Detecting Diabetic Retinopathy from Retinal fundus images using DC-CNN

MSc Research Project
Data Analytics

Ananya Pratap Singh Chandel
Student ID: x19237529

School of Computing
National College of Ireland

Supervisor: Prof. Noel Cosgrave

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Ananya Pratap Singh Chandel
Student ID:	x19237529
Programme:	Data Analytics
Year:	2021
Module:	MSc Research Project
Supervisor:	Prof. Noel Cosgrave
Submission Due Date:	16-08-2021
Project Title:	Detecting Diabetic Retinopathy from Retinal fundus images using DC-CNN
Word Count:	6763
Page Count:	20

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	
Date:	

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

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

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

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

Detecting Diabetic Retinopathy from Retinal fundus images using DC-CNN

Ananya Pratap Singh Chandel
x19237529

Abstract

Today one of the major causes of blindness in adults is diabetic retinopathy (DR). It is a progressive disease and can be categorized into four stages according to its severity. Early detection of DR can save individuals from developing permanent blindness. Hence, the governments run several screening programs to prevent DR. DR detection remains a problem due to the limited number of trained clinicians who can perform the diagnosis. Hence need to develop an automated DR detection system is evident. This research aims at developing a novel dual-channel convolutional network (DC-CNN) for the detection of DR using retinal fundus images. DC-CNN performs a binary classification under two labels DR and No DR. The designed DC-CNN utilizes two channels for deeper feature extraction, designed using the VGG16 transfer learning model and customized tuned CNN model. The performance of the designed DC-CNN has been evaluated using evaluation metrics like accuracy, sensitivity, specificity, and F1 score. ResNet50 and Inception v3 are trained on the same retinal fundus image dataset to perform a comparative analysis of the DC-CNN model. The Designed DC-CNN model outperforms all the models to produce an accuracy of 95.23% and sensitivity of 96.94%.

1 Introduction

Diabetic retinopathy has turned out to be a leading cause of loss of vision. It is estimated that by 2040, around 600 million individuals are expected to have diabetes, and one-third of them may have diabetic retinopathy (DR) (Luo et al.; 2021). DR is a disease that affects both the eyes and is most prevalent in patients whose blood sugar or blood pressure is too high. The blood vessels that supply blood to the retina are affected by DR, which affects the light signals sent to the brain. Proliferative DR (PDR) is the advanced stage of DR and may lead to permanent blindness. Mild non-proliferative DR (NPDR) is an early stage of DR that can be corrected if detected on time. Due to the seriousness of the disease, Iceland and the United Kingdom conduct systematic retina screening nationwide to prevent blindness among working-age adults. All professionals have suggested routine DR screening is a mandate, yet there is a bottleneck in screening programs due to limited human assessors.

The advancement in computational capabilities and the use of deep learning algorithms which learn from large available datasets can surpass human assessor capabilities. Several deep learning algorithms have been applied to detect DR with high sensitivity and specificity to minimize the vision-threatening DR. There are several challenges involved while integrating deep learning techniques to detect DR using retinal fundus images. Using deep learning techniques for DR detection is to develop models that can deliver screening at the community level. This research discusses the design and implementation of a novel dual-channel convolutional neural network (DC-CNN). The DC-CNN aims at achieving higher sensitivity and specificity for detecting DR using retinal fundus images.

1.1 Background & Motivation

The majority of the working-age population is a risk of developing diabetic retinopathy (DR). DR is expected to affect over 93 million people across the world. Progression to blindness can be prevented if DR is detected in its early stage, although it's a challenging task to detect as DR shows few symptoms in the early stages. At present, there are several bottlenecks in detecting DR. DR detection can be a time-consuming process as it is detected manually, where clinicians are trained to examine retinal fundus images and, based on the fundus photograph of the retina, examine the disease. The results are often delayed due to the reviewers' late submission of results, which leads to delay in treatment due to lost follow-ups. An affected retina shows vascular abnormalities, which are caused due to the disease. These abnormalities lead to the formation of lesions that the trained Clinicians can examine towards detection of DR.

Since the number of people having diabetes is increasing day by day, there is a strong need for and infrastructure and skilled clinicians who can detect DR to prevent blindness. But the equipment and expertise required are not enough as the number continues to grow, and the rate of diabetes in the local population continues to increase. DR detection is of utmost importance. Hence, the need for a comprehensive automatic method for detecting DR has been recognized long back. Several efforts have been made towards detecting DR by using deep learning algorithms, image classification, machine learning, and pattern recognition. The aim is to achieve a DR detection system that can be used at the community level and detect DR with clinical standards. This motivates the development of the DCCNN model that aims at improving the specificity and sensitivity towards the detection of Diabetic retinopathy.

1.2 Research Question & Objective

1.2.1 Research Question

RQ: To what extent can the designed dual-channel convolution neural network (DC-CNN) provide better sensitivity and specificity in detecting diabetic retinopathy using retinal fundus images?

1.2.2 Research Objective

The research objectives can be seen in the Table 1 below.

Table 1: Research Objective

Index	Description	Evaluation Metrics
Objective 1	A critical review of the existing work done in recognizing diabetic retinopathy and dual-channel neural networks from 2018 to 2021	-
Objective 2	Collect, collate and augment the images before analyzing the data for extracting features that will differentiate DR from Non-DR images	-
Objective 3	Implementation, Evaluation, and Results using DCCNN	Accuracy, Sensitivity, Specificity, and F1 score
Objective 4	Implementation, Evaluation, and Results using Inception v3	Accuracy, Sensitivity, Specificity, and F1 score
Objective 5	Implementation, Evaluation, and Results using ResNet 50	Accuracy, Sensitivity, Specificity, and F1 score
Objective 6	Comparing the developed model with the state of the art	-

2 Related Work

2.1 Related work on Diabetic retinopathy

In ophthalmology, convolution neural networks (CNN) have demonstrated outstanding performance in detecting illnesses such as cataracts, glaucoma, and diabetic retinopathy. Jeong et al. (2020) used Retinal fundus pictures for information extraction about the anatomy of the eye. Before a surgery, assessment of eye health is critical. For determining eye health, axial ocular length (AL) is identified to be a critical characteristic. For feature extraction, a CNN model with parallel convolutions and different kernel sizes were employed. But, it causes issues such as vanishing gradient, which occurs as the model becomes more complex. The problem is addressed with ResNet. ResNet uses a skip connection, which allows for the creation of deeper CNNs without compromising the performance of the model. After that, convolutional layers were stacked on sequentially stacked ReLU layers. Before activation, a stochastic method called batch normalization was employed to alter the input distribution. For training, the proposed model 1296 retinal images were used while using 272 images to test for AL detection. To access the model's ability, MAE is used as a loss function, and for assessing the ability of the regression model, R2 is used. The developed model detected AL with R2 of 0.67 and MAE of 0.9. The AL determined the clinicians could use it for an early diagnostic. Similarly, DR can be detected using fundus images to give a preliminary diagnostic for ophthalmological examination.

