Retinal Fundus Image Classification using LSTM - Convolution Neural Network

MSc Research Project
Data Analytics

Rahul Vaydande
Student ID: X20181663

School of Computing
National College of Ireland

Supervisor: Vladimir Milosavljevic

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Rahul Vaydande
Student ID:	X20181663
Programme:	Data Analytics
Year:	2021-2022
Module:	MSc Research Project
Supervisor:	Vladimir Milosavljevic
Submission Due Date:	15/08/2022
Project Title:	Retinal Fundus Image Classification using LSTM - Convolution Neural Network
Word Count:	6202
Page Count:	21

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:	Rahul Vaydande
Date:	14th August 2022

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):	

Contents

List of Figures

List of Tables

Retinal Fundus Image Classification using LSTM - Convolution Neural Network

Rahul Vaydande
X20181663

Abstract

In recent years, the prevalence of diabetes and its associated complication, diabetic retinopathy (DR), has increased alarmingly. Since DR might eventually result in permanent blindness, early detection is essential. As a result, countries undertake a variety of screening programs to prevent DR. Due to the scarcity of skilled professionals who can execute the diagnostic, DR identification remains a challenge. There is hence a clear need to create an automated DR detection system. Convolutional Neural Network (CNN) algorithms are the most often used DNN techniques in the field of medical image classification. The Convolution Neural Network in combination with a Recurrent Neural Networks (CNN-RNN) framework has demonstrated success in an Image classification task across different domains. The suggested hybrid CNN-RNN models combine the benefits of standard CNN with Long Short-Term Memory (LSTM). CNN models collect several distinct characteristics from fundus pictures, and indeed the retrieved features are categorised using LSTM. For this research, I have utilized the Kaggle Open-source dataset. The Kaggle dataset for DR suffers from a class imbalance problem. As a result, augmentation methods are utilized to normalize the dataset. The CNN-LSTM model which was developed produced a descent multi-class classification accuracy of 76% when compared to the pre-trained transfer learning models i.e., VGG19 and ResNet50 which produced accuracies of 75% and 47% respectively.

1 Introduction

Automatic categorization of ophthalmologic illnesses using retinal image processing has become common practice in telemedicine. Previously, manual segmentation was utilized, but it was challenging, time-consuming, labour-intensive, observer-oriented, and required a high degree of expertise. In contrast, computer-assisted identification of ocular problems is less costly, attainable, and purpose-oriented, and does not require a highly experienced physician to evaluate the scans. The development of screening systems is necessary for real-time categorization and on-time diagnosis of eye diseases that could be incredibly useful during the recovery period. There are several eye conditions, and each one may have a distinct origin. Diabetes is a disease that has become more common in recent years, and it can produce eye abnormalities that might lead to vision loss. Diabetes is a widespread condition these days, and it can affect the eyes and cause vision loss (Gargeya and Leng; 2017). In recent years, diabetic retinopathy (DR) has indeed drawn considerable attention on a global scale. The eye is without a doubt the most fragile and vulnerable organ in the human body since it is necessary for vision. Because it affects the blood vessels in the sensitive areas of the eyes, it is a serious issue for people. According to World Health Organization forecasts, over 170 million people worldwide have diabetic retinopathy, with the number anticipated to climb to 366 million by 2030. Various retinal pictures with varying degrees of diabetic retinopathy are shown in Figure 1 for demonstration. Mild DR is a form of the illness that begins with subtle alterations to the blood vessels of the eyes. The patient might be able to beat the illness in this situation and make a full recovery. DR will become moderate if the state of this disease is not addressed. Blood vessel leaking may begin with mild DR. In the following instance, if the illness worsens, it may turn into a proliferative, severe form of DR that might cause complete blindness.

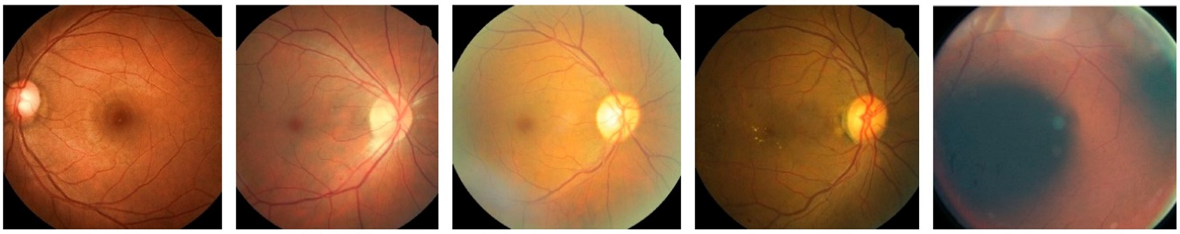

Figure 1: Stages of DR

To be effective to treat DR, one must first recognize it and assess its level of severity. Non-proliferative DR and Proliferative DR are the two predominant types of DR (Gargeya and Leng; 2017). Non-proliferative DR is classified into three stages, which are as follows: Mild NPDR is the mildest kind of DR, and it is proceeded by moderate severe NPDR. PDR is the most severe form of diabetic retinopathy. The five levels of DR severity are: no DR, mild DR, moderate DR, severe DR, and proliferating DR. An early sign of diabetic retinopathy is a lesion that appears as little red particles with a circular shape at the blood vessel terminals on fundus pictures. An indication of mild DR is the presence of microaneurysms, haemorrhages, and/or transudes. The creation of new blood vessels, as well as the previously noted abnormalities, occurs during the proliferative DR stage (Mateen et al.; 2020).Figure 1 depicts colour fundus pictures of a normal retina and increasing degrees of DR severity. The problem of recognizing symptoms early during DR due to visual resemblances discovered between normal fundus picture, relatively moderate DR, and potentially significant DR is a key challenge with DR identification. A loss of eyesight might result if diabetic retinopathy progressed to an advanced stage. Several computer-based strategies have been developed in the literature to help medical practitioners in recognizing DR in real-time

1.1 Motivation

In practice, identifying diabetic retinopathy (DR) involves a significant effort and time, this often culminates into diagnosis and treatment getting delayed, and sometimes in worst cases, leading to partial or total vision loss. More than 90% of people with diabetic retinopathy could likely be completely cured if this problem was discovered early (Zheng et al.; 2012) . Diagnosing diabetic retinopathy can be done manually or using computer-assisted technologies at the moment Manually identifying the illness would need doctors to be exceedingly skilled, and the procedure is typically rather time consuming. This

