Diabetic Retinopathy Detection via Learning Models

Sokolo D. Matsolo[1]

[1] University of Johannesburg, Auckland Park, Johannesburg, South Africa
[2] 222206953@student.uj.ac.za

Abstract. Diabetic Retinopathy (DR), a widespread condition associated with diabetes, remains a leading cause of vision loss worldwide. Early detection and timely intervention are crucial for preventing vision-related complications. The rising global incidence of diabetes highlights the importance of effective DR detection, as unmanaged diabetes can lead to its development. This study delves into the detection of DR using diverse machine learning models, with a specific focus on decision trees, Convolutional Neural Networks (CNNs), and Random Forests. To evaluate the performance of these models, a dataset from Kaggle made up of retinal images representing normal, moderate, and proliferative DR is employed. Three machine learning models are assessed: Decision Tree, Random Forest, and CNN. Our findings reveal that the CNN model outperforms the other models, achieving the highest testing accuracy while also exhibiting superior precision, recall, and F1 score. To mitigate potential overfitting issues, data augmentation techniques are employed, enhancing the dataset's robustness. Additionally, this study underscores the limitations of traditional machine learning models in capturing intricate image features and highlights the pivotal role deep learning, especially CNNs, plays a role in the detection of diabetic retinopathy. It emphasizes the potential for further enhancements in deep learning models to yield even more promising results. In conclusion, this research underscores the critical importance of early DR detection and emphasizes the considerable advantages of deep learning models, specifically CNNs, in this domain. It encourages further research to harness the full potential of deep learning for improved diabetic retinopathy diagnosis.

Keywords: Diabetic Retinopathy, DR detection, Convolutional Neural Networks, Random Forest, machine learning, deep learning.

1 Introduction

Diabetes develops from the body's insufficient insulin control. It comes in two primary types: type I and type II. The cause of type I diabetes is difficult to pinpoint and unavoidable, while type II diabetes can be detected through retinal images, showing indicators like hemorrhages (HEMs), exudates (EXs), cotton wool spots (CWs), and more [5]. Understanding these differences helps with better diagnosis and management of diabetes-related conditions. Untreated type

II diabetes for over five years can lead to Diabetic Retinopathy (DR), a condition damaging the retina's blood vessels falling under ophthalmology. This condition worsens if blood sugar levels remain uncontrolled. In 2011, an estimated 366 million people had diabetes globally, with over half of them untreated. By 2030, this number is projected to grow to 552 million, meaning over 10% of adults worldwide will have diabetes, many unaware of their condition [7]. As the global diabetes population grows, diabetic retinopathy is also on the rise. About one-third of individuals with diabetes are affected to varying degrees [8], leading to significant vision loss and blindness among working-age adults.

To avoid vision loss, early diagnosis and treatment plans for Diabetic Retinopathy (DR) are essential. The lack of early indications for DR, including Diabetic Macular Edema (DME), is becoming increasingly common, making timely detection of DR highly desirable. Non-Proliferative Diabetic Retinopathy, in its various stages of development, is characterized by four symptoms: microaneurysms (MA), hard exudates (HE), CWs, and HEMs [5]. Microaneurysms may appear as tiny, scarlet spots. Computer-aided diagnostic (CAD) systems have recently been developed as a result of quick improvements in hardware and software technology. Clinical data, image-based data, and genetic data are the three categories of data utilized for diagnosis; however, using genetic data in CAD systems is currently not practical [2]. The limitation of these existing solutions is that they take a lot of time and heavily rely on the knowledge of skilled practitioners. For example, a skilled clinician must manually assess digital color fundus pictures of the retina as part of the rent solution, and DR is found by looking for lesions linked to diabetic retinopathy-related ocular anomalies [9].

The significance of diabetic retinopathy (DR) detection is increasingly acknowledged in a growing body of literature. This study's objective is to investigate various methods introduced for this purpose, with a particular focus on Convolutional Neural Networks (CNNs) and Random Forests. CNNs, being part of the deep learning domain, exhibit remarkable capabilities in processing medical images. Recent advancements such as corrected linear units and dropout techniques have significantly improved their performance when compared to early network architectures dating back to the 1970s. In contrast, Random Forests represent an ensemble classification approach that utilizes decision trees on randomly selected data subsets, demonstrating impressive accuracy in classifying high-dimensional data domains. The construction of random forests typically involves selecting a random set of features at each node to form decision tree branches, followed by combining individual trees through a bagging method to create the random forests model [10].

