Deep Learning Models for Automatic Diabetic Retinopathy Diagnosis

Nidhushan Rajendran[1], Mukunthan Tharmakulasingam[2]
[1]Department of Computer Engineering, University of Jaffna, Sri Lanka
[2]Department of Electrical and Electronic Engineering, University of Jaffna, Sri Lanka
[1]2021e189@eng.jfn.ac.lk

Abstract: Many individuals are losing their vision due to diabetic retinopathy. Early detection is crucial for effective treatment, as advanced stages are challenging to cure. Diabetic retinopathy progresses through five stages: None, Mild, Moderate, Severe, and Proliferative. However, it is impractical for doctors to examine each patient's eye condition individually due to time constraints. This study explores the deep learning models for early-stage disease detection using deep learning models. To automatically recognize the stage of diabetic retinopathy, Convolutional Neural Networks (CNN), Inception-v3 and VGG16 models are trained on a large dataset containing approximately 3,662 images, and the performances are compared. This study shows that the VGG16 model outperforms the next-performing Inception model by 19% in F1-score and 21%in Accuracy, demonstrating high precision and enabling timely consultation with medical professionals upon classification.

Keywords: Diabetics Retinopathy, Disease Diagnosis, CNN, Inception-v3, VGG16

Introduction

Retinopathy is a disease that affects the retina blood vessels and is caused mainly due to diabetics. Despite the different types of retinopathy, they all involve these vascular changes. It is a common issue for diabetic patients, where the blood vessels of the retina become damaged due to high blood sugar levels. Diabetic retinopathy has five stages: none, moderate, mild, proliferative, and severe, as shown in Fig.1. In the final stages, vision loss may occur ultimately, with 2.6% of diabetic retinopathy patients experiencing vision loss[4]. Therefore, it is vital to diagnose diabetic retinopathy and its stage at an earlier stage to avoid vision loss.

Specific terminologies are used to detect diabetic retinopathy via fundus images. Lesions refer to areas where blood and fluids have leaked, appearing as spots. These lesions can be red or bright. Microaneurysms (MA) and haemorrhages (HM) are examples of red lesions[4]. Different stages of diabetic retinopathy can be identified using various lesions with the support of an ophthalmologist, as detailed in Table 1. Since diabetic retinopathy is prevalent primarily in developing countries with limited access to ophthalmologists, this paper explores the potential of diagnosing the condition through deep learning approaches.

Existing deep learning models face challenges in detecting the disease at its early stages, coupled with low performance. To tackle these challenges, three deep-learning models have been suggested for automatically detecting diabetic retinopathy from fundus images based on deep learning. Three models, CNN, VGG16 and Inception-v3, are utilized.

Figure 1: Five stages of diabetic retinopathy in fundus images: (A) Without DR, (B) Mild, (C) Moderate, (D)Severe, (E) Proliferative DR Image credit: [7]

Table 1: DR levels and corresponding lesions(reference[4])

DR Level	Lesions
No DR	No
Mild DR	MA only
Moderate DR	More than just MA, HM
Severe DR	Prominent intraretinal microvascular
Proliferative DR	neovascularization

Related Works

Khalifa et al.[1] investigated diabetic retinopathy (DR) identification using deep transfer learning models with the APTOS 2019 dataset[8]. They explored five different neural network architectures: AlexNet, ResNet18, SqueezeNet, Google Net, VGG16, and VGG19, focusing on models with fewer layers to balance performance and computational efficiency. By augmenting the dataset, they aimed to enhance model stability and reduce overfitting. Their study provides insights into the potential of deep transfer learning for DR identification and underscores the importance of robust training strategies.

Das et al.[2] introduced a CNN for classifying normal and abnormal patients using fundus images. Their method involved extracting blood arteries with a maximal principal curvature technique, followed by adaptive histogram equalization and morphological opening. Remarkably, their model achieved 98.7% accuracy and 97.2% precision on the DIARETDB1 dataset[10]. This highlights the effectiveness of their approach in accurately distinguishing between normal and abnormal patients based on fundus image analysis.