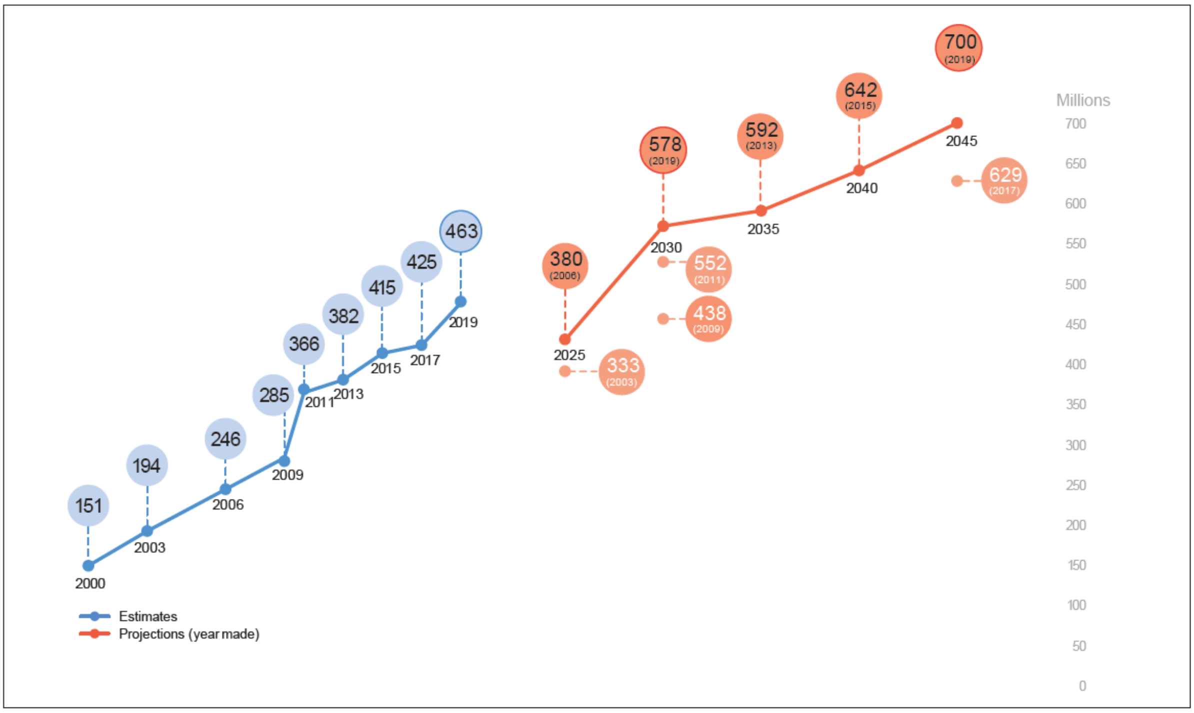

Figure 2: Diabetes prevalence estimation

method has been demonstrated to be inefficient and costly in various circumstances, particularly when many scans are provided. The limits of manual diagnosis in compared to computer-aided diagnosis are also made clear by the increased resources required. However, computer-assisted diagnostics is particularly effective in swiftly and affordably detecting a little amount of diabetic retinopathy evidence (Doshi et al.; 2020). Therefore, it lessens the need for manual work in the DR diagnostic. In the field of medical imaging and diagnostics, deep learning, and artificial intelligence (AI) have lately become more prevalent. In the early treatment of diabetic retinopathy, artificial intelligence possesses two significant advantages over traditional assessment: it carries a lower chance of human mistake, requires lesser effort, and is significantly effective at detecting lesions in the eye. Recognizing diabetic retinopathy is significantly impacted by artificial intelligence, including machine learning and deep learning. Although considerable research has been performed and substantial breakthroughs have also been made with in the field of recognition of diabetic retinopathy, the condition unfortunately affect a large segment of the population. Even today, many areas there are people who still struggle to get their diagnosis done owing to a lack of resources and technology. appropriate diagnosis, and treatment are also delayed and ignored in some cases. Additionally, it has been eventually discovered that classification of illnesses has occasionally been inaccurate, causing people to continue to suffer and eventually suffer complete vision loss, indicating that there is need for improvement (Li et al.; 2020). As seen in the graph in figure 2, if this problem is not resolved, it could have significant impact. To help in the classification of diseases, start employing fewer human resources, and provide experts with a powerful and effective base, medical image recognition technology must be used.

1.2 Research Question

- How does the performance of the traditional CNN-LSTM compare with pretrained CNN models trained on same dataset?

Objectives	Description	Metrics
1	A detailed evaluation of present DR categorization techniques i.e., basic CNN, Transfer Learning, and CNN-LSTM.	
2	Gather, compile, and augment the retinal scans before processing the data to extract attributes that distinguish DR images into different stages of DR	
3	CNN-LSTM model implementation and analysis	Accuracy, Precision, Recall, and F1 score
4	VGG19 model implementation and analysis	Accuracy, Precision, Recall, and F1 score
5	ResNet50 model implementation and analysis	Accuracy, Precision, Recall, and F1 score
6	Comparing the performance of constructed model with the transfer learning models	

Table 1: Research Goals

1.3 Document Structure

The remaining of the report is organized as follows. A overview of the literature is given in Section 2 for a number of concepts, including the use of CNN, Transfer Learning and CNN-LSTM in research and the current methods for DR detection and the challenges associated with it. The project's methodologies and design specification are summarized in Section 3 and 4. It explains and contrasts the method adopted for the implementation of this research. Section 5 comprises of the actual implementation of the different model utilized. The Section 6 and 7 contains the Evaluation and comparison of results. Finally the last section 8 and 9 has the Future works and conclusion.

2 Related Work

Retinal scans are analysed by a medical practitioner by merely analyzing each of them and, which oftentimes is challenging for them to identify lesions and, thus, making the overall diagnosis a challenging task. Diabetic retinopathy's early symptoms are so subtle that even some ophthalmologists sometimes overlook them. In the discipline of ophthalmology, artificial intelligence (AI) is now playing an important role in the detection of serious diseases like diabetic retinopathy. The automated system offers a far wider range of advantages and uses than the traditional method. Compared to a manual approach, an automated system can diagnose diabetic retinopathy more effectively, rapidly. As a result, automatic identification of DR is essential. The automated systems are basically made up of AI/ML model and in this section, we will be discussing some of the research conducted in the field of DR Detection.

2.1 DR detection using CNN

In the scientific world, DR classification has drawn a lot of interest. For DR detection the first model which I have reviewed, (Gondal et al.; 2017) proposed a CNN model, for training and testing, they used two publicly accessible datasets, Kaggle and DiaretDB1.

The three phases which are categorized as referable DR are moderate, severe, and proliferate whereas the No-DR and Mild-DR are considered non-referable DR. Their CNN model's performance on DiaretDB1 was evaluated utilizing a basic binary classification which obtained an output with sensitivity of 93.6 percent and specificity of 97.6 percent. The following study follows a similar methodology but the study only classifies the images in DR and NO-DR which is not I am going to implement. The next framework developed by (Wang et al.; 2017) categorizes images as abnormal, healthy, referable, or non-referable DR. It achieves good AUC scores of 0.978 and 0.960 on normal and referable DR, respectively. They use three networks in the method they propose: main, attention, and crop. The core network implements the Inception model trained on ImageNet; the attention network employed emphasized on types of lesions in the images as well as the crop network cropping the image with the highest attention. Though the author constructed a model and produced good results in both studies, utilizing different datasets and conducting simply binary classification is an under-utilization of resources. To prevent this, I have deployed a model for a multi-class categorization that allows for in-depth diagnosis.

(S. et al.; 2020) in their research stated that feature extraction from the fundus images can be crucial and the difficult process of automated analysis of DR. When it comes to image categorization performance, convolutional neural network algorithms surpass traditional algorithms. To extract properties from retinal images for use in machine learning classifiers, their research used a special CNN architecture. The model was evaluated using a variety of classifiers, including SVM, J48, Naive Bias and Random Forest on datasets from MESSIDOR and Kaggle. Each classifier's efficacy was assessed by comparing its accuracy, precision, and specificity. With a 2-class accuracy of 99.89%, the J48 classifier achieved the best results on the MESSIDOR and other datasets, according to the study, which used a variety of classifiers. This paper demonstrated that the basic CNN may be combined with different models to increase classification performance for DR diagnosis. This is a good thing because the author used machine learning techniques that are simple to understand. The sole drawback to this study is that Boolean categorization does not appear to be a useful method of diagnosing DR since it provides no additional information to the qualified doctor. The next study, which I have studied, has a more complicated architectural design and is therefore more challenging to comprehend.