Fundus pictures might be used to diagnose diabetic retinopathy thanks to advances in deep learning technology (DR). The majority of systems rely on standard fundus images. Ultra-wide fundus (UWF) embodiments were utilized by Oh et al. (2021), as the retinal surface covered by UWF was 82 percent. Using ultra-widefield fundus images and applying deep learning, they presented a diabetic retinopathy diagnosis method by utilizing a picture segmentation software ETDRS 7S. ResNet-34 was utilized as a residual network with deep learning architecture that comprised 34 layers for classification and DR detection. Because the UWF fundus pictures were so large, pixels with a lot of intensity were disregarded. Adam optimizer was used for optimization while keeping the learning rate to be 0.0001. Based on the UWF photographs, a DR detection system was implemented, which resulted in an AUC of 91.50 and 83.3 percent accuracy in DR detection. The technique had a flaw in that UWF photos needed to be aligned to avoid obstructions like eyelids and eyelashes. Such stumbling blocks could be avoided by using normal retinal fundus images. This motivates using retinal fundus images for diabetic retinopathy detection.

Artificial neural networks are usually not used for image classification purposes due to the development of various other new deep learning architectures. Harun et al. (2019) focused on detecting Diabetic retinopathy by classifying images as DR or non-DR using artificial neural networks. They used Bayesian Regularization (BR) and Levenberg-Marquardt (LM) to train their Multi-Layered Perceptron (MLP) for performing the classification of data. For classification, the network utilized nineteen features that were extracted from fundus images as inputs. The model was evaluated by varying the number of hidden nodes for analysis. It was found that the use of LM led to poor performance when compared with MLP, which was trained using BR with classification performance of 67.47% testing and 72.11% training. This study showed potential for using BR in other artificial neural networks. Poor classification efficiency was observed on blurry and low contrast images, which could be used by utilizing new neural model architectures like a Dual Channel Convolutional Neural Network (DC-CNN).

A clinical technician usually diagnoses fundus images by looking at it, which makes it hard for them to recognize the presence of lesions, hence making the detection of the disease difficult. Gayathri et al. (2020) realized that automated detection of DR could be a challenging task in which feature extraction can play a crucial role. Compared with older hand-crafted methodologies, Convolutional neural networks have a superior performance in image classification efficiency. Their work used a novel CNN architecture that was used to extract features from retinal images and later used as input for machine learning classifiers. The model was evaluated using various classifiers like J48, Random Forest, SVM, Naïve Bias on image datasets like MESSIDOR and from Kaggle. Classification efficiency was calculated by comparing precision, specificity, and accuracy for every classifier. The study showed that the J48 classifier performed the best classification on MESSIDOR and other datasets with an accuracy of 99.89% for binary classification. This study showed potential for combining CNN with other models to improve the classification accuracy for DR detection.

Deperlıoğlu and Köse (2018) utilized deep learning and image processing to diagnose

Diabetic retinopathy by using retinal fundus images. They used enhancement techniques like histogram equalization, V transform algorithm, and HSV for enhancing retinal fundus images. Lastly, the retinal fundus images were passed through a Gaussian low pass filter. Once the images were pre-processed, a Convolutional Neural network was used to perform classification. Four hundred images were used to assess the performance of the proposed model from the Diabetic retinopathy detection database of Kaggle. The classification was performed for every stage of image pre-processing. To find the average of values, twenty experiments were performed at each stage. Experiments resulted in an accuracy of 97% and specificity of 93.33%, and sensitivity of 96.67%. Results showed high efficiency towards detection of Diabetic retinopathy using retinal fundus images. The image pre-processing techniques used in this research can be used with other Hybrid models that might help in enhancing the results further.

A technique called Microaneurysm Retinal vein Haemorrhage Exudate (MRHE) for feature extraction and hybrid pre-processing was proposed by Zubair et al. (2020). MRHE used Edge detection (FEED) and feature enhancement to extract image features involving very little complexity. An efficient deep convolutional neural network called the D-CNN model was used to classify DR. Salient features like MA's, haemorrhages and retinal veins were used to train the D-CNN model that was extracted using image pre-processing techniques on raw images. The proposed novel model was able to classify DR on a Structured analysis of the retina (STARE) database that comprised of retinal fundus images. The proposed model was compared with existing DR classification models like ANN, SVM, etc. This study suggests new models can be developed enhancing the performance of existing CNN models that may outperform the existing architectures to attain better results towards the classification of Diabetic retinopathy using retinal fundus images.

Amalia et al. (2021) utilized a mix of two deep learning architectures to identify Diabetic Retinopathy using retinal fundus images: Long Short-Term Memory (LSTM) and Convolution Neural Network (CNN). GoogleNet was the CNN model utilized in their paper. A summary of the features in retinal fundus pictures was included in the output. The picture characteristics were fed into LSTM as a vector with a sentence description. Two deep learning architectures, CNN and LSTM, were used to describe and identify DR, with the model achieving 90% accuracy. The model's output, a descriptive phrase, would aid radiologists in their diagnosis. The study makes no mention of the disease's severity. The research shows potential for combining deep learning models with retinal fundus pictures to detecting diabetic retinopathy by fundus imaging.

2.2 Use of Dual Channel Convolutional Neural networks

For Image classification, deep neural networks are capable of giving deep extracted features. Yang et al. (2018) performed image classification of hyperspectral images (HIS) by designing a dual-channel convolutional neural network (CNN). The first channel was designed to extract hierarchical spectral features to gain maximum advantage from HSI images, while the second channel is used to extract the hierarchical spatial-related feature. The two channels were designed using two customized DenseNets. The designed model was trained experimentally by dense growth rates and several widen factors to attain optimal performance and tuning hyperparameters. The designed dual-channel neural architecture, when compared with baseline models, resulted in higher classification accuracies. Although the performance achieved by the proposed network was highest, the proposed network took longer training time. This study motivates the idea of designing a dual-channel convolutional neural network to extract extra features from retinal fundus images and classify it to detect Diabetic retinopathy.

Poliyapram et al. (2019) proposed a new dual-channel CNN for the classification of Polarimetric synthetic aperture radar (PolSAR) images. A dual-channel CNN is proposed to extract abundant spatial information from a PolSAR image and improve classification results. Both the channels are designed using two separate CNN architectures. These channels can extract two sets of features that are later concatenated to achieve the final classification result. The research showed improved classification results, but further by adding more labeled samples for training could have been improved. Proposed dual-channel CNN showed promising improvements over other single-channel models. This design can be improved by using transfer learning in one of the channels for extracting more features.

