Classification of Severity Levels in Diabetic Retinopathy in Ultra-wide Field Colour Fundus Images using Hybrid Deep Learning Models

Msc Research Project
Data Analytics

Shivani Chandak
Student ID: x20186762

School of Computing
National College of Ireland

Supervisor: Mr. Taimur Hafeez

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Shivani Chandak
Student ID:	x20186762
Programme:	Data Analytics
Year:	2022
Module:	Msc Research Project
Supervisor:	Mr. Taimur Hafeez
Submission Due Date:	15/08/2022
Project Title:	Classification of Severity Levels in Diabetic Retinopathy in Ultra-wide Field Colour Fundus Images using Hybrid Deep Learning Models
Word Count:	5460
Page Count:	19

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:	17th September 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):	

Classification of Severity Levels in Diabetic Retinopathy in Ultra-wide Field Colour Fundus Images using Hybrid Deep Learning Models

Shivani Chandak
x20186762

Abstract

Diabetic Retinopathy (DR) is one of the leading causes for permanent vision loss. Blindness can impact the economic growth of a country as well as the quality of life of an individual. Early diagnosis and treatment can prevent individuals from permanent loss of eye-sight. DR is detected through a fundus imaging technique during eye examination. While the traditional fundus images of the retina provide only up to 60° view of the retina, ultra-wide field color fundus (UWF-CF)images provides up to 200° view of the surface of retina which makes it more promising and reliable. This research aims to compare and contrast the ability of deep learning models: VGG-19, VGG19-RF and VGG19-SVM to classify the severity levels in Diabetic Retinopathy using UWF-CF images. The performance of the models have been evaluated using accuracy, sensitivity, specificity and Cohen's kappa. All the models were trained on same dataset under default settings to perform comparative analysis. The results depicted that VGG-19 outperformed the hybrid models by achieving an accuracy, sensitivity, specificity and Cohen's kappa score of 80% , 95%, 80% and 0.75 respectively.

Keywords— Diabetic Retinopathy, Ultra-wide field fundus, VGG-19, VGG19-SVM, VGG19-RF

1 Introduction

Diabetic Retinopathy is considered to be one of the prime causes for permanent vision loss. (Luo and Mou; 2021) estimated that approximately 600 million people are likely to have diabetes by 2040 of which one-third of them are likely to have Diabetic Retinopathy (DR).

Diabetic Retinopathy is a condition that impacts the vision of an individual and is more common in those with high blood pressure or blood sugar. The diseases leads to development of abnormal blood vessels which results in formation of a scar tissue that causes a division between the retina and the back of the eye, because to which the patient experiences severe vision loss.

Diabetic Retinopathy can be cured if it is detected at an early stage. Due to the severity of the condition, Iceland and the United Kingdom implemented extensive routine retinal screening to prevent vision loss in working-age individuals (Danielsen; 1982). Ophthalmologists have recommended a mandate routine for retinal screening but due to the lack of human resources there has been a bottleneck in the system. With the advancement in tools and technology, use of artificial intelligence techniques and deep learning methodologies can fill in this gap. Over the years, several deep learning techniques have been implemented for detection of Diabetic Retinopathy using traditional retinal fundus images. However, this research aims to utilise ultra-wide field color fundus (UWF-CF) images of the retina instead of conventional retinal fundus images. The objective of the research is to assess the efficiency of hybrid models against deep learning algorithms to detect the grading levels of Diabetic Retinopathy (DR) in UWF-CF images.

1.1 Background & Motivation

In 2019, an estimate of 463 million individuals across the world were living with diabetes leading to an emerging pandemic and a serious health issue (Saeedi P; 2019). It is estimated that 700 million people will have diabetes by 2045, which will cause a severe increase in Diabetic Retinopathy cases (Saeedi P; 2019). The prevalence of Diabetic Retinopathy has been a major public health concern as it can lead to permanent vision loss if not addressed at an early stage. According to (Brown; 1999) and (Lynne Pezzullo and Shickle; 2018), blindness can deteriorate the quality of life of an individual and impact the economy as well. Of the common ocular retinal diseases that result in certified blindness, Diabetic Retinopathy is of epidemiological importance because it is the leading cause of vision loss.

1.2 Why use Ultra-wide field color fundus imaging technique?

Diabetic Retinopathy causes development of lesions on the retinal surface which result in vision loss. Early detection of the disease at an asymptomatic stage can allow effective treatments and prevent permanent loss of eyesight. The process of diagnosing Diabetic Retinopathy is manual as Ophthalmologists perform a thorough dilated eye exam to assess the disease.

Traditional color fundus images provided up to 60° view of the retinal surface. Initially, these images were used to examine retinal diseases. (Takahashi and Kawashima; 2017) captured a wider surface area of the retina with the help of non-mydriatic 45° color fundus images of four-field to analyze Diabetic Retinopathy staging using deep learning methodologies. The study proved that use of four-field fundus images illustrated better performance than traditional fundus images or single field fundus images. The process of acquiring a four-field fundus image requires effort and is time-consuming. However, retinal imaging techniques have advanced over the years.

At present, ultra-wide field color fundus (UWF-CF) images are used extensively as it provides a larger view of the surface of retina up to 200° in a single shot without need for dilatation of pupillary. A single shot is capable of capturing the posterior pole as well as the peripheral surface of the retina. The images include peripheral neovascularization and ischemic areas because of which they are widely accepted for detection of disease. Numerous studies have demonstrated the efficiency of the Optos system (Optos, California, USA) in the detection, classification, and diagnosis of retinal vascular disorders.

1.3 Research Question

Can hybrid models: VGG19-SVM and VGG19-RF outperform VGG19 to identify the severity of Diabetic Retinopathy (Figure 1) in ultra-wide field color fundus images?

(Engelmann and Bernabeu; 2022) states the importance of ultra-wide field color fundus (UWF-CF) images over the traditional color fundus images. Extensive research work has been performed by implementing deep learning models on traditional color fundus photographs to detect Diabetic Retinopathy. Some research work involved detecting the presence of Diabetic Retinopathy (Oh and Yoon; 2021) , while others involved classification into different retinal vascular diseases using deep learning models (Abitbol and Souied; 2022). Recently, (Liu and Galdran; 2022) implemented EfficientNet based models for grading of Diabetic Retinopathy. However, the implementation of hybrid models to detect the severity of Diabetic Retinopathy using ultra-wide field color fundus (UWF-CF) images has not been explored comprehensively.