Materials and Methods

Data Description

The datasets, sourced from the Kaggle website[9], consist of 3662 training images. These images are coloured and have dimensions of 224 by 224 pixels. Within this dataset, a collection of high-quality retinal images representing the five distinct stages of Diabetic Retinopathy (DR), as shown in Figure 2, is used in this study.

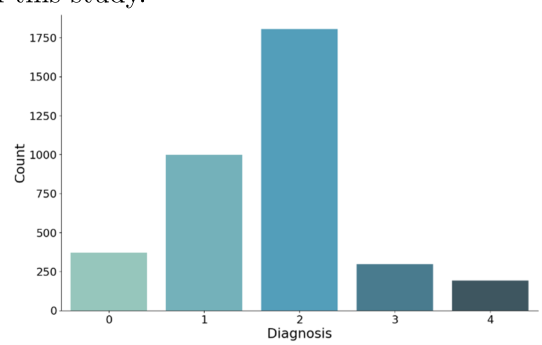

Figure 2: counts for each DR level are presented where Class index: 0-mild, 1-Moderate, 2-No-DR, 3-Proliferate-DR, 4-severe

Proposed Framework

The proposed framework has five main steps: (1) image augmentation, (2) image normalization, (3) data splitting, (4) data balancing, and (5) model training.

Step 1: Image Augmentation:

Image augmentation is a technique used to artificially increase the diversity of training data by applying random transformations to the images. In our framework, we employ image augmentation to improve the generalization of the model and reduce overfitting. We apply random horizontal and vertical flips as well as random rotations to the images using TensorFlow's data augmentation API.

Step 2: Image Normalization:

Image normalization is essential for ensuring that the pixel values of the images are within a specific range, typically [0, 1] or [-1, 1]. In our framework, we normalize the pixel values of the retinal images to the range [0, 1] using TensorFlow's Rescaling layer. This step stabilizes the training process and improves the convergence of the model.

Step 3: Data Splitting:

In machine learning, it is essential to split the dataset into training, validation, and test sets to evaluate the performance of the model. In our framework, we split the retinal image dataset into training, validation, and test sets using a custom function implemented in TensorFlow. We ensure that the distribution of classes is consistent across the splits to prevent bias in the evaluation of the model.

Step 4: Data Balancing:

Imbalanced datasets, where one class is significantly more prevalent than others, can lead to biased models. This study addresses this issue by balancing the dataset using stratified sampling during the data-splitting step. This ensures that each class is represented proportionally in the training, validation, and test sets, thereby improving the robustness of the model.

Step 5: Model training VGG16, conventional CNN and Inception V3 models are explored in this study.

VGG16 model

The VGG16 model is a widely recognized CNN architecture for image classification tasks. Trained on the ImageNet dataset, VGG16 can be repurposed to categorize images into five classes relevant to diabetic retinopathy detection. This adaptation process involves fine-tuning the model, compiling it, and then training it on the specific dataset.
The VGG16 model consists of convolutional layers designed to extract features from input images progressively. It is structured into five successive blocks. Beginning with basic features like edges and colours in the initial block, it gradually advances to more advanced features through subsequent blocks. The final block focuses on extracting the most abstract representations from the input data. By systematically processing the input images through these blocks, the VGG16 model effectively learns hierarchical representations of the input data, facilitating accurate classification for diabetic retinopathy detection.

CNN model

This model is a convolutional neural network (CNN) constructed for image classification. It has many convolutional layers followed by max-pooling layers for feature extraction, and then features are flattened and passed through fully connected layers for classification. The model is trained using the Adam optimizer with categorical cross-entropy loss.

- Convolutional Layers: The model starts with a Conv2D layer with 16 filters, each with a kernel size of 3x3 pixels. The activation function is ReLU, and they are responsible for detecting patterns in the input images.