With the advancement in computer vision applications, it can also be applied in the domain of civil infrastructure. Nowadays, inspection and monitoring of concrete structures is assisted by computer vision techniques. Kumar and Ghosh (2020) proposed a crack detection system based on CNN for detecting cracks in concrete structures. The crack detection system was designed using the proposed Dual Channel Convolutional Network (DuCCNet) model that works on two channels working parallelly. The model was optimized further to deal with the vanishing gradient problem and increase stability. The first channel was implemented with 21 hidden layers, while the second channel is designed using seven layers. The proposed dual-channel model resulted in high validation accuracy of 92.25 towards the classification of concrete defect data.

Identifying smoke is important for fire prevention and safety warning systems used in the industries. Due to the complicated color, texture, and shape, it remains a challenge to detect smoke from an image. Gu et al. (2019) addressed this problem by designing a dual-channel CNN. The first channel of the network is composed of multiple convolutional layers, and max-pooling layers selective batch normalization is applied to prevent overfitting and accelerate training. The first channel is utilized for extracting detailed features like the texture of smoke. In the second channel, along with convolutional and max-pooling layers, skip connection, and global average pooling are used to avoid overfitting and vanishing gradient problems. The second channel captures basic information from a smoke image like contour. Finally, both the layers are concatenated to complement each other's performances. The results obtained from the designed dual-channel CNN beat the state-of-the-art performance by resulting in an accuracy of 99.5% over a public dataset. This research motivates the idea of using two channels for extracting features from retinal fundus images and then concatenating them to complement each other's performance for the detection of diabetic retinopathy (DR).

3 Methodology

An enormous amount of data is being generated from multiple sources. Extracting useful information from the data that would help in decision-making is crucial. Currently, there is an abundance of datasets for diabetic retinopathy detection. This research is carried out using Knowledge Discovery in Databases (KDD) with few modifications. KDD helps find, transform, and extract meaning from raw retinal fundus data towards detecting diabetic retinopathy by designing a new dual-channel convolutional neural network. Figure 1 represents the modified KDD Methodology used in the research.

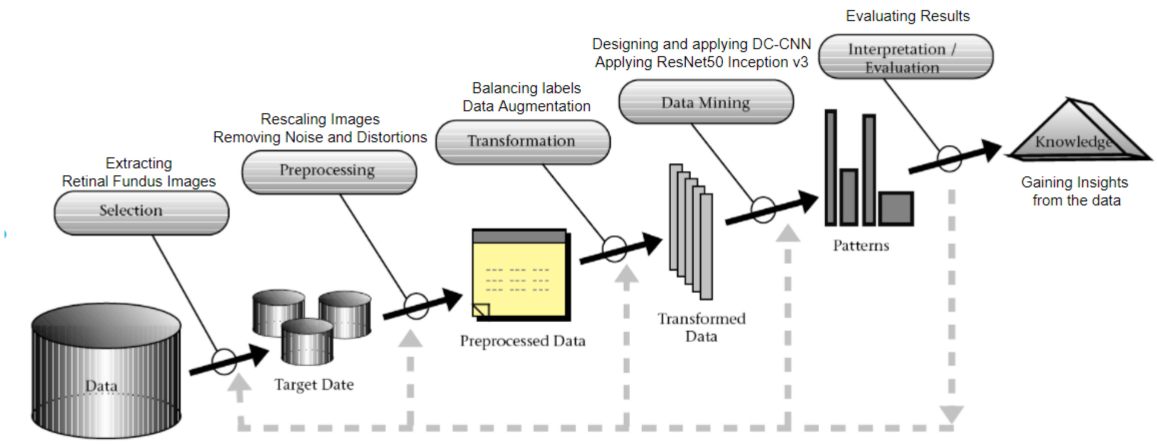

Figure 1: KDD Methodology

3.1 Data Collection

To detect diabetic retinopathy (DR) from retinal fundus images, the dataset is collected from Kaggle, a public repository. The images are obtained from the Diabetic Retinopathy Gaussian filtered dataset, a subset of the dataset named APTOS Blindness Detection[1] that contains retinal images taken under different imaging conditions by fundus photography. Clinicians labeled all the images in the dataset as per the severity from 0-4, Where 0 represented no DR (1805 images), 1 represents mild DR (370 images), 2 represents moderate DR (999 images), 3 states severe DR (193 images), and 4 represents proliferative DR (295 images). Along with the images, a train.csv file has image names, related severity, and stage of DR. Figure 2 shows the distribution of images and labeled categories.

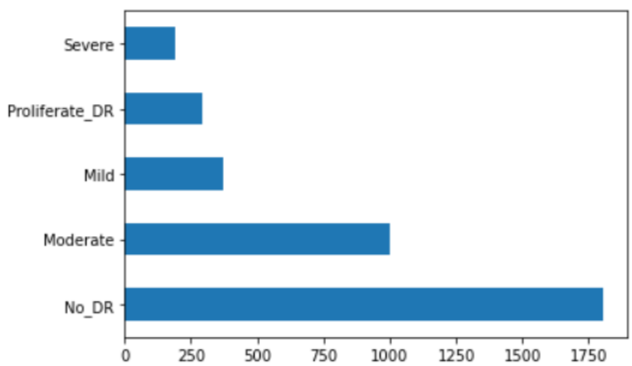

Figure 2: Number of images as per severity

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

3.2 Data Preprocessing

Data pre-processing is a data mining technique that significantly transforms raw data into an efficient, useful and more understandable format. The dataset used has images and labels stored in CSV format. The first task performed was to map the images to their appropriate labels as the research aims at performing binary classification for detecting diabetic retinopathy (DR. The dataset has been categorized under two labels, one having DR while the other with no DR. This also solved the imbalanced number of images per label, as shown in Figures 2 and 3. Retinal images collected from different fundus cameras are prone to noise and at times underexposed or overexposed to prevent these problems. The data selected is gaussian filtered and free from noise and detail. The standard image size of 224 X 224 pixels is used throughout this research as it is the standard size for most pre-trained and deep learning models.

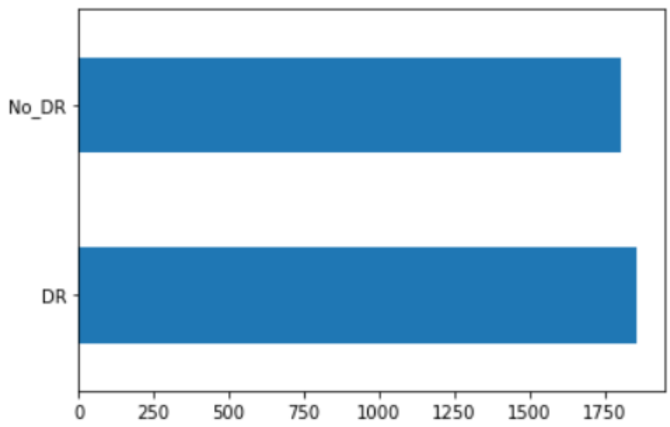

Figure 3: Balanced images under two labels

3.3 Data Transformation

Data transformation involves the conversion of data from one structure or form to other. It plays a crucial role in data management and integration activities. Data is initially divided into subfolders according to the severity of DR mapped with its labels in a CSV file. The data was split into train and test using stratified random sampling in an 80/20 ratio. After splitting the data images were copied in the created working directory to store images under train, test folder with labels as No DR and DR. For improving the performance of any deep learning or machine learning model data augmentation is performed, it helps the model to generalize better by avoiding overfitting. The accuracy and performance of a model can be improved if the data is sufficient. Image data generator from Keras library is used for performing data augmentation. The images are reshaped to 224X224 pixels. To increase the convergence and stability, the images are normalised by using the rescale function of the Image data generator and shuffled randomly. Some images were augmented by performing zoom and horizontal flip to increase the number of images for training. Train and test generators are created to identify 2,949 images belonging to 2 classes and 773 images belonging to 2 classes for testing. A processed and labeled retinal image can be seen in figure 4.