Disease Severity Level	Findings
Grade – 0: No apparent retinopathy	No visible sign of abnormalities
Grade – 1: Mild – NPDR	Only presence of Microaneurysms
Grade – 2: Moderate – NPDR	More than just microaneurysms but less than severe NPDR
Grade – 3: Severe – NPDR	Moderate NPDR and any of the following: • > 20 intraretinal hemorrhages • Venous beading • Intraretinal microvascular abnormalities • No signs of PDR
Grade – 4: PDR	Severe NPDR and one or both of the following: • Neovascularization • Vitreous/preretinal hemorrhage

Figure 1: International Clinical Diabetic Retinopathy (DR) Severity Scale

1.4 Research Objective

The objective of the research has been explained in Table 1.

1.5 Report Structure

The research paper follows the a structured format and consists of different sections. Section 2 of the paper critically evaluates the recent literature related to the domain, Section 3 briefs about the methodological approach carried out in the research, Section 4 consists of the project design and flow, Section 5 comprehensively describes the implementation and evaluation of the models, Section 6 consists of a comparative analysis of the models developed and other state-of-art models that were previously implemented and Section 7 provides a conclusion and scope of future work.

Table 1: Research Objective

Index	Description	Evaluation Metrics
Objective 1	A critical review of the existing research work implemented in the domain in recent years	-
Objective 2	Data collection and Data Augmentation	-
Objective 3	Implementation and Evaluation of results using VGG19	Accuracy, Cohen's kappa, Sensitivity, Specificity
Objective 4	Implementation and Evaluation of results using VGG19-SVM	Accuracy, Cohen's kappa, Sensitivity, Specificity
Objective 5	Implementation and Evaluation of results using VGG19-RF	Accuracy, Cohen's kappa, Sensitivity, Specificity
Objective 6	Compare the efficiency of the developed models	-

2 Related Work

This section of the research critically evaluates and reviews the recent studies associated with Diabetic Retinopathy, Ultra-wide field color fundus (UWF-CF) images and application of hybrid models.

2.1 Related Work on Diabetic Retinopathy

(Oh and Yoon; 2021) demonstrated that, statistically, the utilization of the 7-standard field (7SF) photographs recovered from UWF-CF photography performed better than the optic disc and macula-centered image in the early treatment of Diabetic Retinopathy research. Distracting elements like eyelashes and skin were automatically segmented out of ETDRS 7SF using the proposed approach of DR detection. The ResNet-34 model21 was used to categorise images submitted to the DR detection challenge, which made use of a segmented region of interest. The study also evaluated the efficacy of the current DR detection strategy in comparison to one that just uses ETDRS Fields 1 and 2 (F1–F2). The ETDRS F1-F2 image could be utilised in place of the UWF-CF and conventional fundus images, however, ETDRS 7SF is considered to be the most relevant region for detection and diagnosis of diabetic retinopathy. Although the The model achieved an accuracy of 83.82%, it does have several drawbacks, including the fact that study only examined a subset of the retina (the ETDRS 7SF) for DR detection, as opposed to the full retina as seen in UWF-CF imaging. The research could be expanded by automatically segmenting a sizable retinal surface, encompassing the mid and distant retinal periphery, from the ultra-widefield color pictures, the research may be expanded. The planned research would integrate this expansion to examine the importance of the areas when other retinal vascular illnesses are also taken into account. Further research is required to validate the technique utilizing data from many centers, multiple devices, and multiple nationalities.

(Amalia and Sarwinda; 2021) combined Convolutional Neural Networks (CNN) with Long Short-Term Memories (LSTM). CNN was used to detect lesions on traditional retinal fundus images, and subsequently, LSTM was utilized to generate descriptive phrases regarding the lesions that have been found. Post proper training and validation, LSTM derived its input from the results produced by CNN. Upon completion of the training phase, the desired outcome was to have a model that can convert images of the retinal fundus into sentences. The study achieved an accuracy of around 90% but it considered only distinguished between a normal retinal image and a Diabetic Retinopathy image. The research proposed that training the data with more classes of DR such that the severity of the disease can be classified can be considered as future work.

(Nagasawa and Mitamura; 2019) evaluated the possibility of detecting treatment-naive Proliferative Diabetic Retinopathy (PDR) by combining ultrawide-field fundus pictures with a deep convolutional neural network (VGG-16 DCNN). Specifically, the research was focused in determining if the condition existed in the UWF-CF images. K-fold cross validation technique was applied to the dataset along-with data augmentation methods such as brightness, noise addition, gamma correction etc. This eventually, increased the learning amount of data by 18 times. The DCNN model was trained with a total of 378 photographic photos, of which 132 were Proliferative Diabetic Retinopathy (PDR) and 246 were not PDR. The model demonstrated an AUC score of 0.969, in addition to a high degree of sensitivity 94.7% and specificity 97.2%, in its performance. However, the study was only limited to naive PDR because the stage requires immediate treatment. Another drawback of the research was not taking account of Diabetic Maculopathy, which also causes disturbances in vision.

(Abitbol and Souied; 2022) used ultra-wide field colour fundus photography to assess a deep learning model's ability to distinguish between Diabetic Retinopathy, Sickle Cell Retinopathy, Retinal Vein Occlusions, and healthy eyes. UWF-CFP was used to train a multi layer DenseNet121 deep convolutional neural network to distinguish between images of vascular disease and those of healthy people. Retinal vascular anomalies were seen on the UWF-CFP pictures in 169 cases, compared to 55 of the healthy controls. A k-fold cross-validation procedure was performed on each image. Standard augmentation practices were used to improve the Adam optimizer's training results. The model was found to be 88.4% accurate after 10 iterations. Despite the accuracy of 85.2%, the DR AUC was 90.5%. The AUC and accuracy of RVO were above-average (88.4 per cent). The SCR's accuracy and area under the curve were quite high, at 96.7%. For those in good health, the ROC was found to be 88.5%, with a 95% confidence interval for accuracy. Even though the study accounted for other retinal vascular diseases, the biggest drawback was use of a small dataset for training the model and lack of external dataset for proper testing of the model. The research also used pseudo-colour UWF-CF images which artificially enhanced certain features while diminishing other artifacts.

The primary purpose of the research by (Khan and Myung; 2022) was to identify the optimal strategy for developing a deep learning model that can accurately predict systemic properties from fundus images. Everyone who scheduled an eye checkup for Diabetic Retinopathy between March 2020 and March 2021 was included in the study. The DenseNet201 framework was used to analyse 35,126 fundus pictures and 1,000,000 images from the ImageNet database in each model. More specifically, 1,270 fundus images were used to teach an AI in this study. Its efficacy was evaluated by measuring the region under the receiver operating characteristics curve (AUROC). Models trained on ImageNet data, rather than retinal images, outperformed those trained on retinal images, therefore it focused the significance of fundus shots in assessing a patient's overall health. The findings show that generalising data from picture databases can help even narrow-scope DL models. This study's findings suggest that fundus pictures can reveal important information about a patient's underlying systemic characteristics. When it comes to DL model performance, even domain-specific models can benefit from transfer learning from more broad picture sets. Simply by photographing the fundus, one can gain valuable insight on the underlying systemic characteristics. A DL model can reliably accurately predict age, ethnicity, gender, and ARB/ACE medication use. The model's highest AUC was found to be 0.926.

