extremely imbalanced. The impact of the imbalance dataset is evident in the results obtained during the evaluation of training and testing phases. In training, accuracy was 91.00%, but in testing, it decreased to 81.00%.

Das et al. (2019) proposed a new method for feature extraction and classification that leveraged the VGG19 model. Image data is augmented by flipping, cropping, and rotation before training the model for Feature Extraction and classification. This not only helps to overcome the problem of using a small dataset, but it also helps to improve the model's robustness. Epochs ranging from 25 to 150 with an increment of 25 and a learning rate of 0.1,0.01,0.05,0.001 are used to fine-tune the model. Accuracy and Loss are the two evaluation methods considered. The best accuracy of 93.50% was reached at the 150th epoch and 0.001 learning rate, out of all the values chosen for learning rate and epochs.

A comparative study was conducted by Smaida and Serhii et al. (2019) between the transfer learning model and the custom-built CNN model. Models were trained to distinguish between

4 classes, with 3 different eye diseases and one normal category. In comparison, transfer learning models InceptionV3 and VGG16 achieved 81.00% and 79.40% accuracy respectively, the accuracy of the custom CNN model was the lowest at 53.10%. Models were able to correctly classify images of Diabetic Retinopathy, Myopia, and Normal class, but not images of Myopia. Sarki et al. (2020) implemented the VGG16 and InceptionV3 models to classify eye diseases into two categories: mild and early-stage. Multiple datasets were combined for this purpose, resulting in two datasets with normal stage and mild stage multiple diseases. Four different optimizers were used to train the models: Adam, RMSprop, SGD, and Adagrad. On both datasets, the VGG16 model with Adam optimizer attained the highest accuracy of 84.88% and 84.42%, respectively.

Ensemble learning is a technique that helps us achieve an improved result by training multiple models at one time and by amalgamating results achieved to make a decision. For this domain, the classification of labels to their respective classes would be decided by a vote of all the models that have been developed. By overcoming the limitations of individual models, the model's accuracy and robustness get improved. This method was used to classify Retina Diseases by Jiang et al. (2019). InceptionV3, Resnet152, and InceptionResNetV2 were the pre-trained models applied in this study. The integrated model achieved higher accuracy than the individual models. The drawback of such a model is that it requires extremely high computational power also, model's interpretability gets hampered.

2.3 Classification of Eye Disease using Deep Learning

Above section 2.2, gives review of research on pre-trained transfer learning models. This section presents the studies which are done in the field of Deep Learning.

Ahmad and Hameed (2020) developed a model for the hierarchical classification of eye diseases that incorporates Artificial Neural Networks (ANN). The colour histogram-based and texture-based feature extraction techniques are used in the pre-processing stage. The Hue Minimum Maximum Difference (HMMD) colour space was used to extract colour-based characteristics at the pixel level. The weights for each layer of the ANN are selected from the range of (1 to -1) and (0.5 to -0.5). As this dataset includes non-fundus images, using this approach to Fundus images will be interesting.

The study presented by Lin J., Cai and Lin M (2021) combined the advantages of Graph convolution networks and self-supervised learning for multi-label classification of fundus images. A multi-class graph network contains seven layers that perform feature extraction and classification. The researchers were able, not able to obtain significant results with this methodology. Zaheer et al. (2019) also considered ANN for the classification of eye diseases. The dataset comprises 402 images which consist of 37 features and 41 disease labels. Three different activation functions were used for training the neural network sigmoid, Gaussian and Arctan.

Berrimi and Moussaoui (2020) implemented a deep learning model that outperforms the pre-trained transfer learning models. 3 Convolution layers are included in the proposed architecture, which are followed by a Max Pooling layer, Dense layer, and Output layer. The following hyperparameters were used: Adam Optimizer with 80 batch size, 15 epochs, 0.001 learning rate, and an early stopping factor of 10. The Dropout and Batch Normalization layers were added to this architecture to improve it even further. With optical coherent tomographical images, Bhadra and Kar (2020) implemented a similar method. The proposed algorithm consists of Set of two convolution layers followed by a Max Pooling layer along with Adam optimizer. As the images are obtained from topographical scans and research has achieved significant results, architecture can be applied to fundus images as well.

A model composed of a DC network and a spatial correlation module was proposed by Li et al. (2020). The DC network acts as a feature extractor, while the Spatial Correlation Module learns the correlation between the extracted features, which are then fused using the concatenation, and the classification layer performs multi-label eye disease classification.

While working on a project that includes multi-class classification, class imbalance and fewer images to train the model are frequently encountered issues. By combining 3 different datasets containing images of 12 different eye diseases, Karthikeyan et al. (2019) were able to obtain significant results. Each of the 5 blocks in the deep learning model included a network of convolution layers followed by a maxpool layer. During the training phase, the model had an accuracy 95%. The model was tested using real-time generated fundus images from patients which achieved 92% accuracy. The model is evaluated using multiple evaluation metrics such as Accuracy, Precision, TPR, Specificity FPR, F1-score for individual classes. It is missing the important evaluation metrics sensitivity that is of utmost importance in the medical domain.

2.4 Image Classification using CNN-RNN

This section will examine the studies that were performed in the domain of Image Classification using the CNN-RNN model.

In 2020, the whole world was fighting against the pandemic, researchers were finding accurate and time-saving methods of identifying the causes of the deadly disease. Implementing classification models to classify x-ray images was one of the solutions. In this problem domain, the CNN-RNN approach has proven effective. Islam et al. (2020) were able to identify x-ray images into three categories: Covid 19, Normal, and Pneumonia, using the developed model. The accuracy, specificity, and sensitivity of the custom CNN model and the LSTM as RNN were 99.40%, 99.20%, and 99.30%, respectively.