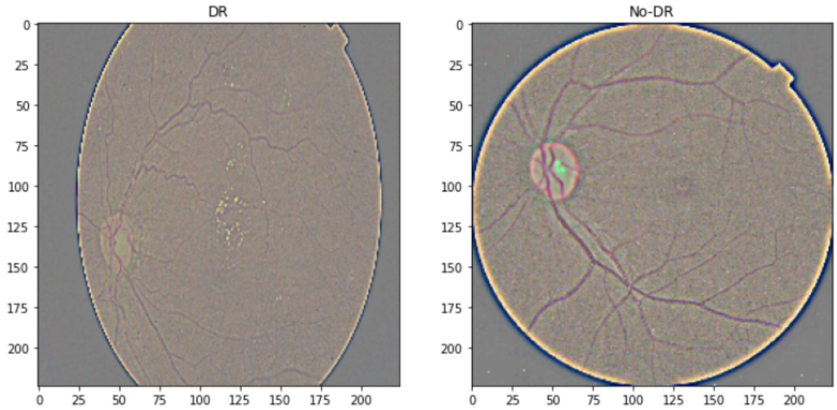

Figure 4: Rescaled retinal fundus images with labels shaped to 224*224 pixels

3.4 Designing model and data mining

Discovering correlations, patterns, and anomalies using a large dataset and predicting outcomes gives a broad idea about data mining. This research aims at predicting diabetic retinopathy (DR) by designing a dual-channel convolutional neural network model (DC-CNN). Proposed DC-CNN is made using two channels that work parallel to each other, and are later combined using fully connected layers to produce output. The first channel aims to utilize the capabilities of transfer learning by using VGG16, a pre-trained model on ImageNet, to extract features and for the second channel using a convolutional neural network. Along with DC-CNN, ResNet 50 and Inception V3 models are applied on the retinal fundus images to detect DR and for comparing the performance of the designed DC-CNN model. All the models are trained using TensorFlow and Keras, frontend and backend.

3.5 Evaluation and Interpretation

Performance evaluation and interpretation is an essential step in any Deep learning lifecycle. It allows us to choose the best model in terms of performance metrics and compare the performances of the model. In the medical domain, sensitivity and specificity play an important role in the detection of disease. The models are evaluated and compared based on the performance metrics like confusion metrics, accuracy, f1 score, sensitivity, and specificity. Lastly, all the results are visualized to create graphs.

4 Project Design Specification

4.1 Project Design Flow

Figure 5 represents the project design flow carried out for the research.

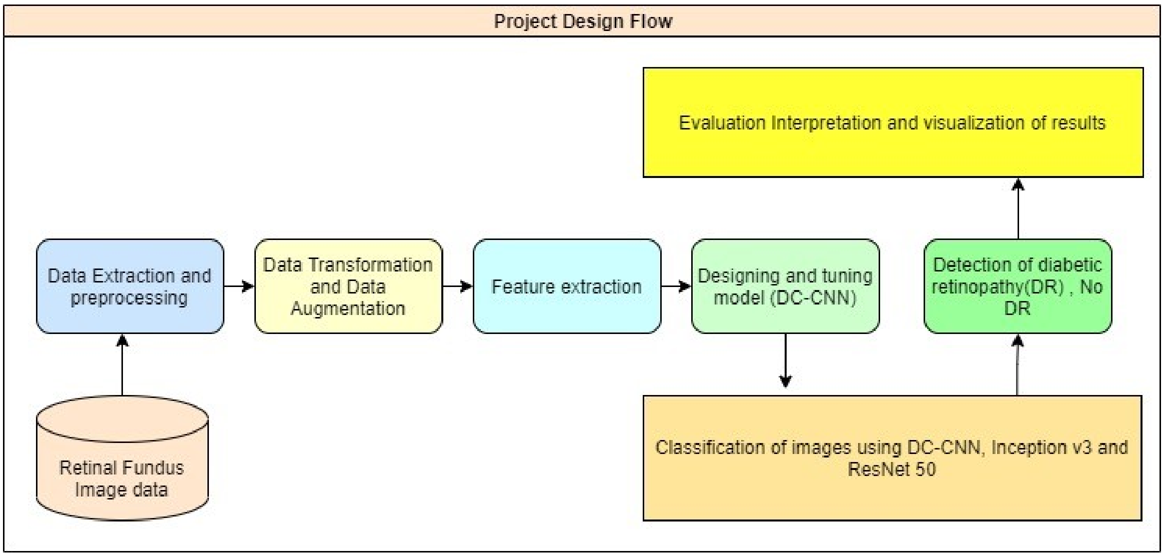

Figure 5: Project Architecture

4.2 Design of Dual-channel convolutional neural network

For feature extraction, deep learning is currently a state-of-the-art technique. Compared with traditional machine learning methods, Deep Neural networks can automatically learn hierarchical representations from large datasets. Convolutional neural networks (CNN) have become a successful architecture for extracting features for most computer vision applications. As we increase the layers in the neural network possibility of extracting a better feature vector increases, but it may lead to problems like gradient explosion or vanishing gradient. These problems lead to the limited depth of a network. A dual-channel architecture was used by Poliyapram et al. (2019) to extract richer features and obtain more spatial features for PolSAR image terrain classification. The dual-channel network obtained two feature sets utilizing more spatial information for giving better classification results. Pre-trained models like VGG16 are trained to classify 1000 different categories and can be used as the backbone for extracting image features and good classification results without utilizing much computational resources.