2.2 Related Work on Ultra wide-field fundus images

Some of the recent research based on ultra-wide field color fundus images using deep learning techniques have been summaries in Table 2.

2.3 Related Work on VGG19 and its hybrid models

A hybrid deep learning model is used in this research by (VigneshKumar and Sumathi; 2021) to identify breast cancer from mammograms. This research presented a deep convolution neural network model - VGG-19 - to enhance the specificity and sensitivity of cancer detection. The reason for using SVM as classifier because it is one of the best models to work with when using small datasets and helps in optimal feature selection. The proposed deep learning model required far less computational effort to perform preprocessing and feature selection than conventional approaches. In addition to the already potent in-built classifier, the recommended model uses a support vector machine (SVM) to increase the detection rate. The suggested performance of the model were assessed and contrasted with those of current detection methods. The suggested approach achieved an accuracy higher than 97%.

(Kaya and Tuncer; 2022) aimed to develop a deep learning-based method for analyzing static plantar pressure images to diagnose ataxia in multiple sclerosis patients (MS). This research presented a fresh, more objective approach to making an early diagnosis of PwMS. For the study, 47 people with Multiple ataxic Sclerosis and 62 healthy people posed for photographs. The processed images were then fed into several pre-trained deep learning models, including VGG16, VGG19, ResNet, DenseNet, MobileNet, and NasNet-Mobile. Parameters were used to generate the feature vector for each model. Static pressure distribution feature vectors were classified using support vector machines, k-nearest neighbours, and ANNs. The effectiveness of the classifier was evaluated by employing cross-validation. Precision, reliability, and F1-measure were used to assess the models. The VGG19-SVM model's F1 score was 95.12% accurate and 94.91% precise. The authors employed digital photographs to conduct a fully automated evaluation of the motor skills of patients suffering from ataxia. This contrast demonstrated the applicability of the suggested method to cases of ataxia with little or no outward symptoms.

(Ali and Abdullah; 2022) first gathered the chest X-ray dataset and then created various models using Machine Learning (ML) and Deep Learning (DL) techniques, notably Convolutional Neural Networks (CNN). K-nearest neighbor (K-NN), Random Forest (RF), VGG19, ResNet, DenseNet121, InceptionV3, and Xception. In order to assess the accuracy of the various models, the images were used to identify the presence of Covid-19 and other various types of lung diseases. The findings showed that VGG19 with fine-

Table 2: **Recent studies involving the use of UWF-CF images**

Author	Model(s) Applied	Performance Parameter	Problem Solved
(Abitbol and Souied; 2022)	DenseNet121	AUC, Accuracy	It was the first research of it's kind to distinguish between three types of retinal vascular diseases DR, SCR and RVO. The model attained an overall accuracy of 88.4%.
(Engelmann and Bernabeu; 2022)	ResNet34	AUC, Sensitivity, Specificity	It used UWF-CF to implement a data driven approach in identifying the regions prone to eight retinal vascular and further classifying them. The model had an AUC of 0.92
(Cai and Scott; 2021)	Inception V4	Sensitivity, Specificity and Accuracy	It focused on detecting sea fan neovascularization which would evaluate the risk of vision loss due to SCR. The model attained an overall accuracy of 97%.
(Li and Yang; 2020)	InceptionResNetV2	AUC, Sensitivity, Specificity	The study focused on detection of Retinal Hemorrhage (RH) since it is the most common symptom in case of retinal ocular diseases. The model achieved an accuracy of 98.4%
(Nagasato and Mitamura; 2018)	VGG16 & SVM	Accuracy, AUC, Sensitivity, Specificity	The study focused on detecting the presence of Central Retinal Vein Occlusion which is one of the prime ocular diseases.
(Li and Xiao; 2019)	InceptionResNetV2, InceptionV3, ResNet50, VGG16	AUC, Sensitivity, Specificity	The study focused on detection of retinal breaks and lattice degeneration which is a retinal disease that leads to severe vision loss.
(Liu and Galdran; 2022)	Singal-b7-model based on EfficientNet, EffcientNet-b4	Weighted Kappa	The study focused on image quality grading of UWF-CF images along-with DR grading using both traditional and UWF-CF images.

tuning, which had an accuracy of 94.34%, was the best model. The effectiveness of seven different categorization algorithms was assessed using Covid-19 images from the dataset. The authors, contrasted deep learning models that have been fine-tuned with those that have not. The findings indicated that fine-tuning the models improved the model's performance since it increased accuracy by 8 to 42%.

Micro expression is a type of covert human expression. In a typical discussion, it is quite difficult to see or hear this emotion since it happens so quickly. (Ibrahim and Irawan; 2021) led to the development of a system that could analyze micro-expression and display the performance of the analysis. The technique made use of a convolutional neural network (CNN) with a feature extractor based on the Visual Geometry Group - 19 (VGG19) architecture and a classifier based on random forest. Because there were numerous picture classification cases, CNN VGG19 was chosen because of its consistently strong performance. The optimal split ratio for the data is 90:10, and the optimal n estimator number is 30, with accuracy values of 99%.

A neurotic disorder called glaucoma causes the optic nerve's dynamic neurodegeneration, which impairs vision. Early glaucoma diagnosis and routine screenings with a specialist for glaucoma diagnosis may prevent it. Examining intraocular pressure and the optic Cup-Disc-Ratio allows for the diagnosis of glaucoma (CDR). This work uses computer-supported analysis from fundus pictures to accomplish innovative automated glaucoma identification. (Raja and Pitchai; 2021) developed a VGG-19 network design based on Support Vector Machines to obtain the simulation results. For glaucoma detection, the CDR threshold value of 0.41 had been employed. Fundus pictures with a CDR of 0.41 or above were considered to be glaucoma-affected, whereas those with a CDR of 0.41 or below were considered to be non-affected. The suggested glaucoma identification method utilized digital color fundus pictures. A classification precision of 94% was achieved for the collection of 175 fundus pictures.

2.4 Research Gap

Most of the studies focus on identifying the presence of the disease using UWF-CF images and not staging. While a recent study by (Liu and Galdran; 2022) focuses on the severity levels in Diabetic Retinopathy, it does not test the ability of hybrid models in its implementation. Hybrid models have proven to be efficient for feature extraction and classification in various different domains. This research aims to bridge the gap of assessing the performance of hybrid models to distinguish between the severity levels in Diabetic Retinopathy using UWF-CF imaging technique.