Breast Cancer is one of the domains where image classification can be performed to automate the process of cancer detection. Yan et al. (2018) introduced a hybrid CNN-RNN model that uses a pre-trained model InceptionNet V3 as a feature extractor and a bi-direction LSTM as an image classifier. The model is capable of capturing both high-level and low-level features in images. The model outperforms the current state-of-the-art model in this domain and achieves balanced accuracies across multiple Labels. Another study using the CNN-RNN hybrid model for lung cancer detection was proposed in the medical domain. A CNN is comprised of a 22-layer network for feature extraction and an LSTM for image categorization (Mhaske, Rajeswari and Tekdae 2019).

Automated Image captioning of the remotely captured image is the research area where CNN-RNN achieved significant results over the state-of-the-art encoder-decoder method. The research implements a similar approach of encoder-decoder but CNN performs the work of encoder and RNN acts as a decoder. The visual features of the image are extracted using CNN, the raw vector of the feature extraction is passed to RNN that translates them into multiple captions. The best caption is selected based on ground truth (Hoxha, Melgani and Slaghenauffi 2020). A similar approach was used by Sumbul and Demir (2019) where K branch CNN is used with a combination of Bi-Directional Long Short-Term Memory for Multi-Label Remote Sensing Image Classification. Researchers compared the CNN-RNN model with CNN-GA, CNN-KNN, CNN-SVM, CNN-MLP, the proposed study performed better than the compared hybrid models with an accuracy of 99.8%.

2.5 Research Gap

The studies reviewed in related work pose several demerits such as varied accuracies are achieved for multiple labels; it should be constant. Some of the models were exhibiting overfitting due to which training accuracy was high but not achieved similar results on the test. One of the most essential evaluation factors for a medical sector that poses a risk to the patient's health is Sensitivity. Models achieved significant Specificity but Sensitivity was lesser in comparison. Models were trained on imbalanced and small size datasets. Complex models were less interpretable.

3 Research Methodology

This research follows the modified Knowledge Discovery in Databases (KDD) process to complete the Research objectives. Now we will look at the stages involved in methodology to classify the fundus images of the diseases into the correct category as shown in Figure 2.

3.1 Knowledge Discovery in Databases (KDD).

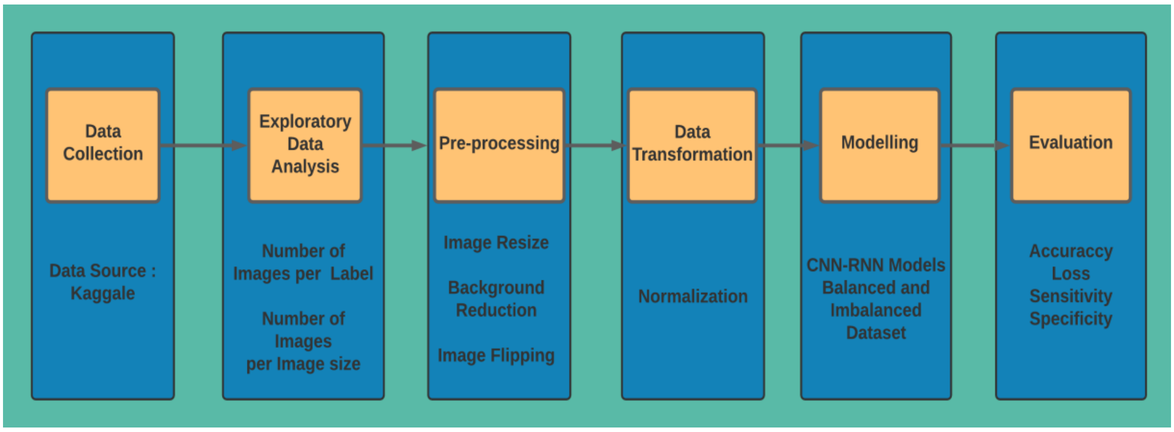

Figure 2: Knowledge Discovery in Databases Process.

Data Collection: Dataset consisting of images for multiple eye diseases is taken from Kaggle. The dataset contains 600 fundus images divided into 4 classes unequally. All the images are high resolution as shown in Figures 1 and 4.

Exploratory Data Analysis: Firstly, the dataset is imbalanced thus we need to find the number of images belonging to each class. Dataset contains 300,100,100,100 images representing Normal, Glaucoma, Cataract respectively. As seen in Figure 3.

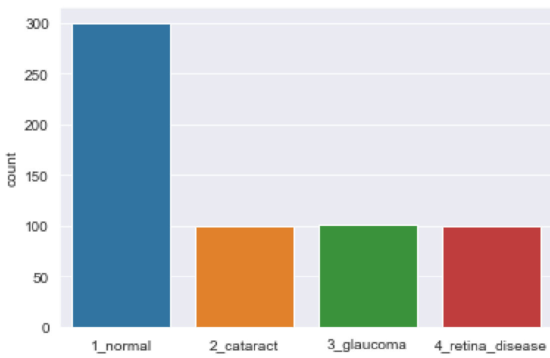

Figure 3: Number of Images per Class

After looking at the fundus images they appear to be having identical image sizes but after verifying the sizes of each image programmatically found that there are images with 3 different height and width dimensions (1224,1848), (1632, 2464) and (1728, 2592). After further analysis found that 40, 158, and 403 images belong to (1224,1848), (1632, 2464), and (1728, 2592) dimensions respectively.

Pre-Processing: This is an important step to get the data in model acceptable shape, format and achieving higher accuracy. The image filename and folder name in which it is stored in the file system denotes its label. All the images were imported and assigned the label according to their folder index. Now as found in exploratory data analysis, images are having different dimensions. To avoid additional pre-processing steps the images belonging to the highest dimensions were kept as it is also the larger number of images belongs to the highest resolution. The remaining images were resized to the highest dimension so that all will be having identical sizes.