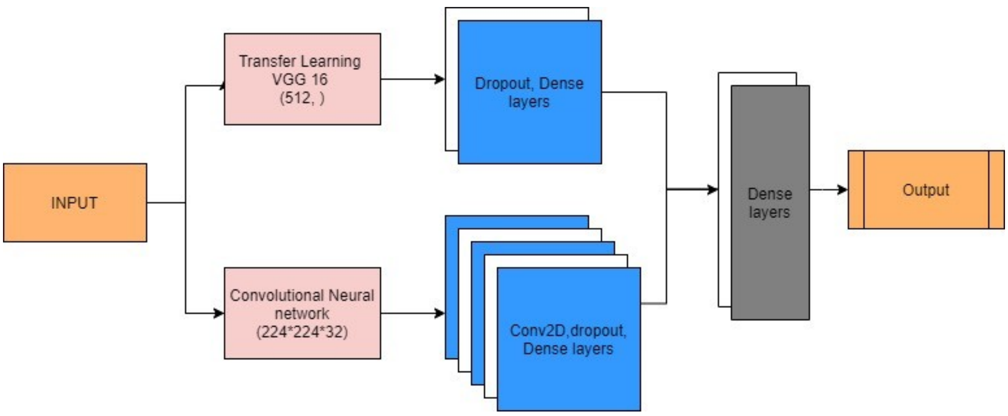

Figure 6: Architecture of dual-channel convolutional neural network

We followed dual-channel architecture based on the best understanding of reviewed dual-channel architectures to design a convolutional neural network with two channels. The first channel utilizes the VGG16 architecture based on transfer learning and can extract images, generalized and even detailed features. The second channel utilized the capabilities of the Convolutional neural network (CNN). The architecture's first channel consists of a simple feed-forward neural network that leverages the input features (i.e., 512 features) of VGG 16. The second channel in the architecture consists of custom-designed convolutional neural network architecture that leverages the input features 224 * 224 * 3 (i.e., 1,50,528 features) of CNN, as shown in Figure 6. After collecting the inputs, the network separates into two channels for processing, and equal output of features is generated (i.e., 256 units). Later, the outputs are combined using two fully connected dense layers to perform binary classification of diabetic retinopathy under two labels, i.e., DR and No DR.

5 Implementation

5.1 Dual-Channel Convolutional Neural network (DC-CNN)

5.1.1 Implementation

Feature extraction and data preparation for both the channels (VGG16 and CNN)

The first channel of designed DC-CNN uses VGG16 architecture. For preparing data for the first channel, the feature extractor function is used to extract essential features from the images by utilizing the ImageNet weights of the VGG16 architecture. After applying feature extractor on both train and test diabetic retinopathy directories, images with extracted essential 512 features are obtained. Originally the images contained 1,50,528 features (224*224*3), from which extracted 512 features are passed to the first channel of the DC-CNN model.

The second channel is designed using customized CNN architecture. No data preparation is required to input images in the second channel of DC-CNN. The input images are read from the directories created during data transformation, and 1,50,528 features with input image shape of (224*224*3) is fed into the second channel of the DC-CNN model.

Dual-Channel – CNN Architecture

The first channel consists of a simple feed-forward neural network that leverages the input of 512 features of VGG Net 16. It further consists of the following layers in the feed-forward neural network:
• Dropout Layer: The input features consisted of 512 features or units. We tried to reduce these to 256 units to make them compatible with our convolutional neural network architecture and merge their output in the end. Also, the model fit faced overfitting issues due to 512 units in the output. Hence a dropout of 50 percent is performed on the outgoing hidden units or nodes.

• ReLu Activation Function: This function is implemented to nullify the impact of negative features by replacing them with 0 and considering only positive impacting features. Another functionality provided by ReLu is maintaining the original non-linearity of the images.

• Dense Layer: Finally, after disabling the 50 percent of incoming hidden units to Dense Layer, the final hidden units in this layer were initialized to 256 later to merge it with the output from the second channel.

The second channel in the architecture is designed using a customized CNN that uses 150,528 input features from retinal images. Hyperparameter optimization is performed using Keras Tuner to determine the number of layers and units utilized for the designed CNN architecture. Different configurations with different layers and filters are used for creating a model. Later Randomised search from Keras tuner is used to determine the best model by considering validation loss as an objective function across different models. Keras tuner is trained using original DR train and test images to determine the best model with the least validation loss. The CNN used in the second channel is implemented by taking into consideration the best model produced by Keras tuners randomized search. Along with tuned parameters, few additional parameters are tweaked based on acquired knowledge and test runs to get optimal performance from the network and making it compatible with respect to the first channel. It further consists of the following layers in the Convolutional Neural Network architecture:

• Input Data: Consisted of all the input training retinal fundus images of the shape (224*224*3)

• Convolutional Layer: The model consisted of 4 convolutional layers with hidden units obtained by the best model of Keras Tuner. These layers were utilized to develop accurate feature maps by leveraging appropriate feature detectors.

- The first convolutional layer uses the minimum number of hidden units, i.e., 32, and feature detector or filter size of (3,3), which is efficient in working with the input shape of images.
- In the second convolutional layer, hidden units are gradually increased to 64, and the feature detector or size of the filter increases to (5,5).
- Finally, Layer 3 and 4 were set with maximum hidden units to capture 128 most essential features to be passed to the final output layer.

• Pooling Layer: Max Pooling is implemented on the feature maps obtained from the Convolutional Layer to focus on the important features or extract features containing the maximum value.

- The filter is set to (2,2) throughout the convolutional layers, thereby ensuring that we capture almost all the important features in the images.

• Dropout Layer: The dropout Layer is added to overcome the final model's overfitting. Instead of dropping the maximum percentage of hidden units from the single-layer, 25 percent of hidden units from each layer are dropped to maintain uniformity in the model and improve the fit.

• Flatten Layer: After all the convolutional, pooling, and dropout layer implementation. The final 15,488 features were flattened into a 1D array to create input for the next layer.

• ReLu Activation Function: This function was implemented to nullify the impact of negative features by replacing them with 0 and considering only positive impacting features. Another functionality provided by ReLu is maintaining the original non-linearity of the images and preventing vanishing gradient problems.

• Dense Layer: For obtaining the same number of features as that of the first input channel, Fifty percent of features are randomly dropped out. And the initialized units in the last layers are set to be 256.

Once the architecture of both the channels is finalized, and equal output of 256 units is obtained for both channels, it needs to be combined for performing classification of retinal fundus images. The output is combined using fully connected dense layers with a dropout of 50 percent to disable hidden units from the combined dual-channel network and propagate only relevant features to the fully connected network. Lastly, a dense layer with one hidden unit generates the output using the sigmoid activation function. The sigmoid activation function is used as the network will perform binary classification of retinal fundus images. The images are classified under two labels diabetic retinopathy (DR) or no diabetic retinopathy (No DR). The final DC-CNN model is compiled using Adam and RMSProp Optimizer, in which Adam optimizer resulted in the best results. The loss function used for the model is binary cross-entropy since the model is performing binary classification.

Learning curves (accuracy and loss on train test data) and accuracy are used to evaluate the performance of the model. Finally, after running multiple instances of the model from 50 to 100 epochs and batch sizes of 16,32,64 with different hyperparameters and configurations, the model is trained on 50epochs and batch size of 64 as it provides the best optimal results in classifying retinal images to categories DR and No DR.