3 Research Methodology

To achieve the goals of the research, the suggested study uses a Knowledge Discovery in Databases (KDD) approach. The idea focuses on the implementation of suitable data mining techniques and refers to the process of finding information in data. The major areas of interest while thinking about this process include pattern recognition, databases, statistical analysis, artificial intelligence, knowledge acquisition of systems, and data visualization techniques. Figure 2 demonstrates the modified approach that is being used in the study.

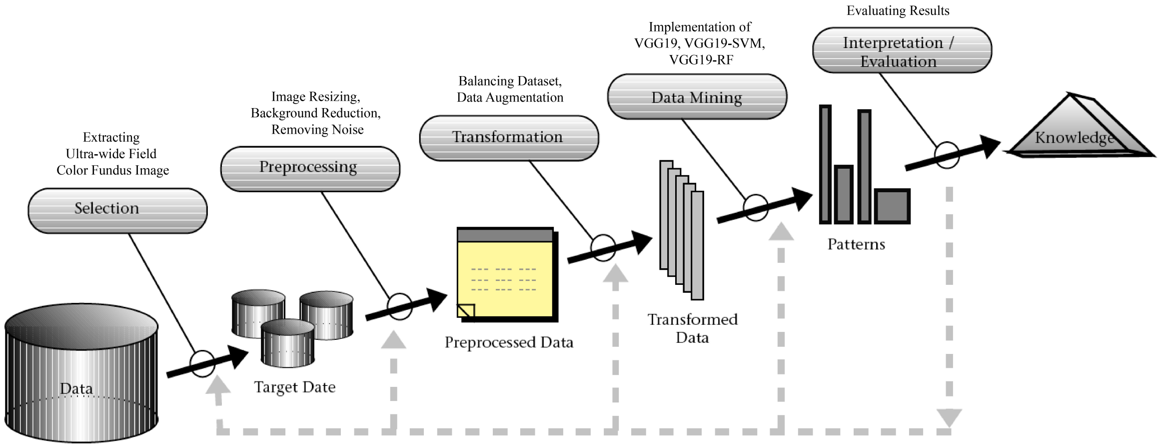

Figure 2: Knowledge Discovery in Databases (KDD) Process

3.1 Data Collection

In order to determine the severity of Diabetic Retinopathy (DR) from ultra-wide field color fundus images, the DeepDR Diabetic Retinopathy Image Dataset (DeepDRiD) is used. This dataset comprises of 154 images for training and an additional 50 images for testing. In the training dataset, of the total images, 44 images belong to Grade 0 (No apparent retinopathy), 43 images belong to Grade 1 (Mind-NPDR), 44 images belong to Grade 2 (Moderate-NPDR), 17 images belong to Grade 3 (Severe-NPDR) and only 6 images belong to Grade 4 (PDR), refer Figure 3 for reference. Optomap P200Tx (Optos, Dunfermline, UK) was used to capture ultra-wide field retinal pictures, all of which are focused on the optic disc. The photographs were in the jpg file type and had a resolution of 3900 x 3072 pixels. Each picture was roughly 2 MB in size. The UWF-CF images in the database were photographed by a retinal specialist at an eye clinic in the Department of Ophthalmology of the Sixth People's Hospital, a Shanghai Jiao Tong University affiliate, between January 2019 and October 2019. Over 200 images were collected from 128 individuals in this process.

Figure 3: No retinopathy, Mild-NPDR, Moderate-NPDR, Severe-NPDR, PDR (L to R)

3.2 Data Pre-processing

A data mining approach called data pre-processing dramatically changes raw data into a format that is effective, practical, and easier to understand. This is a crucial step to getting the data into the right structure, format, and accuracy for the model. Its label

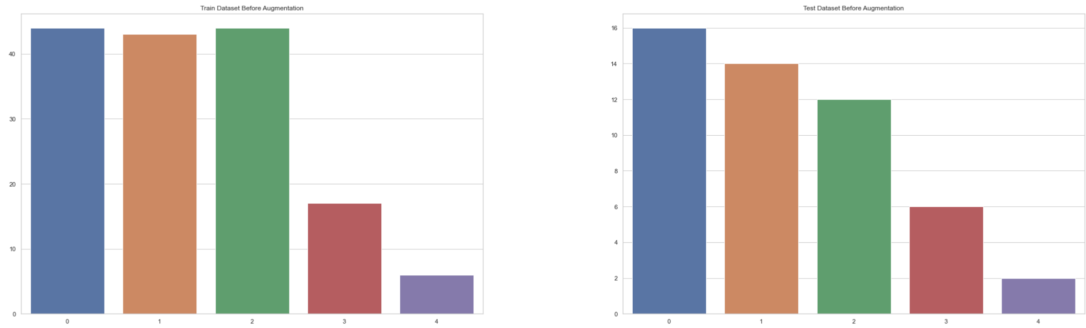

Figure 4: Imbalanced Data Before Augmentation

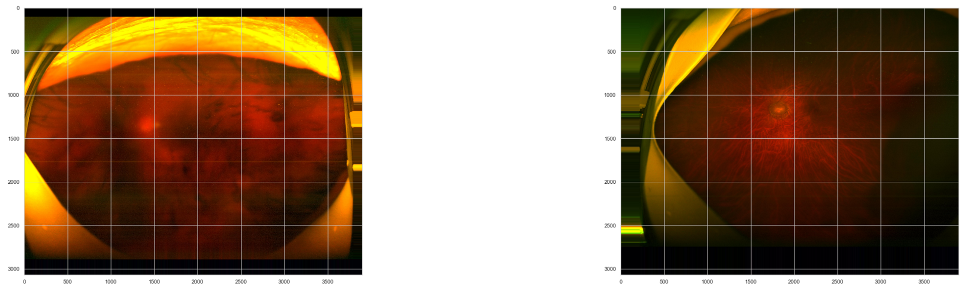

Figure 5: Images after Data Augmentation

was indicated by the image filename and folder name saved in a CSV file. Each image was imported and given a label based on its folder index. Training models having high resolution (3900,3072) require more computational power. Each image is reshaped to a standard size of 120x120.

3.3 Data Transformation

The translation of data from one structure to another is known as data transformation. It is vital to data management and integration processes. The UWF-CF images are divided into sub-folders based on it's severity from Grade 0 to Grade 4 for both training and test datasets. The two objectives of this step was - having a balanced dataset and an increased number of images. Initially, the dataset has unequal number of images in every class for both train data and test data as seen in Figure 4. Balanced dataset refers to the concept where every class has equal number of images and helps building a more refined model. Data augmentation techniques such as rotation, zoom, horizontal flip, vertical flip etc (Figure 5). were applied to achieve a variety in the dataset. Smaller datasets generally result in over-fitting of the model and therefore require data augmentation methods. The data augmentation techniques were applied separately on both training and test datasets to avoid any biased results (Figure 6). Data augmentation was carried out using Image Data Generator from the Keras library. The UWF-CF images were re-scaled and randomly shuffled to enhance the convergence and stability of the model. K-fold validation (k=5) was used in 80:20 (80% training, 20% validation) ratio for every epoch such that images were generated in batches. Since, the images are composed of 8-bit channels of RGB (Red, Green, Blue), they were normalized by dividing it by 255 such that they could be represented on a binary scale.