Fundus images contain black background which should be kept minimal as possible as the model will consider the background as part of an image for training. As seen in Figure 4, 0:430 and 2190: 2592 pixels width wise contains black background portion. This black background portion is removed from both sides as seen in Figure 5, Image dimension after this pre-processing will be (1728, 1852).

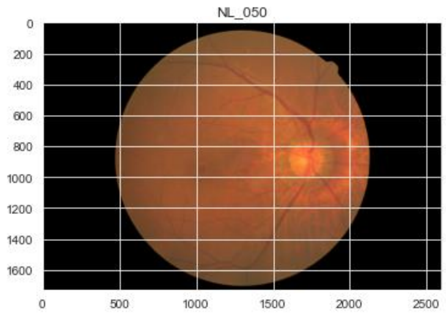

Figure 4: Before Background Reduction

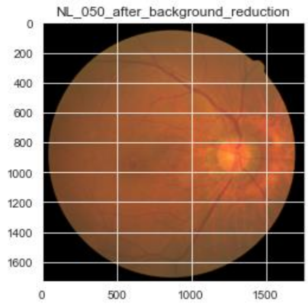

Figure 5: After Background Reduction

Images are resized to (224 x 224) as training the models with a high resolution of (1728,1852) would require high computation power, also transfer learning models accepts input image size of (224 x 224).

One of the objectives of the research is to train the model on balanced data and compare the results with an imbalanced dataset. To create a balanced dataset that will have equal images in all classes, images of glaucoma, cataract and retina disease are flipped vertically and horizontally and vertically. After this augmentation step, all classes contain 300 images each.

Figure 6: Original, Flipped Horizontally, and Flipped Horizontally and Vertically (L to R)

Both balanced and imbalanced datasets are divided into 70:20:10 ratio for Training, Validation and Testing respectively. One of the key aspects of this partition process is that after this process, all the three phases will have equal number of images in all four classes. Figure 7 and

8, represents the number of images for all three phases in imbalanced and balanced dataset respectively.

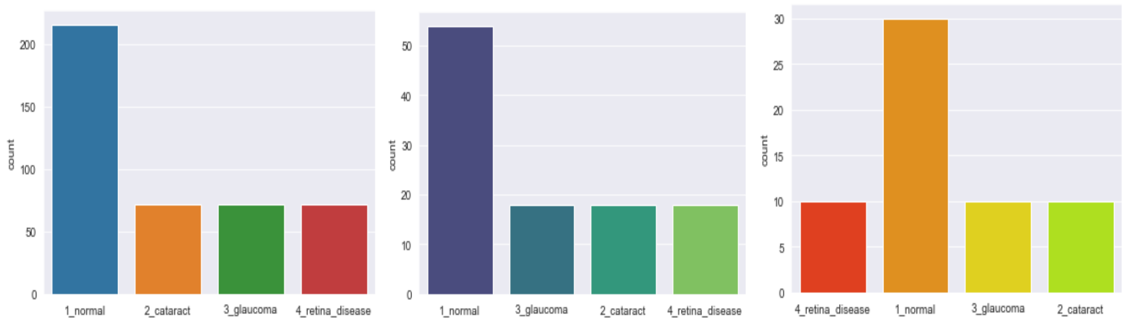

Figure 7: Left: Training set Middle: Validation set Right: Test set (Imbalanced Dataset)

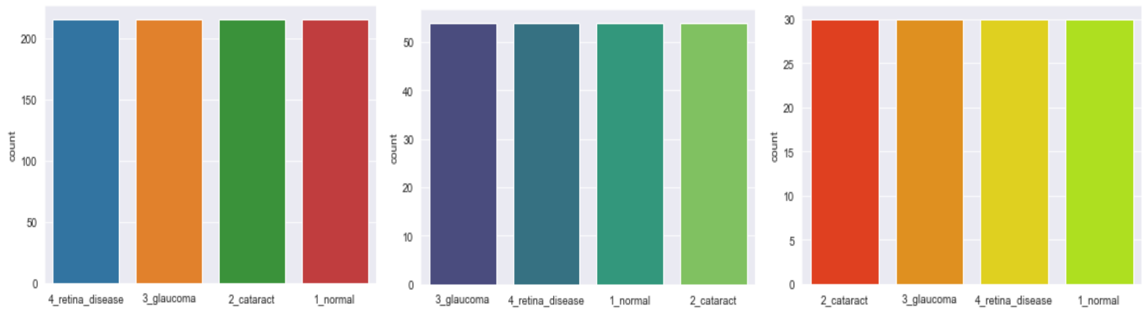

Figure 8: Left: Training set Middle: Validation set Right: Test set (Balanced Dataset)

Data Transformation: Data transformation step used in this research is Normalization. The images are represented by 8-bit channels of red, green, blue colours as there could be $2^8 = 256$ possible values for each colour. This becomes the combination of 256 x 256 x 256. The normalization is done by dividing the images by 255 so that image will be represented using a range of 0-1.

Modelling: In this step, the selected transfer learning models InceptionV3, DenseNet169, and InceptionResNetV2 are performing the feature extraction and LSTM module is used to classify the extracted features. All of the 6 models built uses pre-trained weights of their respective transfer learning models. All the 6 models are implemented by incorporating K-fold Cross Validation at k=5.

Evaluation: Accuracy, Specificity, Sensitivity, and Loss are the evaluation metrics that are used. The model's Sensitivity, also known as the true positive rate, is the capacity of the model to classify those who have the illness. In this application, the goal is to evaluate patients for medical intervention, which necessitates a model with a high Sensitivity. The true negative rate is the model's ability to accurately classify those who do not have the disease (Nugroho et al., 2017).