However, while neural networks benefit from features like rectified linear activations and increased GPU computational capabilities for complex image recognition tasks [6], CNNs often face criticism for their lack of interpretability, often labeled as "black box" models. To address this issue, generating activation maps can enhance the transparency of the decision-making process in CNNs by helping determine whether a neuron requires stimulation. This approach assists in validating the suitability of the received data for the input, ultimately fostering trust in the model's predictions [9]. This self-explanatory aspect can serve as motivation for practitioners to investigate the underlying causes of the disease in individual patients.

2 Background

By integrating feature learning and conventional learning, Chowdhury et al. proposed the usage of the Random Forest (RF) approach, while Pratt et al. proposed the Convolutional Neural Network (CNN) approach to DR detection, and have shown to outperform naive methods like decision trees. These algorithms can automatically learn features from raw photos and anticipate patterns. The principal limitation of the experimental approach is that, if a small subspace from high-dimensional data during the forest construction process is randomly selected, informative characteristics have a high possibility of being missed. As a result, when a substantial number of "weak" trees are formed in a random forest, the forest stands a high risk of making an incorrect deduction, which is mostly due to the classification power of such "weak" trees [2].

In CNNs, things become better. The deep architecture is made up of numerous levels, each of which performs a non-linear change on the outputs of the one before it. The network can extract complicated patterns from low-level to high-level features because of this hierarchical representation. As a result, the network is capable of making complex judgments. Feature extraction and prediction methods are incorporated into a single model, boosting the features' discriminative potential. The CNN model is trained using output labels, which allows the extracted features to be tailored to the specific task without the need for operator intervention [6].

LayerCAM, as introduced by [3], is a technique for generating class activation maps (CAMs) that pinpoint essential areas within an input image for predicting a specific target class. It accomplishes this by determining gradients of class scores for each spatial position in a feature map. Locations with positive gradients are deemed influential for the target class prediction, while those with negative gradients are considered irrelevant. These positive gradient locations are assigned weights based on their gradients, while negative gradient locations are assigned zero weights. Subsequently, the activation values in the feature map are multiplied by their corresponding weights, ensuring that important regions receive more prominence, while unimportant regions are suppressed. The final CAM is produced by combining these weighted activations along the channel dimension. An activation function is applied to ensure that only positive values contribute to the CAM, ultimately highlighting the image regions critical for predicting the target class [3].

3 Method

3.1 Dataset

The Kaggle DR detection challenge dataset [1] includes colour fundus photos to represent normal, moderate, and proliferative DR, respectively. The dataset provides access to a collection of high-resolution retinal images captured under diverse imaging circumstances. As shown in Fig. 1 below, each subject has both left and right eye images available, identified by a subject ID and the eye side (left or right).

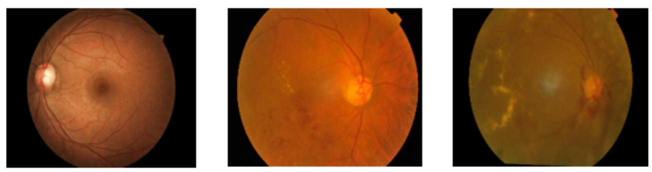

Fig. 1. Examples of retinal images from the dataset.

3.2 Experimental Setup

Decision Tree

Decision Tree is a supervised learning algorithm used in tasks that involve both classification and regression. Its structure is hierarchical, resembling a tree and consisting of key components such as a root node, branches, internal nodes, and leaf nodes. These nodes, guided by the available features, conduct evaluations to form homogenous subsets, which are denoted as leaf nodes or terminal nodes. These terminal nodes encompass all potential outcomes within the dataset.

In order to prepare the data for training and testing, we initially loaded images from the "Diabetic Retinopathy in Eye Images Dataset" and standardized them to a uniform size of 200×200 pixels. We then converted each image into a flattened array, creating a feature vector that was employed as input for the Decision Tree model. The dataset was then divided into training and testing subsets, with 30% of the data being set aside for testing to assess the model's capacity to generalize. The next step involved training the Decision Tree classifier using the training data and utilizing it to make predictions on the test set. Finally, we calculated the model's accuracy, recall, precision, and F1-Score as part of our evaluation process.

Random Forest

This ensemble method is used for both classification and regression tasks. It constructs a multitude of decision trees during the training phase and combines their outputs to make predictions. This method is utilized in diabetic retinopathy detection.

In the experiments conducted, 100 decision trees were generated and the dataset was divided according to established procedures. Also, 45 images were reserved for testing, ensuring that the training and testing sets were disjoint. The training set was further split into training and validation sets using an 8 to 2 ratio for all trees. The final performance score was calculated as the mean recognition rate per class. We computed and reported the accuracy, recall, precision, and F1-Score of the model, offering a comprehensive assessment of its performance and ability to generalize when confronted with unfamiliar data.