An approach for classifying DR in a hierarchical manner was developed by (Ahmad and Hameed; 2020) using ANN. The pre-processing stage makes use of texture and color histograms, as well as feature extraction algorithms. Color-based characteristics were extracted at the pixel level using the HMMD also known as Hue Minimum Maximum Difference color system. Weights to every ANN layer were selected out of a spectrum of 1 to -1, as well as (0.5 to -0.5). It will be interesting to see how this approach performs on fundus images given that this collection includes non-fundus images. Haemorrhages upon fundus images are an important factor in the development of DR and a big component of the disease. In this case, accurate and fast haemorrhage detection are essential for timely patient with diabetes. Haemorrhages can be noticed in the deep and the superficial of retina i.e., the superficial layer of the retina exhibits bright red and linear haemorrhages, whereas the deep layer of retina has stronger red and rounder haemorrhages.

(Wu et al.; 2019) suggested to use a technique based on human visual features and two-dimensional (2D) Gaussian fitting. To begin, the natural colour retina image is pre-processed using brightness normalization and contrast limitation adaptive histogram equalization. Following background estimate and watershed segmentation, potential haemorrhages are retrieved. Finally, visual properties of potential haemorrhages are extracted using 2-dimensional Gauss fit and human vision characteristics. Furthermore, depending on visual characteristics, haemorrhages are extracted from probable haemorrhages. The suggested approach is examined using 219 retina scans from the DIARETDB dataset. According to the experimental results, the overall average sensitivity was 100%, and the specificity, and accuracy was 82%, 95.42% respectively. Leveraging images of the fundus images, (Deperlioglu and Kose; 2018) incorporated deep learning models with image processing techniques to categorize DR. They used image improvement techniques such histogram equalization, the HSV and V transform method, and ultimately, they implemented a Gaussian low pass filtering technique for the fundus images to improve the detections of the retina. Utilizing CNN, categorization was carried out after pre-processing the data. Utilizing 400 photos from the DR Detection database on Kaggle, the effectiveness of the suggested model was assessed. Categorization was used at every phase of the image pre-processing process. Twenty trials were conducted in each stage to obtain the aggregate of all findings. This research's output had an sensitivity of 96.7%, specificity of 93% and an accuracy of 97%.The findings showed that retinal fundus pictures were a highly efficient approach for DR identification. To further enhance the performance, the picture pre-processing methods used in this study can be applied to other hybrid models, therefore, In my research I have utilized a dataset which has pre-processed Gaussian filtered images.This demonstrates how important picture pre-processing is since it increases the model's accuracy.

2.2 DR detection using Transfer Learning

An evaluation of the different classification models that use conventional deep learning techniques makes up the first portion of the related work. This section comprises research papers that used the Transfer Learning technique to classify eye diseases. A deep learning model called transfer learning leverages a model that has been previously trained on a different dataset as the basis for a new model to tackle a new challenge. In this the author (Gangwar and Ravi; 2021) proposed a hybrid deep learning transfer-learning approach for DR detection based on pre-trained Inception-ResNet-v2. To creating the hybrid model, a customized block of CNN layers was placed on top of Inception-ResNet-v2. Based on the Messidor-1 dataset for diabetic retinopathy and the APTOS 2019 blindness detection, researchers assessed the performance of the suggested model. In comparison to previous published results, the model performed much better. On the Messidor-1 and APTOS datasets, they reported test prediction accuracy around 72.33% and 82.18%, respectively. This research is a splendid example of modifying a CNN model and building new models on top of it to create a novel model which can be used for DR detection and as well perform better. The author (He et al.; 2020) in their research suggested an approach that incorporates feature extraction, unilaterally and bilaterally weighting, as well as a combination of both. ResNet is employed main Features Extraction Component in this approach, which itself is divided up under 3 sections, each dedicated to a specific purpose. The output of this componet would be sent to the Feature Splicing Unit, which, as previously stated, utilizes 3 types of attention weights. Additionally, the weights are increased, mixed, and spliced to produce a fusion of all. The last module is in charge of categorizing the feature vectors. Eight categories are represented in the collection of images, which show individuals with various eye conditions. Using 1,167 scans and threefold cross-validation, the model performs multi-label classification. The model was evaluated using metrics such as Kappa, F1 Score, and AUC. Res-Net101 was the major feature extractor used by the model, which resulted in output with the 93.40 % accuracy. Additionally, despite the model's great degree of accuracy, a model of this complexity is less interpretive.

2.3 DR detection using Hybrid CNN-LSTM Model

While CNNs can extract the key features from images, LSTMs can selectively retain patterns for a long time. This CNN-LSTM structure might perform better than traditional CNN classifiers when utilized for image classification. [1]. (Amalia et al.; 2021) combined two deep learning model, Long Short-Term Memory (LSTM) with Convolution Neural Network (CNN)—to detect DR using retinal fundus data. GoogleNet operated here as CNN model in their research. The result was a conclusion of the features in retinal fundus scans. The image features and a written description were provided to LSTM as a vector. The model achieved 90% accuracy in identifying and characterizing DR using a hybrid of CNN with LSTM. A descriptive statement generated by the model would aid experts with their diagnostic. The report makes note of the disease's severity. By combining deep learning algorithms, the study illustrates the prospect of diagnosing diabetic retinopathy via fundus imaging more efficiently. Similarly, Author (Wu et al.; 2017) suggests an approach for handling DR images that makes use of a generative captioning process to create a small sequence that summarizes the abnormal portions in retinal images. The deep recurrent architecture-based generative model employed in the generative approach for images incorporates convolution neural networks (CNN) for objects recognition and identification with long-short-term memory (LSTM) allowing machine interpretation and sequence generation to generate meaningful sentences summarizing an image. The model constructed on the DIARETDB1, Messidor, and DIARETDB0 datasets works effectively and produces fluent sequences. Furthermore, the data show that the accuracy is up to 88.53 percent, and the detection accuracy is higher than 90 percent. For me to employ a similar strategy from the above two research in my work, I would have to manually caption each one of the images in use for training my model, which would have been a tiresome process.

Diabetes is diagnosed in this research (G et al.; 2018) by analysing Heart Rate Variability (HRV) data acquired from ECG readings. Though, it uses different data form, I have decided to review this paper because it uses the same proposed model as mine. To automatically identify the abnormalities, researchers used CNN and CNN-LSTM, 2 deep learning networks that use convolutional neural networks (CNN) and long short-term memory (LSTM). Deep learning approaches do not need any feature extraction, in contrast to the standard analytic methods that have been used up to this point. Researchers started with classification, dividing the dataset into testing and training data. The maximum accuracy for test data using CNN-LSTM was 90.9%. With fivefold cross-validation, CNN produced an output with accuracy of 93.6%, while CNN-LSTM implementation produced the highest accuracy of 95.1%. In this research, deep learning algorithms are used for the first time to classify between diabetes and normal HRV. The accuracy gained through cross-validation was the highest value obtained for the automated identification of diabetes utilizing HRV.And finally the last paper which I have reviewed is also performed on a different subject but the technique implemented is very much in line with mine. The purpose of this study (Islam et al.; 2020) was to provide a

[1] `https://machinelearningmastery.com/cnn-long-short-term-memory-networks/`

deep learning technique for autonomously diagnosing COVID-19 from X-ray scans that combines a convolutional neural network (CNN) and a long short-term memory (LSTM). In this system, deep feature extraction is performed by CNN, while feature-based detection is performed by LSTM. The data for this technique consisted of 4575 X-ray scans, containing 1525 images of COVID-19. The experimental findings demonstrate that our suggested system obtained 99.4% accuracy, 99.9% AUC, 99.2% specificity, 99.3% sensitivity, and 98.9% F1-score.