1) $\text{Accuracy} = \text{TP} / \text{Total Predictions}$
2) $\text{Specificity} = \text{TN} / (\text{FP} + \text{TN})$
3) $\text{Sensitivity} = \text{TP} / (\text{TP} + \text{FN})$

4 Design Specification

The CNN-RNN model will be implemented to classify the fundus images according to their correct labels. The feature extractors will be InceptionV3, InceptionResNetV2, and DenseNet169, which are CNN models, and the classifier will be Long Short-Term Memory, which is an RNN model.

Below Figure 9, represent the Design of the model. The model has been divided into 3 tiers, first tier consists of Data pre-processing, augmentation, transformation module. CNN-RNN module is a part of the second tier. Classified output from the RNN module, evaluation of the model, visualisation of the insights and model comparison is in the third tier.

The first tier begins with data selection and pre-processing. Images are divided into 4 different folders as per their respective classes. To import the data into python, 2 arrays would be created which will store image data and labels respectively. Augmentation block will be used so that number of images in each label would be equal for that image will be flipped Vertically and Horizontally and Vertically.

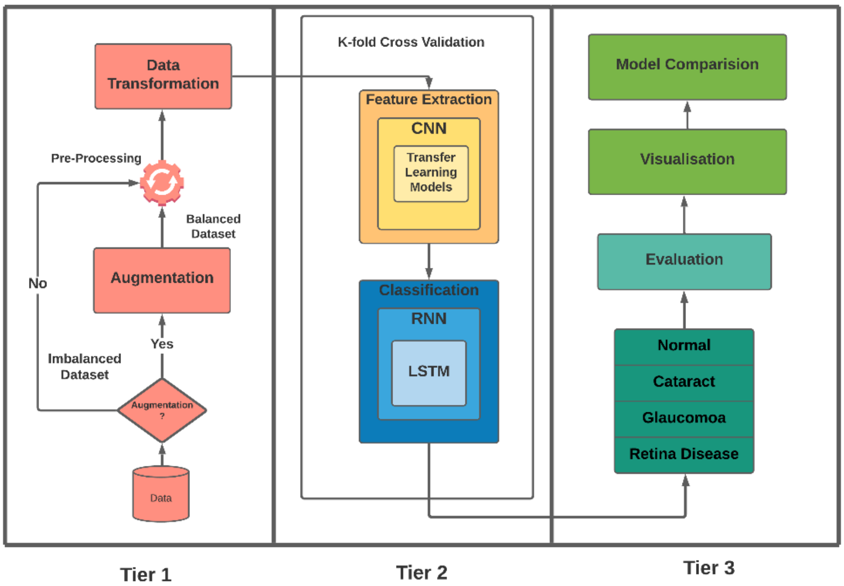

Figure 9: Process Flow Diagram

Firstly, entire dataset is imported. The dataset import function crops the images as shown in Figures 4 and 5, then resize the image to 224 x 224 size. As the function imports the data sequentially from each class the dataset will have Normal class images followed by Cataract, Glaucoma and Retina diseases. Now This imported data is then partitioned into Normal, Cataract, Glaucoma and Retina diseases. These four datasets are randomly shuffled. This process is incorporated in K-fold Cross Validation, usually, K-fold validation divides the data into Train and Test and then selects the next Train and Test set. This process is repeated as per the value of K specified. Here 5 is the value selected for K-fold Validation thus the process mentioned above is repeated 5 times resulting in selecting 5 different sets for Training Testing and Validation. Value of n is the number of images in each class for imbalanced dataset two n values are used one for the Normal class another for rest of the classes. For a balanced dataset, n is kept to 30 as all classes contain equal images After that 10% of images of the total number of images in each class is taken out as Test data in 4 separate objects using k*n to (k+1)*n.

These four dataset objects are then appended into one Test set and later shuffled randomly as append step added images for four classes one after another. Now, the remaining images present in the Normal, Cataract, Glaucoma and Retina diseases dataset are used for Training and Validation by dividing them into 80:20 ratio that would eventually provide 70:20 images of the total number of images. Here 80:20 split is performed on each class therefore, dataset objects belonging to Training are appended Train set and shuffled, a similar process is done for Validation set. To take out the test data indexes are used that allows to select images from k*n till (k+1)*n and for training and validation data first 0 to k*n images are added and then images from (k+1)*n till the last image of the set is selected.

After this step, data will be fed to feature extraction and classification block. The output of the second tier would be images classified into 4 labels. Third-tier will evaluate the model using different evaluation metrics such as Accuracy, Specificity, Sensitivity. The results obtained would be visualised using different visualisation methods.

5 Implementation and Evaluation

This section is focused on discussing the implementation and evaluation of the CNN-RNN models implemented for classifying the fundus images as per eye diseases. Different transfer learning models are used as feature extractors such as InceptionV3, InceptionResnetV2, and DenseNet169 in the CNN part of the model. The output from the CNN module is reshaped and given as input to the RNN module which consists of LSTM layer with Rectified Linear Activation Function (ReLU) followed by Batch Normalization layer. After that one Flatter layer was added and 2 blocks of Dense layer with ReLU followed by Batch Normalization Layer. The output layer is a Dense Layer with 4 units and a Softmax activation function.

5.1 Environmental Setup

The models are trained over a system that contains an i5 11th Generation CPU and 16 GB RAM. The programming language used is Python version 3.8.3 over Jupyter Notebook. Models are created using Keras API imported from TensorFlow. All the models are trained using RMSprop optimizer at the learning rate of 0.01 and clip value of 100. The loss function used for all the models is sparse categorical cross-entropy. All the models are implemented on 50 epochs and 5-fold cross validation. The batch size for the imbalanced dataset is 14 images and the balanced dataset is 27 images.