Convolutional Neural Network

The core of the network comprises three convolutional layers, each serving as a feature extractor. These layers apply filters to the input data, utilizing the Rectified Linear Unit (ReLU) activation function to introduce non-linearity. To maintain spatial dimensions, "same" padding is used in each convolutional layer. Additionally, L2 regularization is applied to the weights of these layers. After each convolutional operation, a max-pooling layer is employed to downsample the feature maps. Following the convolutional layers, there is a flattened layer, which reshapes the 2D feature maps into a 1D vector. This flattened representation of features is then fed into fully connected layers. The first Dense layer consists of 256 neurons and employs ReLU activation as well as L2 regularization to perform further feature extraction and non-linear transformations. The dropout layer is added after this first dense layer. Finally, the output layer is a Dense layer with a softmax activation function, which is well-suited for multi-class classification tasks, as it assigns probabilities to each class.

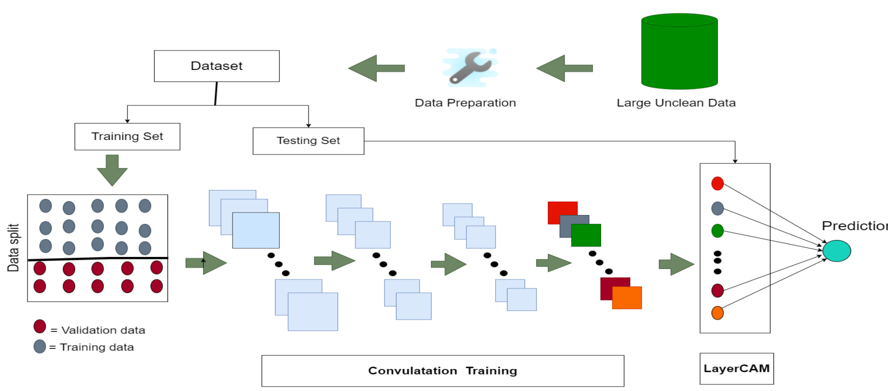

Fig. 2. A Depiction of a neural network is set up.

The original 240×240 pixel color images underwent a series of transformations, including reflection around the X-axis, Y-axis, and X-Y axis. These transformations generated augmented images, which were subsequently used for training with a batch size of 32. To expedite the training process, batch normalization was incorporated, and as a preventive measure against overfitting, dropout was implemented. The choice of the Adam optimizer was made to optimize the network's performance, configured with a learning rate of 0.001 to avoid overshooting, for a total of 35 training epochs. Furthermore, the Categorical Cross-Entropy loss function was adopted for the task.

3.3 Analysis

In this study, we aimed to develop a diabetic retinopathy detection model. We used Decision Trees as our naive model, then Random Forest as our baseline method, and Convolutional Neural Networks (CNN) as our primary approach. For the Random Forest model, we preprocessed retinal images, extracted features, and trained a classifier due to its proficiency with high-dimensional data and complex feature interactions. Our CNN model employed deep learning with convolutional and fully connected layers, utilizing transfer learning on a pre-trained model to leverage its image-related task capabilities and automatic feature learning.

To gauge the effectiveness of our models, we employed a set of key metrics. Firstly, we utilized accuracy as a means to evaluate the overall predictive accuracy of our models, providing us with a comprehensive view of their performance. Additionally, precision served as a valuable metric to assess the models' competence in correctly identifying instances of churn, ensuring that our predictions were accurate and reliable. Moreover, recall played a crucial role in evaluating the models' ability to detect all actual churn cases, thus helping us avoid overlooking potentially critical scenarios. Finally, the F1 score emerged as a vital metric, striking a balance between precision and recall, and considering the critical trade-off between false positives and false negatives. Together, these metrics offered a comprehensive assessment of our models' performance in prediction.

3.4 Tools and Libraries

In this research paper, we conducted the implementation of all the algorithms using Python 3 (version 3.11.2) and leveraged the capabilities of the OpenCV and Scikit-learn libraries within Jupyter notebooks. To create the Convolutional Neural Network (CNN) for detecting diabetic retinopathy, we employed TensorFlow Keras. The Jupyter notebooks were executed on virtual machines provided by Google Colaboratory, which had standard runtime configurations featuring Intel Xeon single-core processors running at 2.3 GHz and approximately 12.7 GB of available RAM. The training of the CNN model was carried out on the GPU runtime, using a single Tesla K80 GPU.