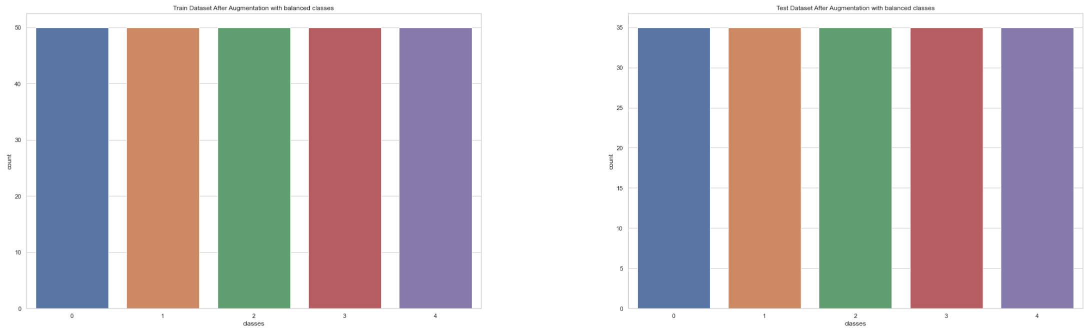

Figure 6: Balanced Data After Augmentation

3.4 Data Mining and Model Design

This research aims to predict the severity of Diabetic Retinopathy (DR) in UWF-CF images by using three selected models: VGG19, VGG19-SVM and VGG19-RF. In case of hybrid models, feature extraction is performed using transfer learning capabilities of VGG19 which is a pre-trained model on ImageNet. The extracted features are then fed as an input to classifiers such as Support Vector Machine (SVM) and Random Forest (RF) for image classification in case of VGG19-SVM and VGG19-RF respectively. All the models are trained and implemented using Keras and TensorFlow.

3.5 Evaluation and Interpretation

It is a crucial step in a life cycle of deep learning models to interpret, assess and compare performance metrices to choose the best model for the stated problem. For this research, the model performance is evaluated by using metrics like accuracy, Cohen's kappa, sensitivity, specificity and confusion metrices. It is essential to analyze the sensitivity and specificity of the models in the medical domain, as the ability of the model to correctly classify both positive and negative cases needs to be examined.

4 Design Specification

Figure 7 depicts the project design flow which has been followed in the research. The process of the research begins with data collection and pre-processing. The images are divided into 5 different folders based on their respective classes i.e. Grade 0 to Grade 4. The images are stored along-with their labels using arrays in python. All the images are then re-sized to attain a standard format. Since, the dataset used for the research is small, data augmentation is performed in the next step to attain balanced number of images in every class. Augmentation techniques like horizontal flip, rotation, zoom etc. are applied to add variety and increase the size of the dataset which will also help in reducing overfitting of the model. K-fold cross validation helps to maximize the performance of the model and minimizes the risk of getting biased results. To yield the images in batches, K-fold cross validation technique is applied where the images are randomly shuffled followed by division into training data and testing data. The process is followed five times as the value of k is chosen to be 5. Post data pre-processing and transformation, the images are then fed into a convolutional neural network architecture, Visual Geometry Group (VGG-19) for feature extraction. VGGNet models tend to perform better than

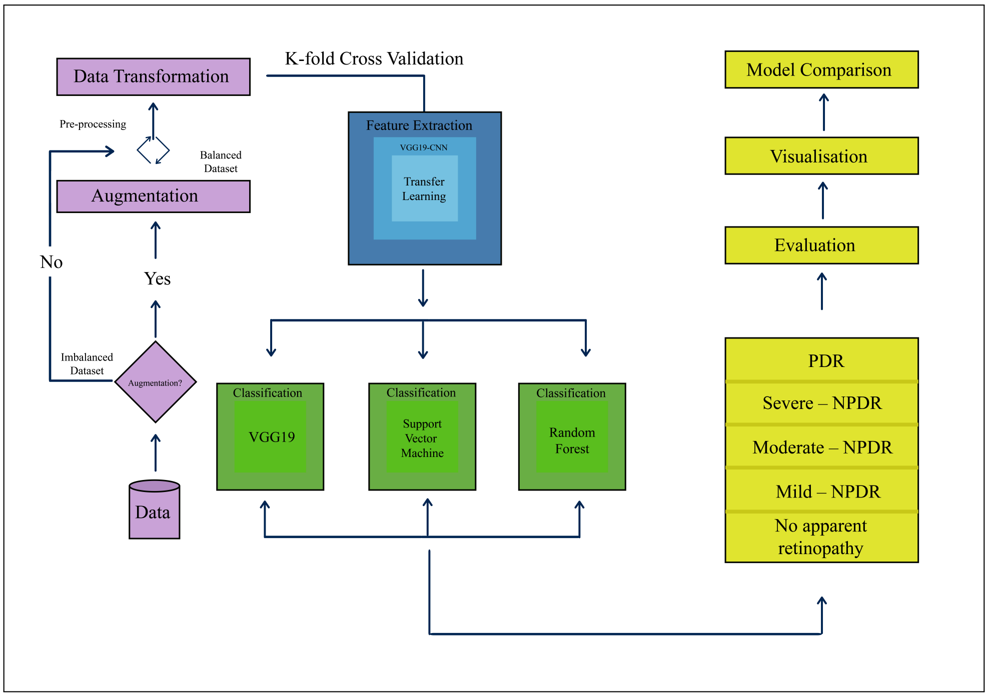

Figure 7: Project Design Flow

some pre-trained models as they substitue numberous small kernel-sized filters with big kernel-sized filters. Three models are implemented in this research: VGG19, VGG19-RF and VGG19-SVM. Random Forest is used as a classifier in VGG19-RF model and SVM is used as a classifier in VGG19-SVM model. Therefore, the output of VGG19 feature extraction is then fed as an input to the classifiers Random Forest and SVM for classification in VGG19-RF and VGG19-SVM respectively. The models are finally assessed on its ability to classify the UWF-CF images into five classes using performance metrics such as accuracy, Cohen's kappa score, sensitivity and specificity. Visuals such confusion metrices are also implemented for better evaluation.