- Pooling Layers: After the convolutional layer, there's a MaxPooling2D layer with a pool size of 2x2 pixels.

Inception V3

Inception v3 is a convolutional neural network architecture developed by Google, famous for its efficiency and Accuracy in image classification tasks. It improves performance by introducing good features such as factorized convolutions and auxiliary classifiers with its deep network structure and extensive use of batch normalization.

This model excludes the top classification layer. It then freezes all the layers of the InceptionV3 network to retain their learned features during training. The Flatten layer is added to convert the model output into a one-dimensional vector. Finally, a Dense layer with a softmax activation function produces the final classification output with five classes.

Performance Evaluation

Accuracy, precision, recall, and F1-scores are calculated using the below equations. Additionally, a confusion matrix is computed to evaluate the model's performance.

As defined in Eq (1), Accuracy measures the proportion of correctly classified samples out of the total samples. Accuracy is a fundamental metric that assesses how well a model performs across all classes.

Precision measures the proportion of correctly predicted positive cases out of all predicted positive cases as defined in Eq (2). It focuses on the Accuracy of positive predictions. When the cost of false positives is high, precision is important.

Recall, also known as sensitivity, measures the proportion of correctly predicted positive cases out of all actual positive cases as defined in Eq (3). It focuses on the ability of the model to predict positive instances. Recall is important when the cost of false negatives is high.

The F1 score is the harmonic mean of precision and recall. It provides a balance between precision and recall, giving equal weight to both metrics as defined in Eq (4). F1 score is important when classes are imbalanced.

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{1}$$

$$Precision = \frac{TP}{(TP + FP)} \tag{2}$$

$$Recall = \frac{TP}{(TP + FN)} \tag{3}$$

$$F1 - Score = \frac{(2 * Precision * Recall)}{(Precision + Recall)} \tag{4}$$

In these equations, TP =True Positives, TN = True Negatives, FP = False Positives, FN = False Negatives.

Results and Discussion

The experiments were conducted using Jupyter Notebook via Anaconda Navigator. In total, 3662 images were utilized for training, carefully categorized into three classes: training, validation, and test sets. During the training phase, a batch size of 16 images was employed to facilitate the learning process. The performance of the three distinct models varied significantly across the board, as listed in Table 2.

Table 2: Comparison of performance evaluation parameters

Model	**Accuracy**	**Recall**	**Precision**	**F1-Score**
VGG16	**0.98**	**0.96**	**0.95**	**0.95**
Inception V3	**0.77**	**0.77**	**0.84**	**0.76**
CNN	**0.77**	**0.77**	**0.72**	**0.71**

The VGG16 model demonstrated an impressive accuracy of 98%, attributed to its capability to discern patterns and features within the images. In comparison, the Inception-v3 model yielded a slightly lower accuracy of 77%. Although it performed admirably, it fell short when compared to the VGG16 model.

```
Classification Report:
              precision    recall  f1-score

           0       1.00      0.94      0.97
           1       0.97      0.98      0.97
           2       1.00      1.00      1.00
           3       0.97      0.92      0.95
           4       0.82      0.93      0.87

    accuracy                           0.98
   macro avg       0.95      0.96      0.95
weighted avg       0.98      0.98      0.98
```