4 Results

Following the general approach, 7,526 images were split into training and validation sets for evaluation purposes. All three models were implemented and tested using testing set,then results were recorded.

4.1 Testing accuracy metric

Decision Tree and Random Forest

The decision tree model displayed an overall accuracy of approximately 55.55%, which emphasizes its inability to correctly classify the majority of instances within the dataset. Notably, the model exhibited modest performance when categorizing specific classes, notably "No DR" and "Moderate - Eye with minor lesions." However, the decision tree model's performance metrics for these classes, including precision, recall, and F1 scores, were relatively low, signifying its challenges in accurately identifying and differentiating instances. The model's performance further worsened when assessed on the "Proliferate" classes.

Table 1. Classes and testing accuracy for decision tree and random forest models

Accuracy/Model	Decision Tree	Random Forest
No DR	0.6	0.78
Moderate	0.44	0.56
Proliferate	0.53	0.50
Total Accuracy	0.55	0.62

Convulational Neural Network

In Fig. 3, a Convolutional Neural Network (CNN) was employed to detect diabetic retinopathy. The model underwent 35 training epochs, during which both training and validation accuracies and losses were monitored. At the outset, the training and validation losses stood at 6.8859 and 0.9461, while the training and validation accuracies were 0.724 and 0.819, respectively.

As the training progressed, the training loss increased to 0.8209, and the validation loss decreased to 0.7745 by the end of epoch 3. Subsequently, from epoch 3 to 35, both the training and validation losses exhibited consistent improvements, ultimately reaching 0.435 for the training loss and 0.445 for the validation loss.

Regarding accuracies, despite initial setbacks, exemplified by a validation accuracy spike at epoch 3, the model achieved a commendable validation accuracy of 85.5% along with a training accuracy of 83.6%. These results highlight the model's ability to overcome obstacles and excel in the detection of diabetic retinopathy.

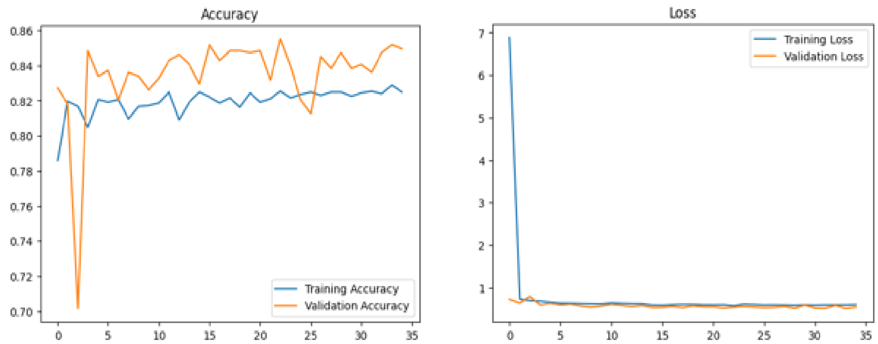

Fig. 3. The CNN accuracy and loss curves

4.2 Performance evaluation

Table 2 facilitates a straightforward comparison of performance metrics across the naive approach, the baseline algorithm, and the proposed CNN.

Table 2. Performance metrics for the models

Model	Decision Tree	Random Forest	CNN
Precision	0.6000	0.6217	0.8673
Recall	0.5609	0.6306	0.7389
F1 Score	0.5846	0.6200	0.7943

The table reveals that the CNN model attained the highest precision, recall percentages, and recall metrics, while the random forest was second to CNN, and the decision tree was the least effective classifier. Thus, the SOTA method underscores the significant enhancement achieved with the deep learning approach in diabetic retinopathy detection and classification.

5 Discussion

To identify the difference between deep learning models and other learning techniques when it comes to the detection of diabetic retinopathy, we conducted a comparative analysis of their performance. The results showed that there was a big difference in how well these methods worked. The proposed CNN model emerged as the top performer, achieving an impressive validation accuracy of 85.5%. In contrast, the Random Forest model lagged behind with a validation accuracy of 65%. The decision tree model, on the other hand, showed the lowest overall accuracy of 57%. These results underscore the significant advantage of deep learning, particularly the CNN model, in the field of diabetic retinopathy detection.