5.2 Implementation and Evaluation of InceptionV3-LSTM

The architecture of the InceptionV3 model consists of 3 Inception A blocks followed by grid size-reduction then 4 Inception B blocks followed by grid size reduction. In between 4 Inception B blocks and grid size-reduction, there is an auxiliary classifier. The end part of the architecture contains 2 Inception C blocks. The model has achieved 78.1% accuracy on ImageNet Dataset. The output from the InceptionV3 model is (5 x 5 x 2048 x Batch Size). This is reshaped into a (25 x 2048 x Batch Size) which is LSTM acceptable input shape.

5.2.1 InceptionV3-LSTM Evaluation Over Imbalanced Dataset

The model achieved an average accuracy of 89.78% and 49.06% in the training and validation phase respectively. Training loss is fairly constant and less than the validation loss. Average of training and validation loss is 1.23 and 13.79 respectively. The average test accuracy achieved by the model is 55.00%. Figure 10, indicates that the model is not able to classify images accurately. The Normal, Cataract and Glaucoma class images are fairly classified correctly but not for Retina diseases. The average test sensitivity and specificity obtained is 65.66% and 85.63% respectively.

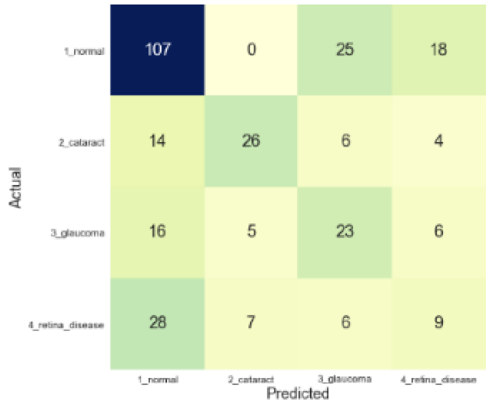

Figure 10: Overall Confusion Matrix for InceptionV3-LSTM (Imbalanced Dataset)

5.2.2 InceptionV3-LSTM Evaluation Over Balanced Dataset

The average training accuracy achieved is 91.90% which is slightly better than the imbalanced dataset. Model is not able to achieve similar accuracy in the validation, the average validation accuracy obtained is 60.60%. The accuracy achieved during the test phase is 64.33% which is a 9.00% improvement from the imbalanced dataset. The average training loss is minimal 0.65 but validation loss is 11.59. As per Figure 11, the model can classify all four class images but accuracy for retina disease is slightly less. The average specificity and sensitivity for the model is 74.33% and 88.41%. Overall model performance on the balanced dataset model is better than the model trained on an imbalanced dataset.

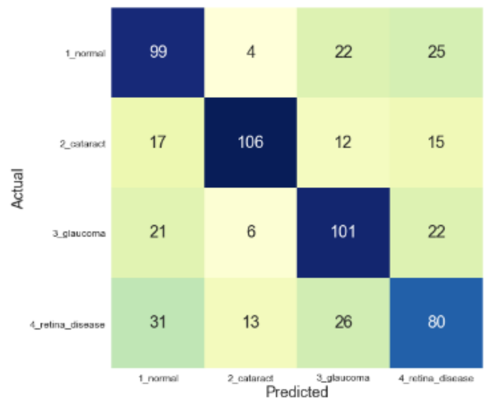

Figure 11: Overall Confusion Matrix for InceptionV3-LSTM (Balanced Dataset)

5.3 Implementation and Evaluation of InceptionResNetV2-LSTM

The overall structure of the model remains the same as InceptionResnetV1 the significant change is in the Inception A, B and C blocks used in version 2. The convolution layers in Inception A, B and C blocks of version 1 contains the same number of kernels 32, 128, and 192 respectively. The output shape of the model is (5 x 5 x 1536 x Batch Size) which is reshaped to (25 x 1536 x Batch Size).

5.3.1 InceptionResNetV2-LSTM Evaluation Over Imbalanced Dataset

The model achieved average training and validation accuracy of 88.44% and 50.00% respectively. The average test accuracy of the model is 55.66% The average loss for the training phase is 1.27 and the testing phase is 14.29. The overall confusion matrix indicates that model is able to identify the images for all classes but accuracy is extremely low. The average sensitivity and specificity of the model is 47.83% and 78.08% respectively.

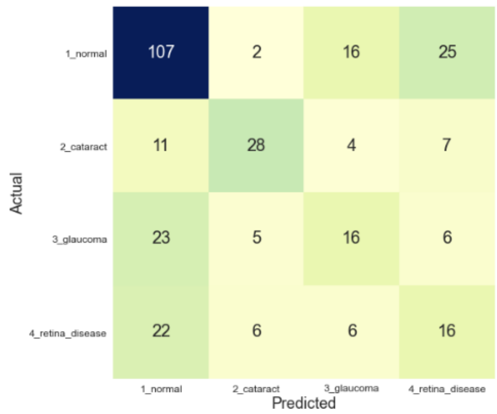

Figure 12: Overall Confusion Matrix for InceptionResnNetV2-LSTM (Imbalanced Dataset)

5.3.2 InceptionResNetV2-LSTM Evaluation Over Balanced Dataset

The average training and validation accuracy obtained is 91.12% and 63.38%. The test accuracy obtained by the model is 62.50%. The average loss for training is 0.73 and the validation loss is 5.37. Overall model sensitivity and specificity is 62.49% and 84.17% respectively. Overall Confusion Matrix represented in Figure 13, indicates that model is able to classify the Normal, Cataract and Retina diseases images with more accuracy but not the Glaucoma class images.