5.1.2 Evaluation and Interpretation

The model is trained multiple times by changing the network's type, size, and the number of layers. Best results are obtained by performing hyperparameter optimization on one channel, consisting of a convolutional neural network using Keras tuner. Along with tuning, the model parameters like activation functions, loss functions, and the number of epochs is varied to attain the best optimum performance by the designed model. Figure 7 represents the loss curve of the model and the accuracy of the model. These Learning curves are used to determine and optimize the performance of the designed DC-CNN model by diagnosing underfitting or the overfitting problem on the training and validation data. The loss curve seen in Figure 7 shows a generalization gap as training loss is almost lower than validation loss, and the training loss gradually decreases to stability point. The learning accuracy curve shows a sudden spike and then increases gradually for both training and validation. Overall, the DC-CNN model illustrates a decent fit when trained for 50 epochs with a batch size of 64. DC-CNN model resulted in an accuracy of 95.22%

for training and 95.23% accuracy for validation data.

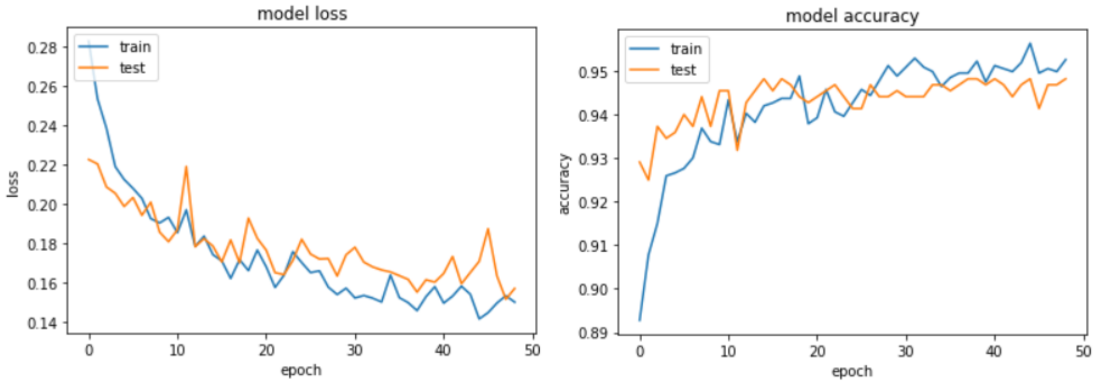

Figure 7: Learning curves for loss and accuracy of DC-CNN model

For Interpreting the model's performance, the trained DC-CNN model is tested with 733 test images to predict images with DR, and no DR. Table 2 represents the confusion matrix for results predicted by the DC-CNN model on 733 test images. The model's sensitivity is calculated to be 96.94%, specificity to be 93.58%, and F1 score of 0.9523.

Table 2: Confusion Matrix for DC-CNN model

	No DR (Predicted)	**DR (Predicted)**
No DR (Actual)	350 (TN)	011 (FN)
DR (Actual)	24 (FP)	348 (TP)

5.2 Inception v3

5.2.1 Implementation

For a wide variety of computer vision solutions (state-of-the-art) tasks, convolutional networks form the core. Inception v3 has been widely used for image recognition and has 78.1% accuracy in the ImageNet dataset. Li et al. (2019) proposed transfer learning approach utilizing Inception v3 network for classifying retinal fundus photographs for detection of diabetic retinopathy. Here Inception-v3 is used as the base CNN. The architecture consists of eleven inception modules, five convolutional layers, two max-pooling layers, one average pooling, and a fully connected layer to perform image-wise classification. Inception-v3 formed a dense structure by sparse grouping nodes, increasing the network's depth and breadth and reducing calculation time. Inception v3 can detect 1000 categories of images as it is trained on the ImageNet dataset from scratch. The inception v3 is used to perform the classification of retinal fundus images for detecting Diabetic retinopathy and comparing its performance with the proposed dual-channel convolutional neural network (DC-CNN). The network is trained using pre-trained weights from the ImageNet database. The input size is 224*224*3, which is kept similar to the pre-processed retinal fundus images. The layers of the base model are not trained, and lastly, a fully connected layer is added to perform binary classification with a sigmoid activation function. The trained model uses default parameters with RMSprop as an optimizer, learning rate as 0.0001, and loss function as binary cross-entropy as it is a binary classification.

5.2.2 Evaluation and Interpretation

The model is trained for 50 epochs to maintain uniformity in all implemented models. A batch size of 64 is selected to train the model as it results in the optimal performance of the model. The accuracy and loss learning curves for the model can be seen in figure 8 below. A steep increase in accuracy and a steep decrease in loss can be seen in the first to fifth epoch. After ten epochs, a gradual improvement in the training and validation(test) accuracy can be seen in the figure 8. While the train and test loss keeps on decreasing and finally attains stability. The model didn't improve the accuracy after the 50th epoch; hence it is not trained further. The final accuracy that the model has attained is 93.96% for training and 91.27% for validation.

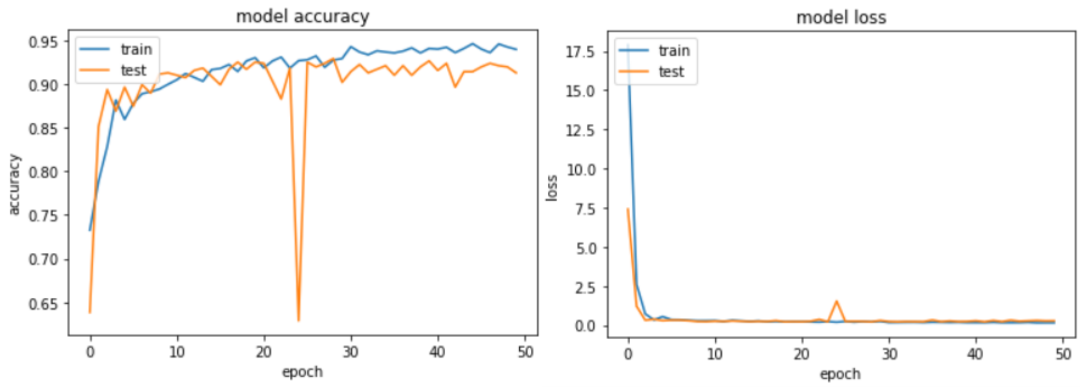

Figure 8: Model accuracy and loss for Inception v3

The model resulted in a validation accuracy of 91.27%. The trained model is then used to detect diabetic retinopathy and check its performance towards disease classification. The model is used to predict over 733 test images. Table 3 represents the performance results in the form of a confusion matrix. The model's sensitivity is calculated to be 89.69%, specificity is 93.04%, and the F1 score for the model is 0.9134.

