Early Detection of Diabetic Retinopathy with Segmentation Model U-Net

ADEYINKA MAYOWA-MAJARO
Department of Artificial Intelligence and Data Science, University of Hull

Abstract - Diabetic Retinopathy (DR) is a serious diabetes complication leading to irreversible vision loss, thus emphasizing the need for early detection and intervention. Leveraging advancements in deep learning, this study investigates the application of a segmentation model based on the U-Net architecture for the early detection of DR. The U-Net model, renowned for its efficacy in semantic segmentation tasks, offers a promising approach to accurately delineating retinal structures indicative of DR-related lesions. Through comprehensive experimentation and evaluation, including the comparison of a standard U-Net segmentation model and U-Net with a VGG16 pre-trained encoder, termed U-NetVGG16, the models' proficiency in achieving high accuracy, precision, recall, Intersection over Union (IoU), and F1-score metrics was demonstrated. U-NetVGG16 excelled, earning a notable IoU of 98.62%, an accuracy of 99.10%, a precision of 99.30%, a recall of 99.40%, and an f1-score of 99.30%. The results highlight the potential of deep learning-based segmentation models in revolutionizing diabetic eye care by facilitating automated and precise identification of DR-related abnormalities. This study contributes to advancing the field of early DR detection, aiming to mitigate the global burden of preventable vision impairment associated with this debilitating condition.

I. INTRODUCTION

Diabetic Retinopathy (DR) is a critical complication of diabetes, with the potential for irreversible vision loss if not detected and treated early. It arises as a complication of diabetes mellitus, affecting the retina due to prolonged high blood sugar levels. DR progresses through four stages: mild Non-Proliferative DR (NPDR) with micro-aneurysms; moderate NPDR with distorted blood vessels; severe NPDR with blocked vessels leading to new vessel growth; and proliferative DR (PDR) with fragile vessels prone to leakage. (Dutta et al., 2021). The World Health Organization (WHO) estimates that over 400 million people worldwide are affected by diabetes, with about one-third at risk of developing DR (WHO, 2020). Early detection of DR gains paramount significance in preventing considerable fatalities. Traditionally, DR's detection relied on manual methods such as stereoscopic colour fundus photography and optical coherence tomography, which required skilled readers and were labour-intensive (Goh et al. 2016). In recent years, artificial intelligence (AI) and deep learning (DL) techniques have revolutionised medical imaging analysis, offering promising avenues for early detection and diagnosis of various diseases, including DR (Athanasios, 2023). The application of segmentation models such as U-Net for the early diagnosis of DR is one such development in this field. The U-Net architecture, initially proposed for biomedical image segmentation, has garnered significant attention for its effectiveness in accurately delineating structures within medical images (Xiao-Xia et al., 2022). By leveraging convolutional neural networks (CNNs) and an encoder-decoder architecture, U-Net demonstrates proficiency in segmenting retinal structures and identifying abnormalities indicative of DR. This paper aims to provide an in-depth exploration of the application of the U-Net segmentation model in the detection of DR irrespective of the severity level. Through a comprehensive review of existing literature, methodologies, and advancements, this paper seeks to elucidate the potential of U-Net as a tool for automated and precise segmentation of retinal images, thereby facilitating the early identification of DR-related lesions.