5 Implementation & Evaluation

This section of the research paper focuses on the implementation and evaluation of the deep learning models for classifying the UWF-CF images as per the severity levels in Diabetic Retinopathy. In the first phase, a deep CNN architecture, VGG-19 is implemented. The second phase involves implementation two hybrid models: VGG19-RF and VGG19-SVM. The output from the VGG19 module is reshaped and given as an input to Random Forest and SVM as input to perform classification in VGG19-RF and VGG19-SVM respectively. The model's ability to determine staging in Diabetic Retinopathy is tested using a separate dataset consisting of 175 UWF-CF images.

5.1 VGG-19

5.1.1 Implementation

VGG-19 is a network architecture that comprises of a total of 19 layers that has 16 convolution layers and 3 fully connected layers. The accuracy of the neural network can be increased by increasing the number of layers as it enhances its ability to process more precise features. The convolution layers are trainable which are essentially 3x3 network layers. It is connected with max-pooling layer which is used to minimise the output size from the convolution layer. The convolution operation is performed on every pixel of the UWF-CF input image. The input data consists of UWF-CF images of the shape (120,120,3). The model uses pre-trained ImageNet weights for training. The model comprises of 4 dense layers on the custom top for classification. The first, second, third and fourth layer consists of 200, 100, 50 and 5 neurons respectively. Relu Activation function was implemented to substitute negative characteristics with 0 and take into account positive features that might have an impact. In addition, ReLu has the ability to maintain the original non-linearity and prevents issues of vanishing gradient in the UWF-CF images. Since, the model has five classes, Softmax Activation function is implemented to produce one output value for every node present in output layer. ”Crossentropy” is used as loss function since image classification is performed in this research. The model also consists of dropout layers which is added to overcome over-fitting of the model. Adam optimizer is used to handle sparse gradients and has faster computation time.

5.1.2 Evaluation

The model achieved an accuracy of 80% and a kappa score of 0.75. The average sensitivity of the model is found to be 0.95 while the average specificity of the model is 0.80. Figure 8 provides a depiction of the number of images that got classified correctly and incorrectly. The model faced an issue while classifying Grade 1 and Grade 4 level of severity which is why the specificity for both the levels is comparatively much lower than others as seen in Table 3.

Table 3: Sensitivity & Specificity for VGG-19

Class	Sensitivity	Specificity
Grade 0	0.922	0.972
Grade 1	0.922	0.600
Grade 2	0.914	0.857
Grade 3	0.992	1.000
Grade 4	1.000	0.571

5.2 VGG19-Random Forest (VGG19-RF)

5.2.1 Implementation

Random Forest is one of the most powerful machine learning techniques for classification problems. In this hybrid model, the UWF-CF images are given as input in the shape of (120,120,3) to the VGG19 component. The loaded layers are marked as non-trainable as pre-trained weights are used. VGG19 is used for feature extraction from the images.

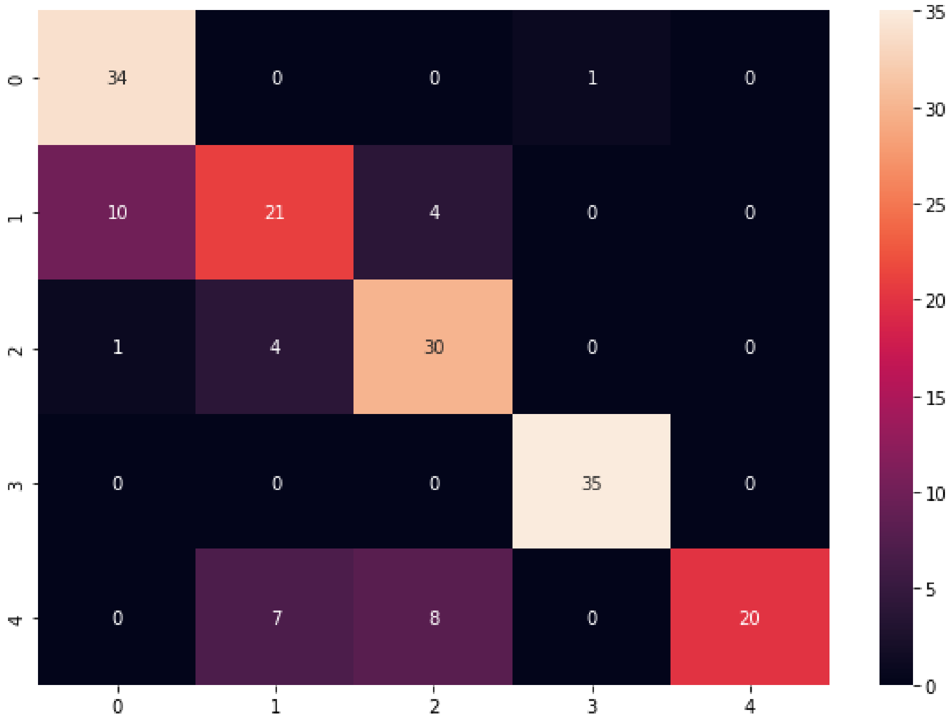

Figure 8: Confusion Matrix for VGG-19

The features are then re-shaped to be fed as an input parameter to Random Forest for classification. Depth is not defined so that the model can refer to the best fit. The data is partitioned up and new variables are generated for each tree as part of the bootstrap sampling procedure used by the method. Every variable is trained, verified, and a conclusion is provided by the algorithm. The majority vote is conducted once each tree has provided its conclusion, and the result is the majority vote's conclusion. Random Forest takes the role of the last fully connected layer of the VGG19 model.

5.2.2 Evaluation

The model achieved an accuracy of 56% and a kappa score of 0.45. Figure 9 depicts that the model as successfully able to classify Grade 3 level of severity accurately and failed to classify Grade 0, 1 and 2 significantly. The average of sensitivity and specificity of the model is 0.89 and 0.56 respectively. The model is capable of classifying the positive cases much more accurately as compared to the negative cases as seen in Table 4.

Table 4: Sensitivity & Specificity for VGG19-RF

Class	Sensitivity	Specificity
Grade 0	0.835	0.485
Grade 1	0.907	0.314
Grade 2	0.921	0.285
Grade 3	0.935	1.000
Grade 4	0.850	0.714