Table 3: Confusion Matrix for Inception v3 model

	No DR (Predicted)	**DR (Predicted)**
No DR (Actual)	321 (TN)	40 (FN)
DR (Actual)	24 (FP)	348 (TP)

5.3 ResNet 50

5.3.1 Implementation

Various computer vision tasks use Residual networks as their backbone, and Residual networks are also known as ResNet. ResNet model allows the training of deep neural networks. Neural networks with over 150 layers can be successfully trained by using ResNet. The architecture of ResNet 50 consists of 50 weighted layers and four stages associated with it. Every stage has a total of 3 convolutional layers and is replicated. A shortcut connection enables skipping these blocks of convolutional layers. This feature helps ResNet in learning global features specific to data. For implementing transfer learning, the parameters for the convolutional layers are kept intact. ResNet is imported using tensorflow.keras library, the data used as an input is already in the standard size of 224*224*3. After defining the input, the model uses pre-trained weights from the ImageNet database. Hence training all layers is not required. The Last layer is replaced with a fully connected layer. For performing binary classification of retinal fundus images under two labels i.e., DR and No DR. network is then trained using Default settings like max-pooling, a learning rate of 0.0001, sigmoid activation function, and binary_crossentropy loss function. A similar architecture was used by Elswah et al. (2020) for detecting DR and determine its grade (e.g., Proliferative Diabetic Retinopathy (PDR), moderate, severe, or mild). The was trained using Indian Diabetic Retinopathy Image Dataset (IDRiD) dataset. Highest classification accuracy achieved by their proposed model is 86.67%.

5.3.2 Evaluation and Interpretation

The ResNet50 model is evaluated by running 50 epochs for training and validation data. A batch size of 32 is used to train the model as it resulted in optimal performance. The accuracy and loss learning curves for the model can be seen in figure 9 below. By observing the learning curves, there was a gradual increase in the accuracy for the train and test of the model till the first 5 epochs, and then the curve attains stability. Similarly, the Loss curve for test and train data shows a steep decrease in the loss value for train and test data. Then it reaches the stability point beyond which no improvements in the model's fit, decrease in loss, or increase in accuracy is observed. Learning curves can also be used to determine dataset properties like whether they are representative relatively. An under representative dataset is the one in which there is no sufficient training information present in the training dataset compared to the validation data used for evaluating it. As shown in figure 9, both curves show an improvement yet there is a gap between both the curves. It can be corrected by including more data in the training set by performing data augmentation. Still, since this model is developed for comparative analysis of the designed DC-CNN model, the size of the train and test datasets are kept the same.

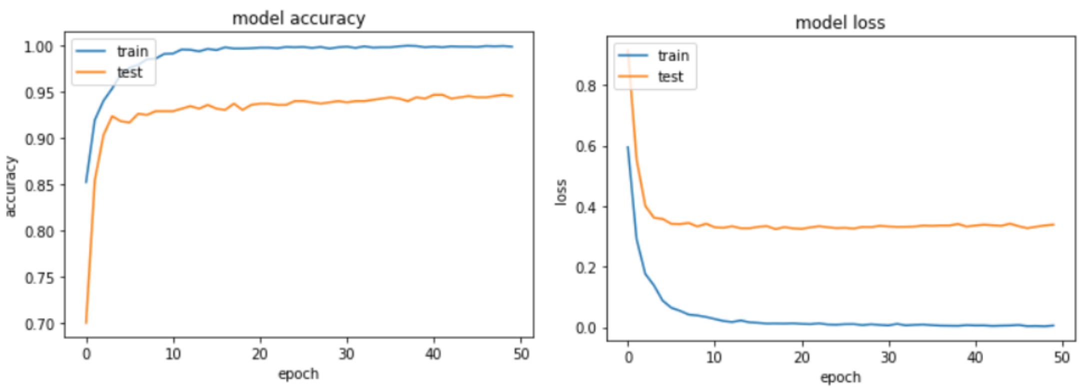

Figure 9: Model accuracy and loss for ResNet 50

The model resulted in a validation accuracy of 94.54 percent. Model is then used to detect diabetic retinopathy and check its performance towards disease classification. The model is used to predict over 733 test images. Table 4 represents the performance results in the form of a confusion matrix. The model's sensitivity is calculated to be 96.11%, specificity is 93.03%, and the F1 score for the model is 0.94.

Table 4: Confusion Matrix for ResNet 50 model

	No DR (Predicted)	**DR (Predicted)**
No DR (Actual)	347 (TN)	14 (FN)
DR (Actual)	26 (FP)	346 (TP)

6 Comparison of Results and Discussions

This section compares the three applied deep learning models, i.e., newly design dual-channel convolutional neural network (DC-CNN), ResNet, and Inception v3. The main objective of the research is to compare the performance of the newly designed DC-CNN with other state-of-the-art models that have been previously applied for detecting diabetic retinopathy (DR) using retinal fundus images. ResNet and Inception v3 architectures have been used previously for the detection of diabetic retinopathy by Elswah et al. (2020) and Li et al. (2019). They have achieved good classification accuracies, which makes them a suitable choice for performing comparative analysis. Figure 10 represents the comparative analysis of all the three applied models. All the models are compared on the basis of accuracy, sensitivity, specificity, and f1 score. As shown from Figure 10, The newly designed DC-CNN model results in the best classification performance of (DR) with an accuracy of 95.23 percent. ResNet50 achieves the second-highest accuracy of 94.54 percent, and Inception v3 achieves the lowest accuracy of 91.27 percent. The sensitivity of a model plays a very important role in developing a solution in the medical domain, the higher the sensitivity better is the model. DC-CNN model achieved the highest sensitivity when compared with other models. The DC-CNN model achieved a sensitivity of 96.95 percent, specificity of 93.58 percent, and f1 score of 95.23 percent.

In contrast, the second-highest performance is achieved by the ResNet50 model with a sensitivity of 96.11 percent, specificity of 93.03 percent, and f1 score of 94.55 percent. Lastly, the Inception v3 model achieved the lowest performance with 89.69% sensitivity, specificity of 93.04%, and f1 score of 91.34%. Overall, on observing the loss and accuracy learning curves on the train and test data, the DC-CNN model was found to fit the data very well, while the other two models just showed decent enough fits with few errors. Hence, it can be concluded from the overall results that the newly designed DC-CNN model outperforms the ResNet and Inception v3 models when trained on the same data with default settings towards predicting diabetic retinopathy from retinal fundus images.