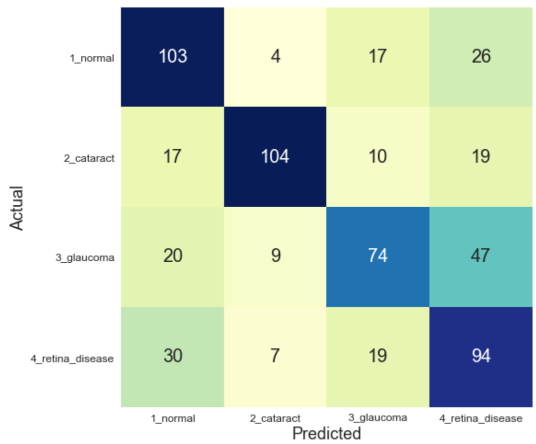

Figure 13: Overall Confusion Matrix for InceptionResnNetV2-LSTM (Balanced Dataset)

5.4 Implementation and Evaluation of DenseNet169-LSTM

The DenseNet169 model is built using 4 dense blocks and output from each block is forwarded ahead using transition layers. The number of layers in dense blocks varies depending on which DenseNet model is being implemented. Transition layers contain a batch normalization layer, convolution layer and 2 average pooling layers. The output of the model is (7 x 7 x 1664 x Batch Size) which is reshaped to (49 x 1664 x Batch Size).

5.4.1 DenseNet169-LSTM Evaluation Over Imbalanced Dataset

The model achieved 89.18%, 47.61%, 56.00% values for average training, validation and test accuracy respectively. The overall sensitivity and specificity achieved was 52.00 % and 78.59% respectively. Average training and validation loss is 1.22 and 12.46 respectively. Overall, the model is able to classify all classes accurately except for Retina disease.

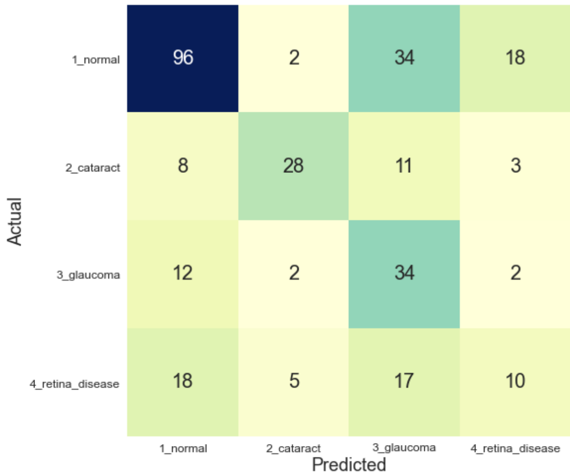

Figure 14: Overall Confusion Matrix for DenseNet16-LSTM (Imbalanced Dataset)

5.4.2 DenseNet169-LSTM Evaluation Over Balanced Dataset

The balanced dataset has achieved better accuracy than the imbalanced dataset. 92.57%, 60.00%, and 69.50% accuracy were obtained during the training, validation and test phase respectively. The average training and validation loss values are 0.64 and 14.06 respectively. The average Sensitivity of the model is 69.50% and the average specificity is 87.75%. Both the values are significantly better than the imbalanced dataset. The overall confusion matrix indicates that model is able to classify all the labels. The classification is more accurate for normal, cataract and glaucoma classes than retina diseases.

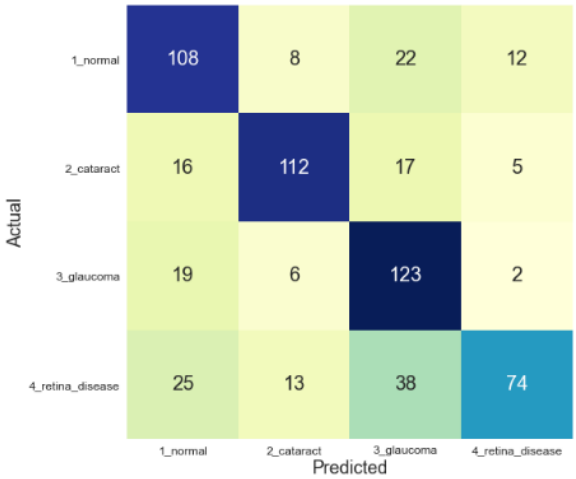

Figure 15: Overall Confusion Matrix for DenseNet16-LSTM (Balanced Dataset)

6 Comparison and Results

6.1 Comparison of Models over Imbalanced and Balanced Dataset (Objective 5)

Model Name	Imbalanced Dataset			Balanced Dataset		
	Training	Validation	Test	Training	Validation	Test
InceptionV3-LSTM	89.78%	49.06%	55.00%	91.90%	60.60%	64.33%
DenseNet169-LSTM	89.18%	47.61%	55.99%	92.57%	60.00%	69.50%
InceptionResnetV2-LSTM	88.44%	50.06%	55.66%	91.91%	63.30%	62.50%

Table 2: Comparison of Models over Imbalanced and Balanced Dataset

From Table 2, we can interpret that model has achieved better accuracy when modelled on a balanced dataset than an imbalanced dataset. All the models implemented on an imbalanced dataset achieved similar values for accuracy. As seen in the confusion matrices of the imbalanced datasets, models were not able to identify images of all the classes but on a balanced dataset, all models were able to classify image of all classes. When the dataset consists class imbalance the model tends to classify all the images with the label that has a greater number of images accurately. The test accuracy of all the models has increased by at least 6% and at the most 14% when implemented on the balanced dataset.