These results corroborate the findings of a great deal of previous research, particularly the work conducted by [6], which served as a precursor to our own investigation, reporting that the final trained CNN model achieved an impressive 95% specificity, 75% accuracy, and 30% sensitivity. It was further emphasized that the network exhibited proficiency in identifying images of healthy eyes. This success can be attributed to the substantial presence of healthy eye data in the dataset, a characteristic that aligns with the observations made in our study. However, it is clear that Pretta et al. had a more robust model than ours. These differences can be attributed, in part, to the possibility that their CNN model benefited from more effective architectural choices in terms of layer configurations and optimization functions, enabling it to extract more intricate hierarchical information from images. Alternatively, the relatively limited training duration of 35 epochs in our case might have led to overfitting.

Lastly, the decision tree and random forest models achieved accuracy rates of 63% and 55%, respectively. Their limited performance can be attributed to their incapacity to capture local image features, which simplifies the assumption of feature independence. However, in the context of diabetic retinopathy detection, these models displayed strong performance in specific categories but exhibited a bias due to the inherent limitations of self-reported data. It's essential to emphasize that both the random forest and decision tree models do not belong to the realm of deep learning models, thereby reinforcing the notion that they are less adept at discerning intricate hierarchical patterns within eye images when compared to state-of-the-art models like the proposed CNN.

6 Conclusion

Diabetes-related vision complications encompass a condition known as diabetic retinopathy (DR). While it may initially present without symptoms or with only minor vision disturbances, this ailment carries the potential to lead to blindness over time. It is alarming to note that over half of the estimated 366 million cases of diabetes worldwide in 2011 remained undiagnosed. This number is projected to surge to 552 million by 2030, signifying that more than 10% of adults globally will be afflicted by diabetes, many of whom are unaware of their condition.

In this study, we used the Kaggle dataset to examine how well deep learning models, random forests, and decision trees perform when it comes to identifying diabetic retinopathy (DR) in a medical setting. Augmentation techniques were used on the dataset to solve the overfitting problem, which resulted in a greater number of photos in the dataset than it had initially. In this investigation, deep learning models were chosen. The CNN model performed better than the others, obtaining the greatest testing accuracy of 85.5% as well as the highest percentages for precision and F1 score. The results also included testing accuracy overall and performance measures, such as precision, recall, and F1 score. Surprisingly, this model achieved its results with a small number of layers, which lowered the computational complexity and training time.

References

1. Diabetic retinopathy detection competition dataset, 2015. Accessed on: July 27, 2023.
2. Amrita Chowdhury, Tamojit Chatterjee, and Sreeparna Banerjee. A random forest classifier-based approach in the detection of abnormalities in the retina. *Medical Biological Engineering Computing*, 57, 08 2018.
3. Peng-Tao Jiang, Chang-Bin Zhang, Qibin Hou, Ming-Ming Cheng, and Yunchao Wei. Layercam: Exploring hierarchical class activation maps for localization. *IEEE Transactions on Image Processing*, 30:5875–5888, 2021.
4. T. Manaka, T. Van Zyl, A. N. Wade, and D. Kar. Using machine learning to fuse verbal autopsy narratives and binary features in the analysis of deaths from hyperglycaemia. *Proceedings of SACAIR2021*, 1(1):90–106, 2021.
5. Dimple Nagpal, S.N. Panda, Muthukumaran Malarvel, Priyadarshini A Pattanaik, and Mohammad Zubair Khan. A review of diabetic retinopathy: Datasets, approaches, evaluation metrics and future trends. *Journal of King Saud University - Computer and Information Sciences*, 34(9):7138–7152, 2022.
6. Harry Pratt, Frans Coenen, Deborah M. Broadbent, Simon P. Harding, and Yalin Zheng. Convolutional neural networks for diabetic retinopathy. *Procedia Computer Science*, 90:200–205, 2016. 20th Conference on Medical Image Understanding and Analysis (MIUA 2016).
7. L. M. Ruta, D. J. Magliano, R. LeMesurier, H. R. Taylor, P. Z. Zimmet, and J. E. Shaw. Prevalence of diabetic retinopathy in type 2 diabetes in developing and developed countries. *Diabetic Medicine*, 30(4):387–398, 2013.
8. Parth Sompura and Shaily Surti. Understanding different causes of blindness with focus on diabetic retinopathy patients: a cross-sectional study. *International Journal Of Community Medicine And Public Health*, 9:867, 01 2022.
9. Zhiguang Wang and Jianbo Yang. Diabetic retinopathy detection via deep convolutional networks for discriminative localization and visual explanation. *CoRR*, abs/1703.10757, 2017.
10. Baoxun Xu, Yunming Ye, and Lei Nie. An improved random forest classifier for image classification. In *2012 IEEE International Conference on Information and Automation*, pages 795–800, 2012.