2.4 Conclusion

The research evaluated in the above section, has motivated me to find a solution for the problem of varying output from the set of reviewed work. When it comes to medical domain, sensitivity plays a crucial role as accurately classifying a positive case of DR as positive is more important than mis-classifying a negative case as positive. The other research gap identified is that class imbalance in the kaggle dataset which is the widely used dataset. To tackle this problem, I have used the data augmentation technique which aid to better train the model. The data source which I will be utilizing in my research will be from public source such as Kaggle which has a pre-processed images which can be used directly. Finally, after sourcing the dataset, I have conducted a study of DR Detection using CNN and RNN-LSTM and then evaluate the model performance depending upon various evaluating parameter such accuracy,precision, Recall and F1 score with different transfer learning models.

3 Methodology

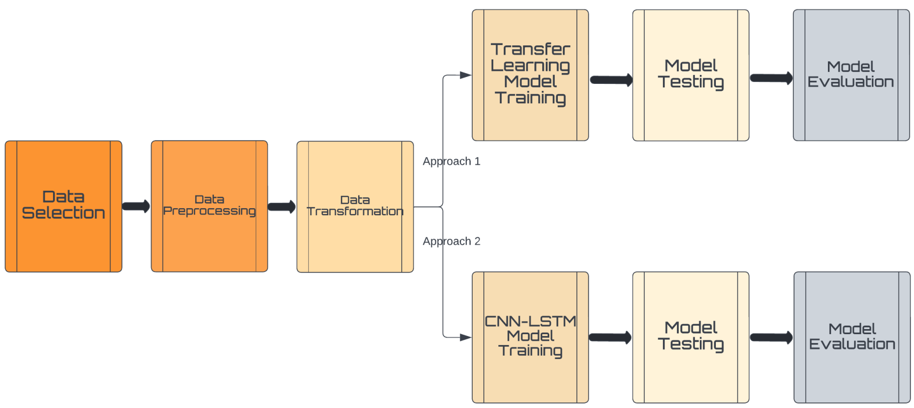

Figure 3: Modified KDD Process

A huge volume of data is now being generated from several sources. To make decisions, it is vital to extract relevant intel from the data as there are now several resources available for detection of DR. With a few modifications, Knowledge Discovery in Databases (KDD) is utilized in this research. By developing a novel convolutional neural network-LSTM model, KDD aids in the transformation of raw fundus images data as well as the extraction of information for the purpose of detecting diabetic retinopathy. The figure Figure 3 shown below depicts the KDD process adopted for this research

3.1 Data Selection

A substantial amount of study has previously been undertaken in the subject of DR detection utilizing various data sources. Instead of utilizing the original dataset from source like APTOS 2019 Blindness Detection, I will be using the same dataset but a subset of the original dataset to obtain required retinal fundus photos for my research.This dataset has mentioned in 2.1 has images which are pre-proccessed with Gaussian Filtering.The Dataset [2] which is utilized in this research had Clinicians assign a severity rating of 0 to 4 to each picture in the collection. The Figure 4 shows the class distribution for the dataset.

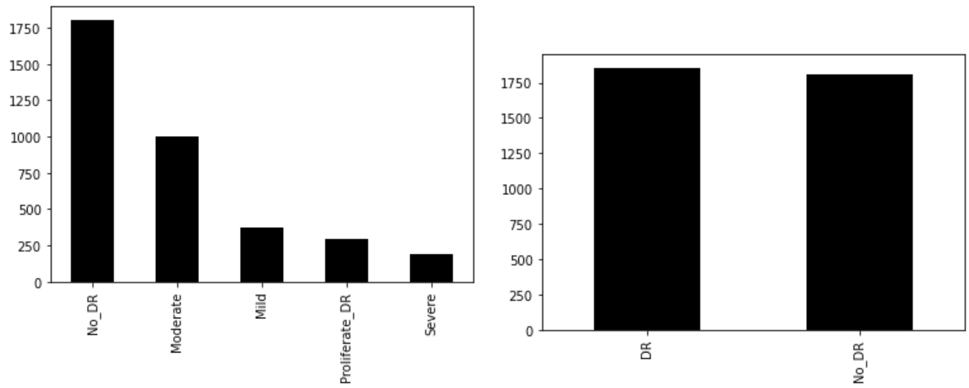

Figure 4: Class Distribution

3.2 Data Preprocessing and Transformation

After conducting the review of the work done in this domain, I came to conclusion that data needs transformation for this research and plays a rather important part. In KDD, data transformation is the process of transforming raw data from one structure to another. This step is certainly important in data management and integration as the data is brought to degree in which it can be directly injected to the DL or ML Model. In my research, In the pre-processing pipeline, data resizing, shuffle, and normalization were conducted. A training set, validation set, and testing set were created from the pre-processed data set. The dataset will then be divided into their respective classes/labels as given below:

- 0 – No DR
- 1 - Mild
- 2 - Moderate
- 3 - Severe

[2] https://www.kaggle.com/datasets/sovitrath/diabetic-retinopathy-224x224-gaussian-filtered

- 4 - Proliferate DR

Data augmentation will be used to boost the performance of deep learning as well as machine learning model, allowing the model to generalize better by avoiding overfitting. If there is enough data, a model's accuracy and performance may be enhanced. For data augmentation, an image data generator from the Keras package will be employed. The pictures will be normalized using the Image data generator's rescale function and mixed randomly to improve the convergence and stability. After successful formation of these datasets, I performed transformations on them and create data loaders for each of the dataset. Transformations that I performed include resizing the images to 224x224 size, random horizontal flip, random vertical flip, and normalization. Kaggle's dataset has a challenge of class imbalance and by using such an imbalanced dataset the model will have a high false negative rate as the no. of No DR class images are drastically more as compared to other classes.Therefore, Data Transformation in this research is a crucial step. The Figure 5 below shows data transformation was performed on the fundus images.

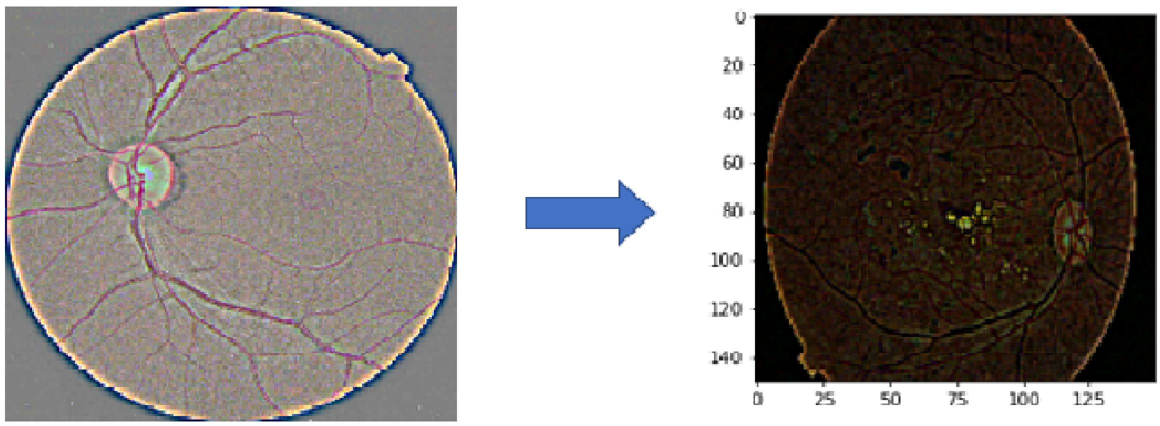

Figure 5: Designing model and data mining

3.3 Designing model and data mining

Data mining may be broadly defined as the process of identifying relationships, trends, and irregularities utilizing a huge dataset and making predictions about outcomes. The goal of this study is to develop a convolutional neural network (CNN) and a long short-term memory (LSTM) network that operate together to predict diabetic retinopathy (DR) as shown in Figure 6, In addition to CNN-LSTM, ResNet 50, and VGG19 models are used to detect DR and compare the performance of the proposed CNN-LSTM model on retinal fundus pictures. CNN models trained on same dataset.