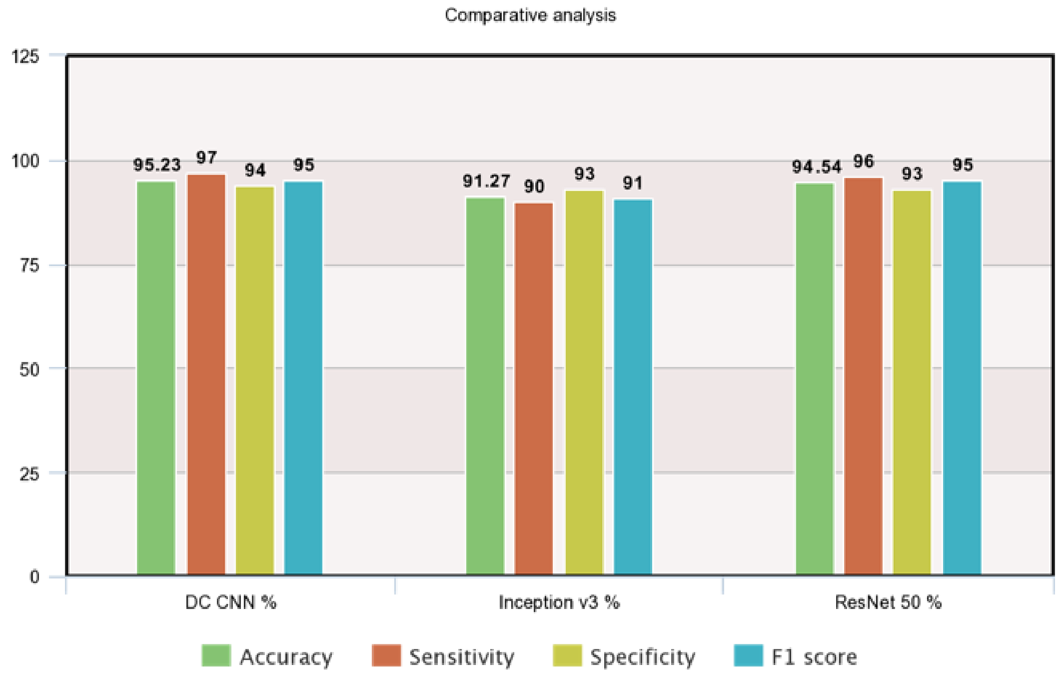

Figure 10: Accuracy, Sensitivity, Specificity, and F1 score of all the models

7 Conclusion and Future Work

Detecting diabetic retinopathy (DR) on time has been a prime concern for all the governing bodies to prevent the young working population from developing permanent blindness. Hence, this research is conducted to improve the overall efficiency of the existing DR detection system by designing a new model called dual-channel convolutional neural network (DC-CNN), which can replace existing manual clinical procedures for DR detection. The designed model aims to classify retinal fundus images in two classes showing the presence and absence of DR (DR and No DR). The designed DC-CNN model uses two channels, and one channel utilizes VGG16 architecture while the second channel is designed on tuned customized CNN architecture. The model's classification accuracy is measured in terms of sensitivity, sensitivity, accuracy, and F1 score, and it is compared with the performance of two pre-established models ResNet50 and Inception v3. All the steps for image pre-processing, transformation and data augmentation are kept standard on performing a comparative analysis of all the three models, i.e., DC-CNN, ResNet50, and Inception v3 for binary image classification of retinal fundus image and DR detection. The designed DC-CNN model outperforms the other two models to provide the highest validation accuracy of 95.23%, the sensitivity of 96.94%, specificity of 93.58%, and f1 score of 95.23%. Hence the objective of the research is achieved by measuring to what extent the designed DC-CNN can detect diabetic retinopathy using retinal fundus images in terms of sensitivity and specificity.

Detection of DR in early stages can prevent blindness in more than half of the population around the world. Due to some bottlenecks in the traditional clinical diagnosis of DR, a considerable population evades regular DR screening. In the future, the development of an automated DR detection system that can perform up to the gold standards of clinical diagnosis can be created by utilizing the vast repositories of retinal fundus images available. Due to limitations in computation power, this research performs a binary classification by designing a new DC-CNN architecture. With enough computational power, DC-CNN can be used to classify the severity of DR in retinal images, and a high-performing model can be deployed at the community level. The designed DC-CNN architecture can also be used for other similar computer vision applications.

Acknowledgement

I want to thank Prof. Noel Cosgrave for clarifying all the doubts and guiding me throughout the supervision to be on the right track. Without which it would have been tough to pursue this research.

References

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*, Vol. 1722, IOP Publishing, p. 012010.

Deperlıoğlu, Ö. and Köse, U. (2018). Diagnosis of diabetic retinopathy by using image processing and convolutional neural network, *2018 2nd International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*, IEEE, pp. 1–5.

Elswah, D. K., Elnakib, A. A. and Moustafa, H. E.-d. (2020). Automated diabetic retinopathy grading using resnet, *2020 37th National Radio Science Conference (NRSC)*, IEEE, pp. 248–254.

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

Gu, K., Xia, Z., Qiao, J. and Lin, W. (2019). Deep dual-channel neural network for image-based smoke detection, *IEEE Transactions on Multimedia* **22**(2): 311–323.

Harun, N. H., Yusof, Y., Hassan, F. and Embong, Z. (2019). Classification of fundus images for diabetic retinopathy using artificial neural network, *2019 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT)*, IEEE, pp. 498–501.

Jeong, Y., Lee, B., Han, J.-H. and Oh, J. (2020). Ocular axial length prediction based on visual interpretation of retinal fundus images via deep neural network, *IEEE Journal of Selected Topics in Quantum Electronics* **27**(4): 1–7.

Kumar, B. and Ghosh, S. (2020). Detection of concrete cracks using dual-channel deep convolutional network, *2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, IEEE, pp. 1–7.

Li, F., Liu, Z., Chen, H., Jiang, M., Zhang, X. and Wu, Z. (2019). Automatic detection of diabetic retinopathy in retinal fundus photographs based on deep learning algorithm, *Translational vision science & technology* **8**(6): 4–4.

Luo, X., Pu, Z., Xu, Y., Wong, W. K., Su, J., Dou, X., Ye, B., Hu, J. and Mou, L. (2021). Mvdrnet: Multi-view diabetic retinopathy detection by combining dcnns and attention mechanisms, *Pattern Recognition* **120**: 108104.

Oh, K., Kang, H. M., Leem, D., Lee, H., Seo, K. Y. and Yoon, S. (2021). Early detection of diabetic retinopathy based on deep learning and ultra-wide-field fundus images, *Scientific Reports* **11**(1): 1–9.

Poliyapram, V., Imamoglu, N. and Nakamura, R. (2019). Deep learning model for water/ice/land classification using large-scale medium resolution satellite images, *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*, IEEE, pp. 3884–3887.

Yang, G., Gewali, U. B., Ientilucci, E., Gartley, M. and Monteiro, S. T. (2018). Dual-channel densenet for hyperspectral image classification, *IGARSS 2018-2018 IEEE International Geoscience and Remote Sensing Symposium*, IEEE, pp. 2595–2598.

Zubair, M., Naik, M. U. K. and Mouli, G. N. C. (2020). Facile diabetic retinopathy detection using mrhe-feed and classification using deep convolutional neural network, *2020 IEEE 15th International Conference on Industrial and Information Systems (ICIIS)*, IEEE, pp. 247–252.