Figure 3: classification report on test images

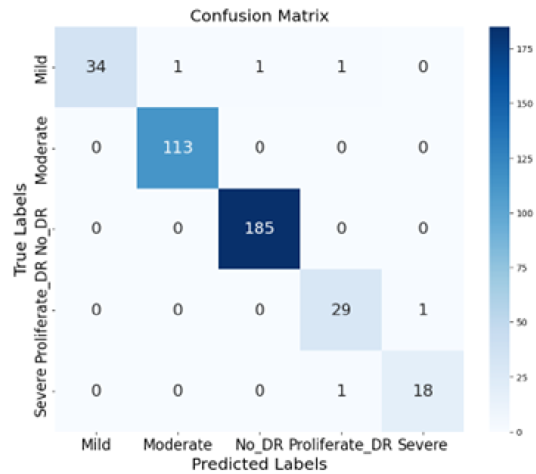

Figure 4: Confusion matrix

The CNN model exhibited the lowest Accuracy among the three, achieving a score of 76%. It struggled to attain results comparable to those of the VGG16 model. Based on these findings, it is evident that the VGG16 model stands out as the clear frontrunner regarding Accuracy and efficacy for the given image classification task. Its superior performance underscores the importance of leveraging pre-trained models with well-established architectures to achieve optimal results in similar image classification tasks. Thus, through this comparative analysis, we can confidently assert that the VGG16 model emerges as the most suitable choice. The classification report and confusion matrix for the VGG16 model are reported in Figure 3 and Figure 4.

Conclusion

This study explored the application of deep learning models for retinopathy detection in eye images. We experimented with three popular architectures: VGG16, CNN, and InceptionV3. Through this evaluation and comparison, we found that VGG16 outperformed the other two models in terms of Accuracy for detecting retinopathy. Additionally, we employed various preprocessing techniques to enhance the quality of the input data. Leveraging TensorFlow's input data pipeline facilitated the seamless integration of these preprocessing techniques, contributing to improved model performance.

Based on our findings, this study recommends adopting VGG16 for retinopathy eye detection tasks. Its superior performance suggests that it effectively captures the features relevant to this application. However, we encourage further exploration and experimentation with hyperparameter tuning and data augmentation techniques to enhance the model's performance further. Additionally, it is crucial to stay updated with advancements in deep learning research, including model architectures and techniques. This will ensure our model's continued reliability and relevance for future studies.

Although this study proposes a suitable deep learning model, these models often function as black boxes, lacking transparency and interpretability. Further research is needed to integrate explainable AI methods to provide understandable results, which is essential in the medical field.

References

[1] Khalifa, N.E.M.; Loey, M.; Taha, M.H.N.; Mohamed, HNET Deep transfer learning models for medical diabetic retinopathy detection. Acta Inform. Med. (2019).

[2] Das, S.; Kharbanda, K.; Suchetha, M.; Raman, R.; Dhas, E. Deep learning architecture based on segmented fundus image features for classification of diabetic retinopathy. Biomed. Signal Process. Control (2021).

[3] Ali, G., Dastgir, A., Iqbal, M. W., Anwar, M., & Faheem, M. A Hybrid Convolutional Neural Network Model for Automatic Diabetic Retinopathy Classification from Fundus Images. (2023).

[4] Alyoubi, W. L., Abulkhair, M. F., & Shalash, W. M. Diabetic Retinopathy Fundus Image Classification and Lesions Localization System Using Deep Learning. (2021).

[5] Alwakid, G., Gouda, W., & Humayun, M. Deep Learning-Based Prediction of Diabetic Retinopathy Using CLAHE and ESRGAN for Enhancement. (2023).

[6] Nazih, W., Aseeri, A. O., Atallah, O. Y., & El-Sappagh, S. Vision Transformer Model for Predicting the Severity of Diabetic Retinopathy in Fundus Photography-Based Retina Images. (2023).

[7] Tomi Kauppi, Valentina Kalesnykiene, Joni-Kristian Kamarainen. DIARETDB1. © Tomi Kauppi, Valentina Kalesnykiene, Joni-Kristian Kamarainen.

[8] Abraao,A.,Oliveira,R.,Rodrigues,F.,& Aires, K.(2019). APTOS 2019 Blindness Detection. Kaggle.

[9] Rath,S.R.(2020). Diabetic Retinopathy 224x224 Gaussian Filtered Dataset [Data set]. Kaggle.

[10] Kauppi, T., Kalesnykiene, V., & Kamarainen, J.-K. (Year). DIARETDB1 [Data set]. Papers With Code. https://paperswithcode.com/dataset/diaretdb1