6.2 Comparison of CNN-RNN Models with Previous Researches (Objective 6)

Dataset	Number of Images	Same/Different Dataset as this research	Classifier	Accuracy	Author
DIARETDB0,HRF Image and STARE	2848	Different	Deep CNN	92.00%	Karthikeyan et al. (2019)
I Challenge-GON Comprehension	1200	Different	InceptionV3	81.00%	Smaida and Serhii (2019)
Messidor, Messidor-2, DRISHTI-GS	3250	Different	VGG16	84.88%	Sarki et al. (2020)
Akdeniz University Hospital Eye Diseases Department	9565	Different	Xception with multilayer perceptron	81.00%	Bulut et al. (2020)
STARE and DRIVE	440	Different	VGG19+Augmentation	93.58%	Das et al. (2019)
Kaggle	600	Same	ResNet50+Augmentation	60.00%	Santra T. (2019)
Kaggle	600	Same	DensNet169-LSTM	56.00%	This Research
Kaggle	1200	Same	DensNet169-LSTM	69.50%	This Research

Table 3: Model Accuracy comparison with Similar Researches

Table 3, contains the comparison of the accuracy obtained in this research against the researches that have performed multi-class classification of the fundus images. Among all the models implemented in this research DenseNet-169-LSTM model achieved the highest test accuracy of 69.50%. The model is not able to achieve the accuracy as compare to models that were implemented on different datasets. The main reason for that could be the number of images used for training the model. This dataset is trained on a relatively smaller number of images than other researches. DenseNet-169-LSTM Model was able to improve the accuracy of the model ResNet50 (Santra T, 2019) that is implemented on the same dataset.

7 Conclusion and Future Work

This research aimed to classify the fundus images according to 4 different categories. To answer the research question and objectives six different CNN-RNN models were implemented. The hybrid CNN-RNN models implemented on an imbalanced dataset were not able to classify the images with four classes accurately. For an imbalanced dataset, the DenseNet169 and InceptionResNetV2 achieved similar 56.00% accuracy. Due to class imbalance, the results of these models are affected. When trained on balanced datasets all of the models exhibited significant improvement from previous results. DenseNet169 was able to classify the images more precisely with an accuracy of 69.50%. In future work, the plan is to combine several similar datasets to increase the number of images in each class. As per the results obtained over the balanced dataset by performing data augmentation by flipping images the results were improved significantly in the range of 6-14%. If multiple datasets are combined then a more precise and robust model can be built. As Cataracts, Glaucoma and Retina diseases are associated with three different parts of the eye Optic disk, Optic nerve and retina thus segmentation of this part from the fundus images can be performed and later trained on the CNN-RNN model. This would be advantageous as fundus images of normal and with disease look similar but there could be a minute part of the image which could be the onset of the diseases. So, if the models are trained on segmented data model can be more accurate in classifying the fundus images.

8 Acknowledgement

The completion of this research project could not have been possible without the supervision of Dr. Christian Horn. I would like to express my gratitude towards Dr. Christian Horn for always pointing me in the right direction and advising me on how to formulate this project. I would also like to thank my family for consistently supporting me throughout the project lifecycle.

References

Acharya, U.R., Kannathal, N., Ng, E.Y.K., Min, L.C., Suri, J.S., 2006. Computer-Based Classification of Eye Diseases, in: 2006 International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE, New York, NY, pp. 6121–6124. https://doi.org/10.1109/IEMBS.2006.260211

Ahmad, H.M., Hameed, S.R., 2020. Eye Diseases Classification Using Hierarchical MultiLabel Artificial Neural Network, in: 2020 1st. Information Technology To Enhance e-Learning and Other Application (IT-ELA). IEEE, Baghdad, Iraq, pp. 93–98. https://doi.org/10.1109/IT-ELA50150.2020.9253120

Berrimi, M., Moussaoui, A., 2020. Deep learning for identifying and classifying retinal diseases, in: 2020 2nd International Conference on Computer and Information Sciences (ICCIS) IEEE, Sakaka, Saudi Arabia, pp. 1–6. https://doi.org/10.1109/ICCIS49240.2020.9257674

Bhadra, R., Kar, S., 2020. Retinal Disease Classification from Optical Coherence Tomographical Scans using Multilayered Convolution Neural Network, in: 2020 IEEE Applied

Signal Processing Conference (ASPCON). IEEE, Kolkata, India, pp. 212–216. https://doi.org/10.1109/ASPCON49795.2020.9276708

Bulut, B., Kalin, V., Gunes, B.B., Khazhin, R., 2020. Deep Learning Approach For Detection Of Retinal Abnormalities Based On Color Fundus Images, in: 2020 Innovations in Intelligent Systems and Applications Conference (ASYU). IEEE, Istanbul, Turkey, pp. 1–6. https://doi.org/10.1109/ASYU50717.2020.9259870

Das, A., Giri, R., Chourasia, G., Bala, A.A., 2019. Classification of Retinal Diseases Using Transfer Learning Approach, in: 2019 International Conference on Communication and Electronics Systems (ICCES). IEEE, Coimbatore, India, pp. 2080–2084. https://doi.org/10.1109/ICCES45898.2019.9002415

Dong, Y., Zhang, Q., Qiao, Z., Yang, J.-J., 2017. Classification of cataract fundus image based on deep learning, in: 2017 IEEE International Conference on Imaging Systems and Techniques (IST). IEEE, Beijing, pp. 1–5. https://doi.org/10.1109/IST.2017.8261463

He, J., Li, C., Ye, J., Qiao, Y., Gu, L., 2021. Multi-label ocular disease classification with a dense correlation deep neural network. Biomedical Signal Processing and Control 63, 102167. https://doi.org/10.1016/j.bspc.2020.102167

He, J., Li, C., Ye, J., Wang, S., Qiao, Y., Gu, L., 2020. Classification of Ocular Diseases Employing Attention-Based Unilateral and Bilateral Feature Weighting and Fusion, in: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI). IEEE, Iowa City, IA, USA, pp. 1258–1261. https://doi.org/10.1109/ISBI45749.2020.9098525