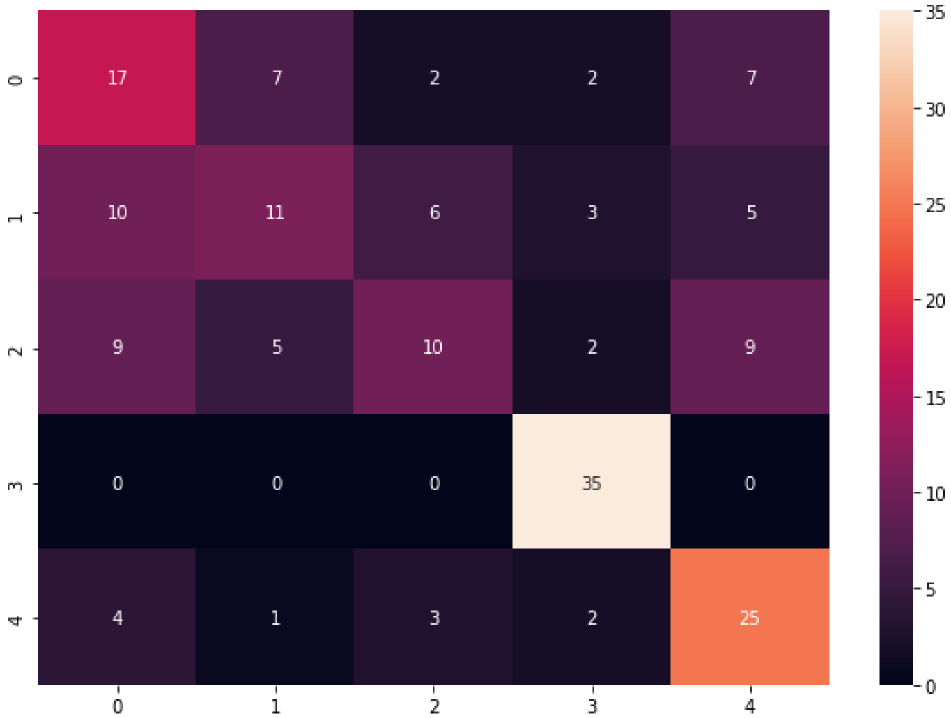

Figure 9: Confusion Matrix for VGG19-RF

5.3 VGG19-Support Vector Machine (VGG19-SVM)

5.3.1 Implementation

Support Vector Machines (SVM) is one of the most widely used machine learning models that is used either as a single model or a hybrid with CNN. ref6 states that the model is known for providing better accuracy in smaller data sets as the performance of the model reduces when the size of the dataset is large. The hybrid VGG19-SVM provides optimal feature selection and is therefore used in numerous research works. The images are shaped into 120x120 and fed into the VGG19. The VGG19 architecture uses pre-trained ImageNet weights for feature extraction. The extracted features are then reshaped and fed as an input to the SVM model. The SVM architecture uses poly kernel and he degree of optimization that it needs to meet is set to 20. This parameter helps to define how much of misclassification needs to be avoided.

5.3.2 Evaluation

Table 5: Sensitivity & Specificity for VGG19-SVM

Class	Sensitivity	Specificity
Grade 0	0.664	0.371
Grade 1	0.950	0.142
Grade 2	0.892	0.371
Grade 3	0.928	0.685
Grade 4	0.835	0.514

The model attained an accuracy of 42% and a kappa score of 0.27. Figure 10 depicts that the model failed significantly to classify Grade 0, 1 and 2 levels of severity correctly. The average sensitivity of the model is 0.85 which is almost twice the average specificity of the model 0.41 (Table 5). This means that the model has the capability to classify the images with the disease correctly and lacks the capability of classifying images without the disease accurately.

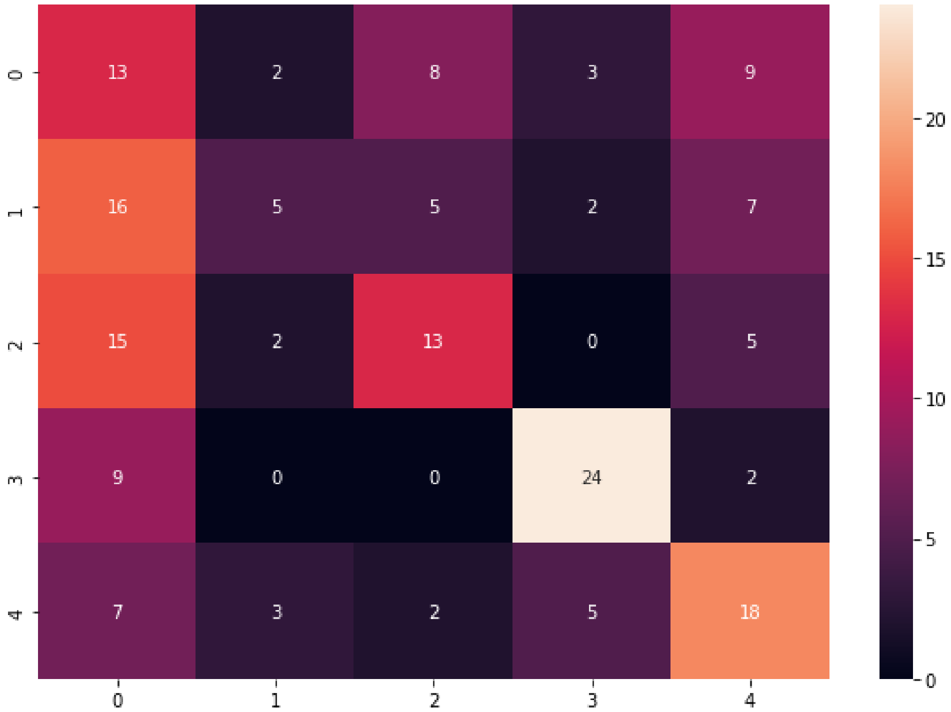

Figure 10: Confusion Matrix for VGG19-SVM

6 Comparison of Results & Discussion

This section compares and contrasts the performance of the three deep learning models that have been applied- VGG-19, VGG19-SVM and VGG19-RF in this research. The performance of the models are also compared with other existing models that have been implemented previously for grading of Diabetic Retinopathy in ultra-wide field fundus color fundus (UWF-CF) images. (Liu and Galdran; 2022) used Singal-b7-model based on EfficientNet and achieved a score of 0.82. In another approach, it used EfficientNet-b4 and attained a kappa score of 0.86. The study also experimented on using the concept of pseudo-loop labelling technique as described in (Tan; 2019) with CycleGAN while implementing EfficientNet-b4 which resulted in a score of 0.90. On the other hand, VGG-19 had a score of 0.75 which is considered to be a substantial agreement. As depicted in Table 6, VGG-19 has an accuracy of 80% which is the highest amongst all the three models. The hybrid models VGG19-RF and VGG19-SVM had an accuracy of 56% and 41% which is not as efficient as compared to other models. The main reason for this could be lack of number of images in the training data. With the increase in the number of images in every class, there could be a possible increase in accuracy. Sensitivity and Specificity are two very important factors that need to be assessed in deep learning models in the medical domain. All the three implemented models have a high sensitivity but the hybrid models have a very low specificity. The Cohen's kappa score of VGG-19 is 0.75 which is almost twice of what the hybrid models have achieved. Overall, it can be concluded that the VGG-19 model outperformed both the hybrid models when trained on the same dataset comprising of default settings towards grading of Diabetic Retinopathy (DR) from UWF-CF images.