3.4 Evaluation and Interpretation

A crucial stage in almost any Deep learning process would be the analysis and assessment of outcomes. This really enables to choose the most effective approach in accordance with its performance metrics as well as to compare the model's performance. Since the targeted

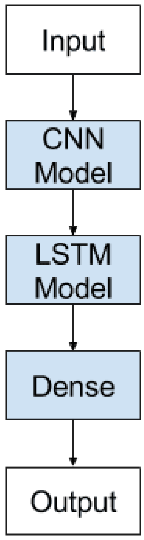

Figure 6: Basic Proposed Model

output in my research is not binary, a conventional confusion matrix is not appropriate. For evaluation purposes, I had to use a multi-class confusion matrix. The performance metrics used to evaluate and compare the models include confusion matrix, precision, f1 score, accuracy, and recall. The final step is to create graphs by visualizing all the results.

4 Design Specification

This section is an basic introduction to implemented system for in this research as depicted in Figure 7, and we will be discussing each phase of design in detail

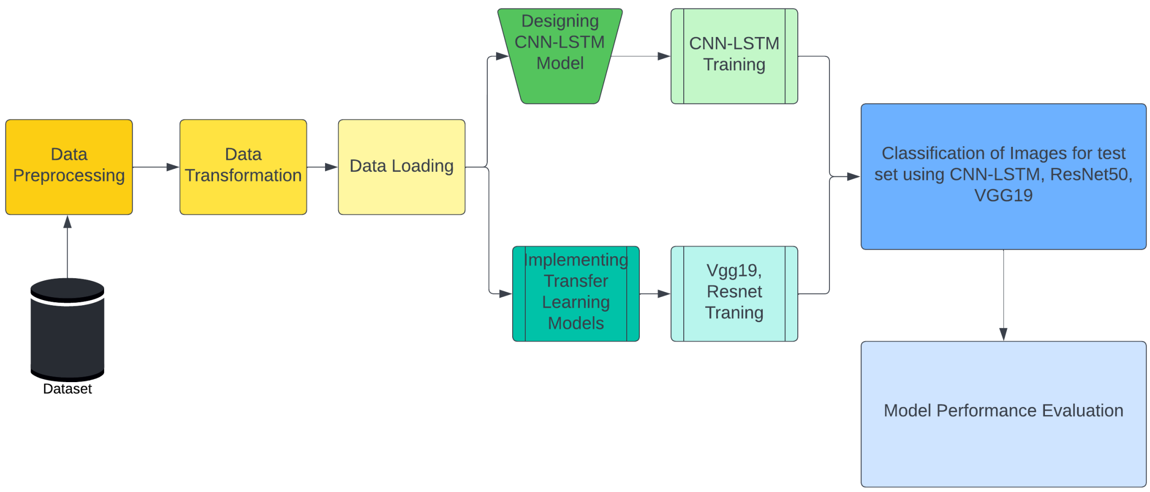

Figure 7: Project Design

4.1 Convolutional neural network (CNN)

Deep learning has now become arguably the most effective feature extraction technique. Deep Neural networks have an advantage over conventional machine learning techniques as they are capable of autonomously develop hierarchical representations from sizable datasets. In numerous applications, including image classification, object detection, and medical image analysis, CNNs have demonstrated exceptional performance. The probability of obtaining a more accurate feature vector increase as the number of layers within the neural network increases, yet this could result in difficulties such as vanishing gradient or gradient explosion. Because of these issues, a network's depth is restricted. A CNN is a specific kind of multilayer perceptron, however unlike deep learning architecture, a simple neural network cannot learn complicated features. The fundamental principle behind a CNN is that it can extract local features from inputs at high layers and send them to lower layers for features that are more complicated. Convolutional, pooling, and fully connected (FC) layers make up a basic CNN model as shown in the Figure 8

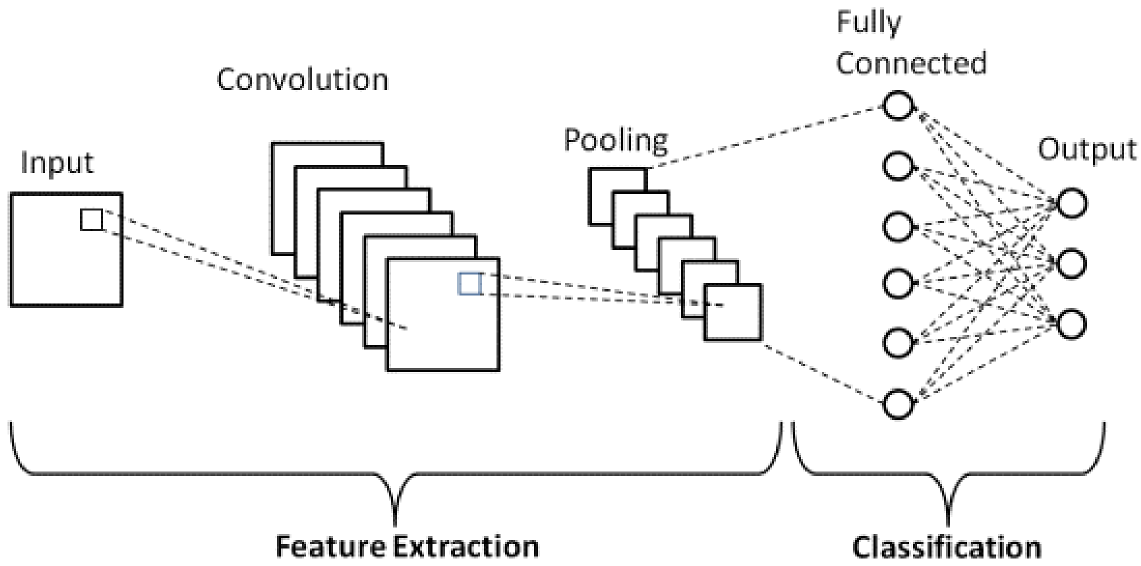

Figure 8: Basic CNN Architecture

4.2 Long Short-Term Memory (LSTM)

Recurrent neural networks are improved by long short-term memory (RNNs). The vanishing and exploding gradient problem are solved with LSTM instead of traditional RNN units. The key distinction between it and RNNs is that it adds a cell state to store long-term states. An LSTM network can remember and link information from the past to information acquired in the present. The LSTM is used in conjunction with three gates: first an input gate, then a "forget" gate, and finally an output gate. The current input is denoted by Xt, the new and previous cell states are denoted by Ct and Ct-1, respectively, and the current and prior outputs are denoted by ht and ht-1 as depicted in the Figure 9.