Hoxha, G., Melgani, F., Slaghenauffi, J., 2020. A New CNN-RNN Framework For Remote Sensing Image Captioning, in: 2020 Mediterranean and Middle-East Geoscience and Remote Sensing Symposium (M2GARSS). IEEE, Tunis, Tunisia, pp. 1–4. https://doi.org/10.1109/M2GARSS47143.2020.9105191

Islam, Md.Z., Islam, Md.M., Asraf, A., 2020. A combined deep CNN-LSTM network for the detection of novel coronavirus (COVID-19) using X-ray images. Informatics in Medicine Unlocked 20, 100412. https://doi.org/10.1016/j.imu.2020.100412

Jiang, H., Yang, K., Gao, M., Zhang, D., Ma, H., Qian, W., 2019. An Interpretable Ensemble Deep Learning Model for Diabetic Retinopathy Disease Classification, in: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, Berlin, Germany, pp. 2045–2048. https://doi.org/10.1109/EMBC.2019.8857160

Karthikeyan S., Kumar P., Madhusudan S, Madhusudan R., Sundaramoorthy S, Namboori, P.K.K., 2019. Detection of Multi-Class Retinal Diseases Using Artificial Intelligence: An Expeditious Learning Using Deep CNN with Minimal Data. Biomed. Pharmacol. J. 12, 1577–1586. https://doi.org/10.13005/bpj/1788

Li, C., Ye, J., He, J., Wang, S., Qiao, Y., Gu, L., 2020. Dense Correlation Network for Automated Multi-Label Ocular Disease Detection with Paired Color Fundus Photographs, in: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI). IEEE, Iowa City, IA, USA, pp. 1–4. https://doi.org/10.1109/ISBI45749.2020.9098340

Lin, J., Cai, Q., Lin, M., 2021. Multi-Label Classification of Fundus Images With Graph Convolutional Network and Self-Supervised Learning. IEEE Signal Process. Lett. 28, 454–458. https://doi.org/10.1109/LSP.2021.3057548

Lyona, C., Menezes, J., Shinde, T., Gavhane, M., Rohatgi, R.M., Chavan, S., 2020. Classification of retinal images in stages of diabetic retinopathy using deep learning, in: 2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN). IEEE, Greater Noida, India, pp. 228–231. https://doi.org/10.1109/ICACCCN51052.2020.9362913

Mhaske, D., Rajeswari, K., Tekade, R., 2019. Deep Learning Algorithm for Classification and Prediction of Lung Cancer using CT Scan Images, in: 2019 5th International Conference On Computing, Communication, Control And Automation (ICCUBEA). IEEE, Pune, India, pp. 1–5. https://doi.org/10.1109/ICCUBEA47591.2019.9128479

Nugroho, H.A., Oktoeberza, W.K.Z., Budiani, R.L., Adji, T.B., 2017. Analysis of Texture-based Features for Image Classification of Retinal Exudates, in: Proceedings of the International Conference on Imaging, Signal Processing and Communication - ICISPC 2017. ACM Press, Penang, Malaysia, pp. 54–58. https://doi.org/10.1145/3132300.3132333

Omar, Z.A., Hanafi, M., Mashohor, S., Mahfudz, N.F.M., Muna'im, M., 2017. Automatic diabetic retinopathy detection and classification system, in: 2017 7th IEEE International Conference on System Engineering and Technology (ICSET). IEEE, pp. 162–166. https://doi.org/10.1109/ICSEngT.2017.8123439

Pandey, A., Patre, P., Minj, J., 2020. Detection of Glaucoma Disease using Image Processing, Soft Computing and Deep Learning Approaches, in: 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC). IEEE, Palladam, India, pp. 1–7. https://doi.org/10.1109/I-SMAC49090.2020.9243596

Purandare, M., Noronha, K., 2016. Hybrid system for automatic classification of Diabetic Retinopathy using fundus images, in: 2016 Online International Conference on Green Engineering and Technologies (IC-GET). IEEE, Coimbatore, India, pp. 1–5. https://doi.org/10.1109/GET.2016.7916623

Qiao, Z., Zhang, Q., Dong, Y., Yang, J.-J., 2017. Application of SVM based on genetic algorithm in classification of cataract fundus images, in: 2017 IEEE International Conference on Imaging Systems and Techniques (IST). IEEE, Beijing, pp. 1–5. https://doi.org/10.1109/IST.2017.8261541

Santra, T., n.d. Automatic Retinal Disease Classification Using Machine Learning and AI 8.

Sarki, R., Ahmed, K., Wang, H., Zhang, Y., 2020. Automated detection of mild and multi-class diabetic eye diseases using deep learning. Health Inf Sci Syst 8, 32. https://doi.org/10.1007/s13755-020-00125-5

Smaida, M., Serhii, D.Y., 2019. Comparative Study of Image Classification Algorithms for Eyes Diseases Diagnostic 4, 9.

Sumbul, G., Demir, B., 2019. A Novel Multi-Attention Driven System for Multi-Label Remote Sensing Image Classification, in: IGARSS 2019 - 2019 IEEE International Geoscience and

Remote Sensing Symposium. IEEE, Yokohama, Japan, pp. 5726–5729. https://doi.org/10.1109/IGARSS.2019.8898188

Wang, J., Yang, L., Huo, Z., He, W., Luo, J., 2020. Multi-Label Classification of Fundus Images With EfficientNet. IEEE Access 8, 212499–212508. https://doi.org/10.1109/ACCESS.2020.3040275

Zaheer, N., Shehzaad, A., Gilani, S.O., Aslam, J., Zaidi, S.A., 2019. Automated Classification of Retinal Diseases in STARE Database Using Neural Network Approach, in: 2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE). IEEE, Edmonton, AB, Canada, pp. 1–5. https://doi.org/10.1109/CCECE.2019.8861588