Table 6: Performance Matrix for all models

Model	Accuracy	Sensitivity	Specificity	Cohen's Kappa Score
VGG-19	80%	0.95	0.80	0.75
VGG19-RF	56%	0.89	0.56	0.45
VGG19-SVM	42%	0.85	0.41	0.27

7 Conclusion & Future Work

Diabetic Retinopathy (DR) has always been one of the prime retinal vascular disease for concern to prevent the young generation from developing permanent vision loss. Diabetic Retinopathy can be cured if diagnosed and treated at an early stage. Hence, this research was conducted to test the ability of hybrid models to distinguish between different stages of Diabetic Retinopathy in comparison to the conventional deep learning model. The models VGG-19, VGG19-RF and VGG19-SVM were designed to classify the staging of diabetic retinopathy into five classes (Garde 0 to Grade 4) using ultra-wide field color fundus (UWF-CF) images. The performance of the models is evaluated using accuracy, Cohen's kappa score, sensitivity and specificity. The steps taken during data pre-processing, transformation and augmentation are kept uniform for all the three models i.e. VGG-19, VGG19-SVM and VGG19-RF in order to do rational comparative analysis. The models VGG-19, VGG19-SVM and VGG19-RF attained an accuracy of 80%, 42% and 56% respectively. As per the results, both the hybrid models were not able to classify the staging of Diabetic Retinopathy accurately. In future, similar datasets can be combined to have more number of images in every class. To create a more realistic environment, other retinal abnormalities can also be taken into consideration. In addition, this research work can be extended by concatenating multiple deep neural networks for achieving better performances.

8 Acknowledgement

I would like to express my sheer gratitude towards Mr. Taimur Hafeez who provided me with constant guidance and support throughout the project and made the completion of this research possible.

References

Abitbol, E., M. A. E. J. M. C. A. F. K. S. O. M. and Souied, E. (2022). Deep learning-based classification of retinal vascular diseases using ultra-widefield colour fundus photographs, *BMJ open ophthalmology* .

Ali, A. and Abdullah, K. (2022). Identification of covid-19 on chest radiographs: A new approach to early diagnosis, *International Journal of Health Sciences* .

Amalia, R., B. A. and Sarwinda, D. (2021). Edetection and description generation of diabetic retinopathy using convolutional neural network and long short-term memory, *In journal of physics: conference series.IOP Publishing* .

Brown, G. (1999). Vision and quality-of-life, *Transactions of the American Ophthalmological Society* .

Cai, S., P. F. U. M. G. M. H. G. and Scott, A. (2021). Deep learning detection of sea fan neovascularization from ultra-widefield color fundus photographs of patients with sickle cell hemoglobinopathy, *JAMA ophthalmology* .

Danielsen, R, J. F. H. T. (1982). Prevalence of retinopathy and proteinuria in type 1 diabetics in iceland, *PatteActa medica Scandinavicarns* .

Engelmann, J., M. A. M. I. P. E. S. A. and Bernabeu, M. (2022). Detection of multiple retinal diseases in ultra-widefield fundus images using deep learning: data-driven identification of relevant regions, *arXiv preprint* .

Ibrahim, A.F., R. S. S. C. and Irawan, B. (2021). Micro-expression recognition using vgg19 convolutional neural network architecture and random forest, *In 2021 4th International Symposium on Agents, Multi-Agent Systems and Robotics (ISAMSR)* .

Kaya, M., K. S. and Tuncer, S. (2022). Detection of ataxia with hybrid convolutional neural network using static plantar pressure distribution model in patients with multiple sclerosis, *Computer Methods and Programs in Biomedicine* .

Khan, N.C., P. C. D. E. C. K. M. V. M. P. D. D. L. T. and Myung, D. (2022). Predicting systemic health features from retinal fundus images using transfer-learning-based artificial intelligence models, *Diagnostics* .

Li, Z., G. C. N. D. L. D. Z. Y. C. C. X. Y. X. F. J. C. Z. X. and Yang, Y. (2020). Development and evaluation of a deep learning system for screening retinal hemorrhage based on ultra-widefield fundus image, *Translational Vision Science Technology* .

Li, Z., G. C. N. D. L. D. Z. Y. C. C. Z. L. X. F. J. C. Z. X. and Xiao, H. (2019). A deep learning system for identifying lattice degeneration and retinal breaks using ultra-widefield fundus images, *Annals of translational medicine* .

Liu, R., W. X. W. Q. D. L. F. X. Y. T. S. J. T. S. L. J. G. Z. and Galdran, A. (2022). Deepdrid: Diabetic retinopathy—grading and image quality estimation challenge, *Patterns* .

Luo, X., P. Z. X. Y. W. W. K. S. J. D. X. Y. B. H. J. and Mou, L. (2021). Diabetic retinopathy screening: A systematic review on patients' non-attendance, *Mvdrnet: Multi-view diabetic retinopathy detection by combining dcnns and attention mechanisms, Pattern Recognition* .

Lynne Pezzullo, Jared Streatfeild, P. S. and Shickle, D. (2018). The economic impact of sight loss and blindness in the uk adult population, *BMC Health Services Research* .

Nagasato, D., T. H. O. H. M. H. E. H. I. N. S. T. K. M. N. M. H. K. and Mitamura, Y. (2018). Deep neural network-based method for detecting central retinal vein occlusion using ultrawide-field fundus ophthalmoscopy, *Journal of ophthalmology* .

Nagasawa, T., T. H. M. H. E. H. N. M. O. Z. Y. Y. O. H. and Mitamura, Y. (2019). Accuracy of ultrawide-field fundus ophthalmoscopy-assisted deep learning for detecting treatment-naïve proliferative diabetic retinopathy, *International ophthalmology* .

Oh, K., K. H. L. D. L. H. S. K. and Yoon, S. (2021). Early detection of diabetic retinopathy based on deep learning and ultra-wide-field fundus images, *Scientific Reports* .

Raja, J., S. P. and Pitchai, R. (2021). An automated early detection of glaucoma using support vector machine based visual geometry group 19 (vgg-19) convolutional neural network, *Wireless Personal Communications* .

Saeedi P, Petersohn I, S. P. (2019). Global and regional diabetes prevalence estimates for 2019 and projections for 2030 and 2045: results from the international diabetes federation diabetes atlas, *Diabetes Res Clin Pract* .

Takahashi, H., T. H. A. Y. I. Y. and Kawashima, H. (2017). Applying artificial intelligence to disease staging: Deep learning for improved staging of diabetic retinopathy, *PloS one* .

Tan, M. Le, Q. V. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks, *ICML* .

VigneshKumar, M. and Sumathi, N. (2021). A hybrid deep learning approach for breast cancer detection using vgg-19 and support vector machine, *International Journal of Mechanical Engineering* .