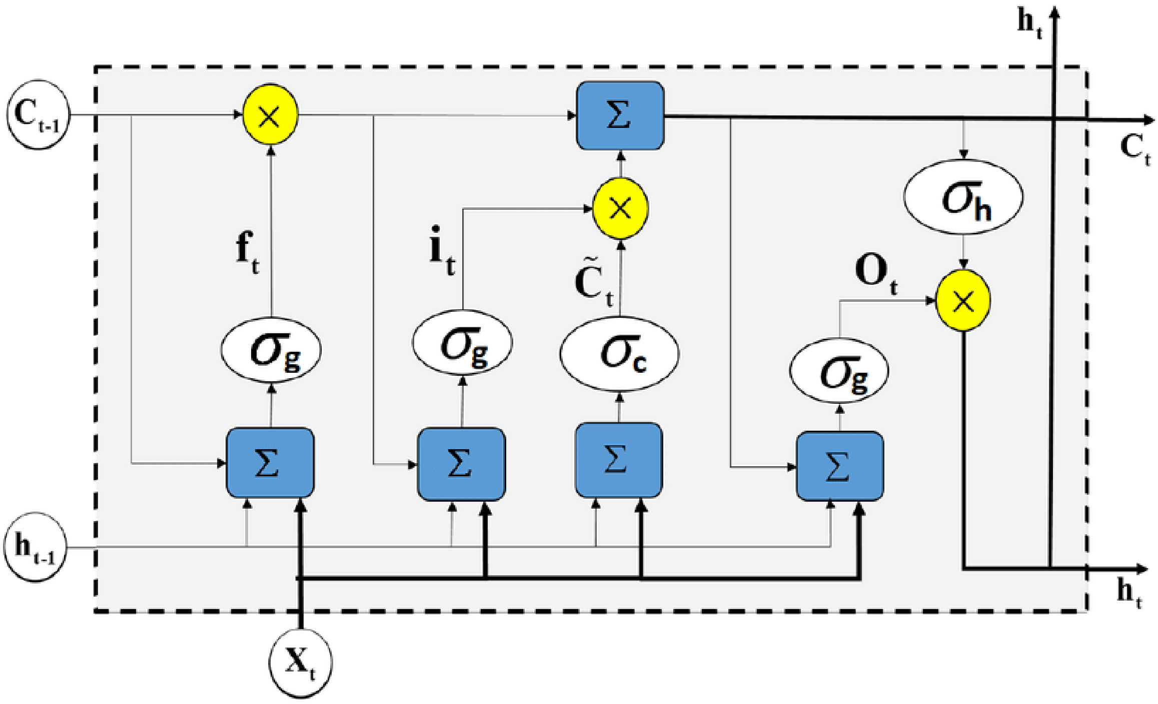

Figure 9: Basic LSTM Cell

4.3 The Process

The images which are stored in the google drive will be automatically loaded on the google collab instance when the drive is mounted. The next step is splitting the dataset into train, test and validation into 75:15:10 ratio and while doing this data augmentation is also performed. This includes image re-scaling to 150x150 for VGG19 and 224x224 for CNN-LSTM and also normalizing the images. The images are flipped horizontally and vertically and shuffled randomly to better train the different models. The data loading is performed in batch size of 32 with 4 workers so as to avoid load on the RAM. These images are then fed directly to the different models. The classifier layer of each model is developed/altered in such a way that they produce a 5-class output depending upon the type of input image. The models are then trained upto a certain epoch and the model with the best accuracy was selected for testing. These outputs are then evaluated on a common metrics of accuracy, precision, Recall and F1 score.

5 Implementation

This section focuses on the development of the CNN-LSTM, and the transfer learning models used to categorize retina images based on severity of DR.

5.1 CNN-LSTM

Using the dataset from Kaggle across five classes, a combinational approach was built in this study to automatically detect retinal retinopathy. Combining CNN with LSTM networks, where CNN is used to extract intricate details from images and LSTM is utilised as a classifier, created the framework for this architecture as shown in Figure 10. Twelve convolutional layers, five pooling layers, one FC layer, 2 LSTM layer, and one output layer using the SoftMax function make up the network's 20 layers. One pooling layer, two or three 2D CNNs, and each convolution block are concatenated. The ReLU function activates the convolutional layer, which has a 3 x 3 kernel size, for feature extraction. Utilizing a max-pooling layer with a size of 2 x 2 kernels

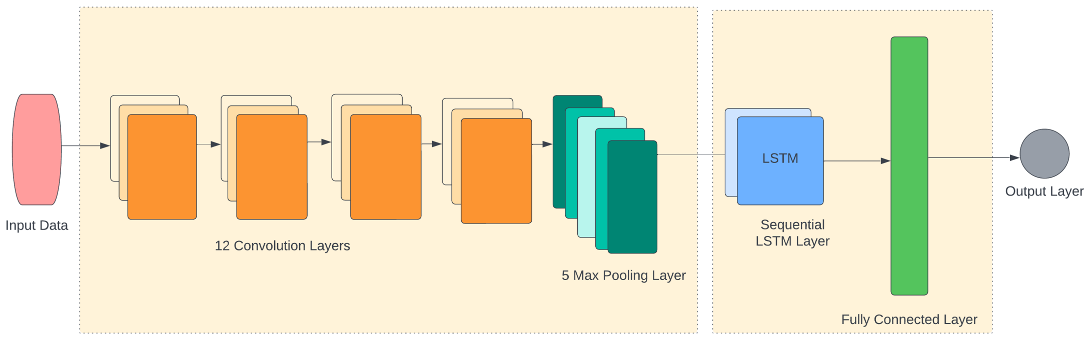

Figure 10: CNN-LSTM Architecture

5.2 Transfer Learning Model

There are 19 layers in the convolutional neural network known as VGG-19. The model has been pretrained using the ImageNet database's more than a million pictures. The pre - trained models can categorize data into Thousand different objects. But, our dataset has only 5 classes to classify instead of the predefined 1000 classes of the VGG19 model hence, I have changed Input to this altered version of the VGG19 model i.e., I have selected the Input layer with 25088 in features to take in as much as information from the previous layers and then I have reduced them to 2048 and passed this to the next layer and gradually reduced to 5 output features. ReLU has been selected as an activation function for the newly designed layers. And to avoid the problem of overfitting, I have also implemented a dropout function with P=0.6. The basic architecture of the VGG19 model is depicted in the Figure 11. The ResNet50 model is also another example of pretrained model trained on the ImageNet dataset and with 48 Convolution layers, 1 MaxPool layer, and 1 Average Pool layer. For this research, I have modified the resnet50 model by taking the same steps as done for the VGG19 model

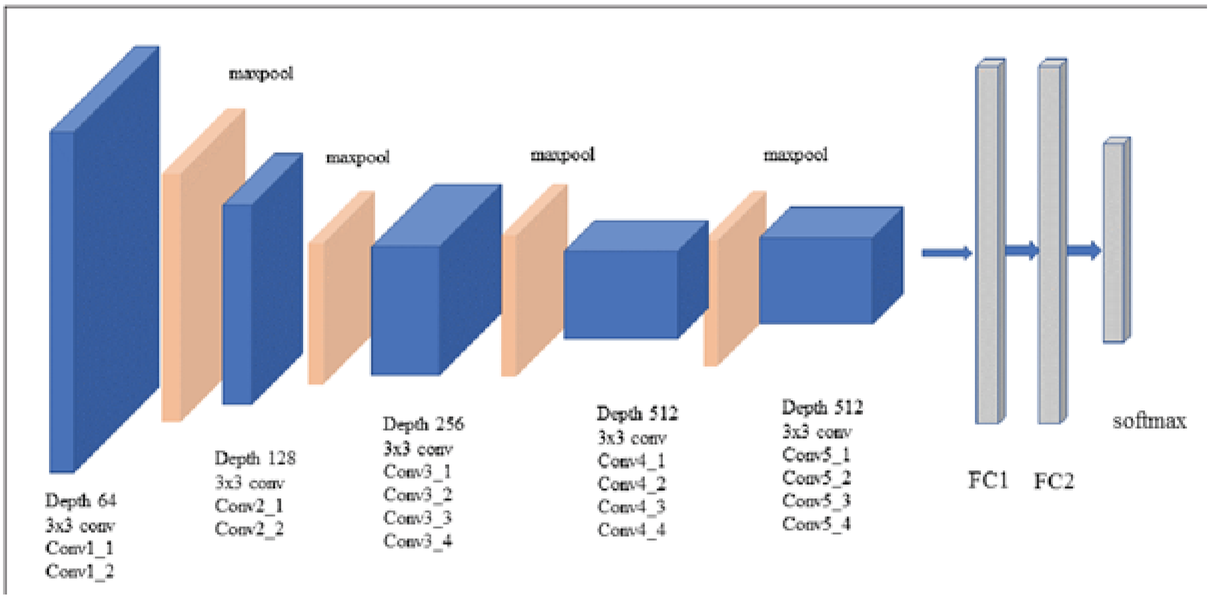

Figure 11: VGG19 Architecture

6 Evaluation

This section of research is focuses on the evaluation of the newly developed CNN-LSTM model with more sophisticated pretrained models to determine the performance of proposed model.

6.1 VGG19 Model

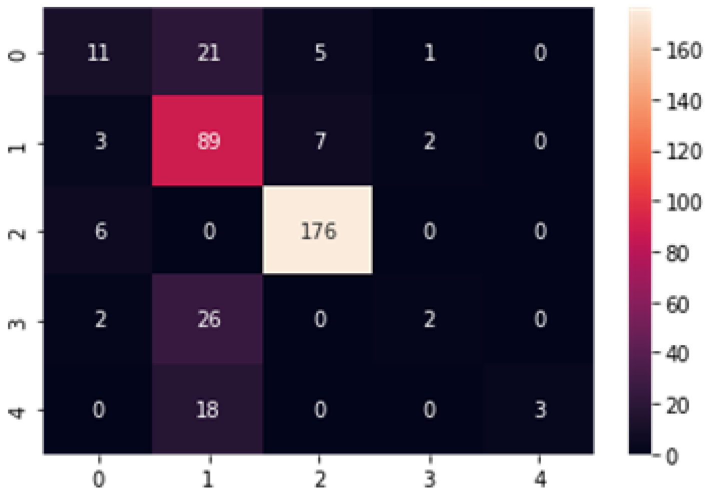

Figure 12: VGG19 Test Dataset Confusion Matrix

The first model which I implemented in this research is the VGG19 model with the modified classifier layer. This model was trained on the training dataset with a learning rate of 0.001 and weight decay rate of 0.0005 to avoid the gradient overshoot and overfitting problem.

As seen from Figure 12,the The model produced an output for the training set with accuracy of 77% i.e., It was capable of correctly classifying 2121 images out of 2744 images and validation accuracy of 76% i.e., 281 correct out of 372 indicating that the model doesn't overfit but will inaccurately detect the images 27% of the time. The training loss is approximately similar that of the validation loss. The average accuracy for the VGG19 model was 76.5%. The classification report provides a brief idea about the performance of the model as accuracy is the only parameter used for comparison of multi-class classification models shown in the Figure 13. The accuracy for this model is good whereas,the Recall and F1 score is below 50% which is not good.

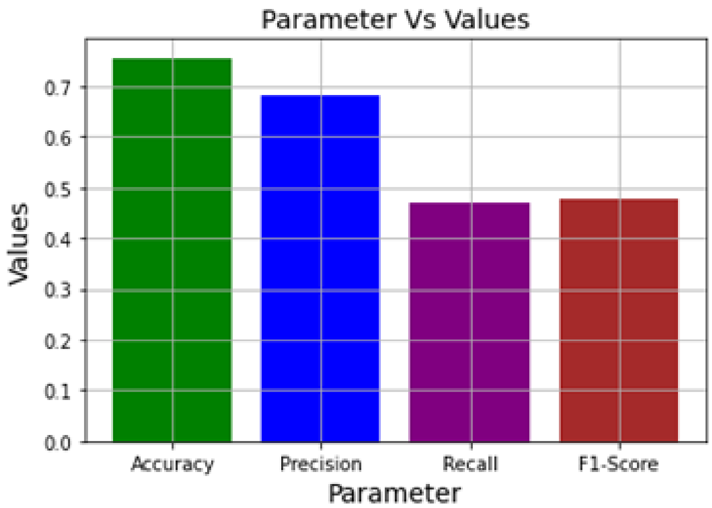

Figure 13: VGG19 Test Dataset Classification Metrics

6.2 ResNet50

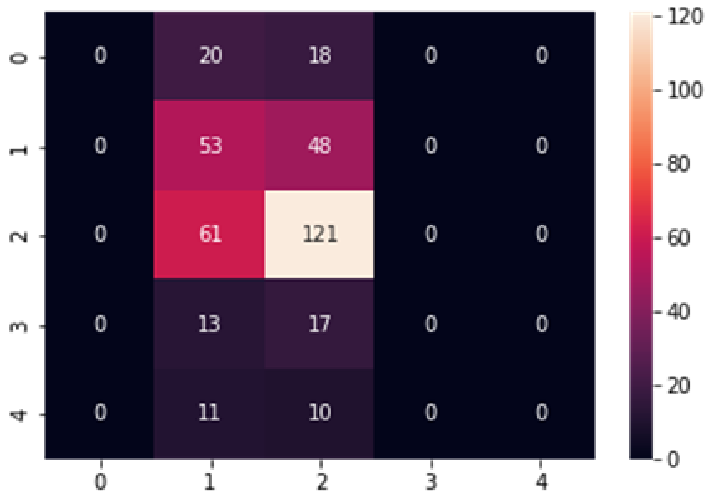

Figure 14: ResNet50 Test Dataset Confusion Matrix

The second model was the ResNet50 model with the modified classifier layer was trained on the training set for 100 epochs and the model each model was saved after every 5 epochs.The 95th epoch model was chosen for the evaluation. This model produced an output with accuracy of 72% on training set and 46% validation accuracy which is quite low and indicates that the models fail to accurately classify fundus images and also the validation loss is high as compared to the training loss as seen from the Figure 14. The model did not classify images belonging to the class 0 (No-DR), Class 3 (Severe DR), Class 4 (ProliferateDR) instead it classified images belonging to these classes to class 1 (Mild DR), Class 2(Moderate DR). The classification report for this model is shown in the Figure 15. The accuracy is greater than the precision which is not a good sign for a classification model. The Recall and F1 score are also below par.

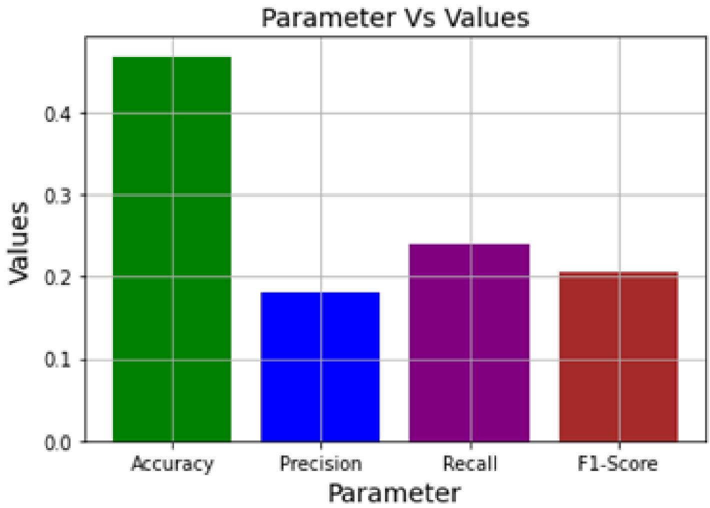

Figure 15: ResNet Test Dataset Classification Metrics

6.3 CNN-LSTM

Figure 16: CNN-LSTM Test Dataset Confusion Matrix

The final model which was implemented was designed from scratch, the model was trained for 150 epoch and the 145th epoch model was chosen for evaluation on the training set which yielded an output with accuracy of 99% which denotes that model performing well as it has already seen the images from the training set . The model's performance can only evaluated on the validation set, therefore, the model was run on the validation set which produced an output with accuracy of 76% as depicted in Figure 16 . The average accuracy for the training and validation set is 87.5% which is quite good when compared to the pretrained VGG19 model. After looking at the classification report in Figure 17 only then I was able to thoroughly compare the outputs for the CNN-LSTM model with VGG19 model. The precision is just above 50% whereas for VGG19 it was approx. 70%. The Recall and F1 score for the CNN-LSTM are slightly more than VGG19.

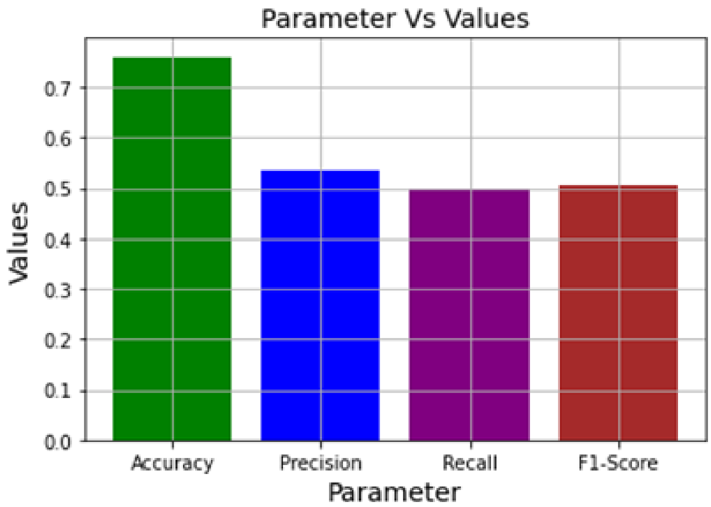

Figure 17: CNN-LSTM Test Dataset Classification Metrics

7 Conclusion

The section of report is focused on comparison of the different models used in this research. The initial steps before model training have been same for all the model such as image

Model	Dataset		
	Train	Test	Validation
ResNet50	72%	47%	48%
VGG19	77%	75%	76%
CNN-LSTM	99%	76%	76%

Figure 18: Model Performance Metric Comparison

pre-processing, transformation and yet the results of each model varies. As no model is a perfect model, but a model can be best fit for a particular set of problem. All the models implemented had one job of classifying different images according to their severity of DR. Hence, each model produced a multi-class output and comparing the results of multi-class classification model is not same as that of a binary classifier because to compare them, we need to check the classification report which includes the micro, macro, and weighted precision, recall and F1 score. It can be easily deduced that the ResNet50 model performed well when executed on the training dataset, but it produced poor results for both test and validation dataset. This performance may occur due to imbalanced dataset and sometimes due to model overfitting. The next model which I implemented was the VGG19 model, The output for all the datasets remained constant indicating that the model doesn't overfit and can accurately predict DR 75% of the time, which is a quite good. The final model of my research was the CNN-LSTM model which I designed and implemented. The model performance on training dataset is exceptional with a 99% accuracy but it dropped to 76% for the test and validation dataset, which is not as poor as compared to the ResNet50 and VGG19 models which are previously trained on large dataset. If the datasets size would have been large, then the model may deliver better performance. The consistent accuracy is a positive indicator that model performs well on unseen data and hence, can be implemented in the real-world scenario. Hence, It can be stated that the research has achieved the objective of how well does the novel CNN-LSTM model is able to classify the retinal fundus images into 5 categories depending upon the severity.

8 Future Works

The goal of every research conducted in this field has been not to eliminate the need of doctor but to aid in their diagnosis. Therefore, the primary goal of this study has been to categorize retinal fundus images into five distinct groups. Three different CNN models were employed to address the research objectives and question. A significant amount of world's population can avoid blindness if DR is detected earlier. A significant population avoids regular DR screening due to capacity constraints and sometime due to lack of availability of necessary equipment in the conventional diagnostic techniques of DR. The massive archives of retinal images which are presently accessible could be used in the long run to enhance DR detection method which may perform in accordance with the highest medical standards of diagnosis The model used in this study was able to categorize fundus pictures with reasonable accuracy, and this performance might be enhanced further by expanding the dataset size. Also, the performance of the model for detecting critical cases of DR can be improved by exposing the model to more data in order to better train it.

9 Acknowledgment

I would like to extend my sincere thanks to Prof. Vladimir Milosavljevic for his guidance and support throughout my research. He has been always patient and helped resolve all the doubts and without whom this research wouldn't have been possible.

References

Ahmad, H. M. and Hameed, S. R. (2020). Eye diseases classification using hierarchical multilabel artificial neural network, IEEE, pp. 93–98.

Amalia, R., Bustamam, A. and Sarwinda, D. (2021). Detection and description generation of diabetic retinopathy using convolutional neural network and long short-term memory, *Journal of Physics: Conference Series* **1722**: 012010.

Deperlioglu, O. and Kose, U. (2018). Diagnosis of diabetic retinopathy by using image processing and convolutional neural network, IEEE, pp. 1–5.

Doshi, N., Oza, U. and Kumar, P. (2020). Diabetic retinopathy classification using downscaling algorithms and deep learning, IEEE, pp. 950–955.

G, S., Kp, S. and R, V. (2018). Automated detection of diabetes using cnn and cnn-lstm network and heart rate signals, *Procedia Computer Science* **132**: 1253–1262.

Gangwar, A. K. and Ravi, V. (2021). Diabetic retinopathy detection using transfer learning and deep learning, pp. 679–689.

Gargeya, R. and Leng, T. (2017). Automated identification of diabetic retinopathy using deep learning, *Ophthalmology* **124**: 962–969.
URL: *https://www.sciencedirect.com/science/article/pii/S0161642016317742*

Gondal, W. M., Köhler, J. M., Grzeszick, R., Fink, G. A. and Hirsch, M. (2017). Weakly-supervised localization of diabetic retinopathy lesions in retinal fundus images.
URL: *http://arxiv.org/abs/1706.09634*

He, J., Li, C., Ye, J., Wang, S., Qiao, Y. and Gu, L. (2020). Classification of ocular diseases employing attention-based unilateral and bilateral feature weighting and fusion, IEEE, pp. 1258–1261.

Islam, M. Z., Islam, M. M. and 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.
URL: *https://www.sciencedirect.com/science/article/pii/S2352914820305621*

Li, X., Hu, X., Yu, L., Zhu, L., Fu, C.-W. and Heng, P.-A. (2020). Canet: Cross-disease attention network for joint diabetic retinopathy and diabetic macular edema grading, *IEEE Transactions on Medical Imaging* **39**: 1483–1493.

Mateen, M., Wen, J., Hassan, M., Nasrullah, N., Sun, S. and Hayat, S. (2020). Automatic detection of diabetic retinopathy: A review on datasets, methods and evaluation metrics, *IEEE Access* **8**: 48784–48811.

S., G., Gopi, V. P. and Palanisamy, P. (2020). A lightweight cnn for diabetic retinopathy classification from fundus images, *Biomedical Signal Processing and Control* **62**: 102115.

Wang, Z., Yin, Y., Shi, J., Fang, W., Li, H. and Wang, X. (2017). Zoom-in-net: Deep mining lesions for diabetic retinopathy detection.
URL: *http://arxiv.org/abs/1706.04372*

Wu, J., Zhang, S., Xiao, Z., Zhang, F., Geng, L., Lou, S. and Liu, M. (2019). Hemorrhage detection in fundus image based on 2d gaussian fitting and human visual characteristics, *Optics Laser Technology* **110**: 69–77.

Wu, L., Wan, C., Wu, Y. and Liu, J. (2017). Generative caption for diabetic retinopathy images, IEEE, pp. 515–519.

Zheng, Y., He, M. and Congdon, N. (2012). The worldwide epidemic of diabetic retinopathy, *Indian Journal of Ophthalmology* **60**: 428.