II. USE OF MACHINE LEARNING METHODOLOGIES FOR THE DETECTION OF DIABETIC RETINOPATHY

Several studies have employed ML and DL methods to detect DR in various stages. Anas et al. (2022) employed a U-Net architecture for DR detection, achieving 93.52% accuracy on DIARETBDO datasets. Similarly, Revathy et al. (2020) explored DR detection using ML methods. The authors employed a feature extraction approach and explored a spectrum of algorithms, including k-nearest neighbours (KNN), Random Forests (RF), and Support Vector Machines (SVM). Their findings emphasized the potential of ML in automating DR detection, achieving a notable accuracy rate of 82% in classifying various DR stages. Parthasharathi et al. (2022) utilised DL architectures effectively. Their study's strength lies in the use of DL architectures for automated diagnosis by creating a CNN model using the VGG19 framework to detect DR and provide information on the severity level of the disease, the study showcased the effectiveness of DL models in achieving 84.41% accuracy. Gunasekaran et al. (2022) focused on attributed relevancy and categorization strategies for effectively categorising diseases related to the retina. They utilised a Deep Recurrent Neural Network (DRNN) system, achieving a 14.69% boost in efficiency compared to traditional approaches. Rakhlin (2018) integrated deep CNNs to diagnose eye fundus images, achieving 99% sensitivity, 71% specificity, and 97% Area under Curve (AUC). Mujeeb et al. (2022) applied ML algorithms such as SVM and Deep Neural Network (DNN), achieving AUCs of 97.11% and 99.15%, respectively. The study detected the existence of DR in retinal images using segmentation techniques. Nathan (2022) predicted DR using various models, including Linear Support Vector Classification (LSVM) and Logistics Regression (LR) achieving 80% accuracy. Thippa et al. (2020) used a hybrid Principal Component Analysis (PCA)-firefly-based DNN model for DR classification. The result of the model was evaluated against the traditional state of the model to establish its superiority in terms of accuracy. Mohamed et al. (2021) proposed a Hybrid Inductive ML Algorithm (HIMLA). The proposed method was evaluated on CHASE datasets to detect DR. The accuracy, sensitivity, and specificity of the approach are 96.62%, 95.31%, and 96.88%, respectively. Penikalapati and Agastyaraju (2022) integrated SVM, PCA, and moth-flame optimisation techniques, achieving 91.5% accuracy. The existing literature strongly suggests that the application of artificial intelligence (AI) and deep learning (DL) techniques holds considerable potential for advancing medical imaging analysis.

III. MATERIALS AND METHODS

- Dataset Collection

The retinal images used in this study were collected from a public data repository; Kaggle (Diabetic_Retinopathy_Balanced | Kaggle). The dataset consists of 49,700 grey-scaled fundus images (512x512 resolution) that are grouped into five categories based on disease severity (Figure 1). Natural features (blood vessels, macula, optic disc) and abnormal features (microaneurysms, exudates, haemorrhages) can be seen in the images. The five classes of DR severity are represented equally in the dataset with each class containing approximately 10,000 images. Segmentation models require binary masks as an input, the Otsu image segmentation algorithm together with adaptive methods was used to generate binary masks from the retinal images; both the retina images and generated binary masks were used for the segmentation task.

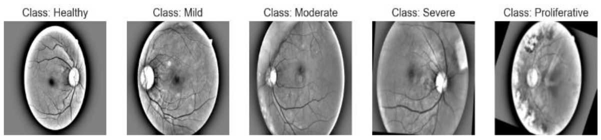

Figure 1: Representative examples of retinal images presented in order of increasing severity from left to right.

- Data Preprocessing

The retinal images were presented in jpeg format, allowing for easy usage, the images were resized to 128x128 resolution to make the images fit into the architecture of U-Net, increase computational efficiency and decrease processing time. The normalization layer was used for image data preprocessing. It scales the pixel values of the input images to the range [0, 1] by dividing each pixel value by the maximum intensity value of 255 to maintain stability and improve convergence speed.

- Model Training

The model training was formulated to optimize the U-Net model parameters and enhance its segmentation performance on the retinal images. The training dataset comprises pairs of input images and corresponding ground truth binary masks, representing the desired segmentation labels and were trained on 80% of the dataset and validated on 20%. During training, the Adam optimizer was employed with a learning rate of 1e-3 to minimize the binary cross-entropy loss between the predicted segmentation masks and the ground truth labels. The training process was conducted iteratively over multiple epochs, with each epoch processing a batch of 32 image-mask pairs. This mini-batch stochastic gradient descent approach accelerates convergence and facilitates efficient utilization of computational resources. To monitor training progress and prevent overfitting, callbacks like model checkpointing, Learning Rate Reduction on Plateau and early stopping were utilized; Model Checkpointing saves the best-performing model weights periodically based on the chosen validation metric (IoU), ensuring that the model with the highest validation performance is retained; The learning rate is dynamically adjusted by a factor of 0.1 if no improvement in the IoU is observed over a certain number of epochs (patience parameter); and finally the early stopping halts training if no improvement in the validation metric is observed within a patience callback of 2, preventing unnecessary computation and mitigating overfitting. All experiments were programmed and implemented using Python 3.11.7 with Keras and Tensorflow (Tensorflow,2023). Anaconda Jupiter notebook was used as the software environment of the experiment and models were run on an Intel(R) Core(TM) i5-1145G7 @ 2.60GHz, 2611 Mhz, 4 Core(s), and 8 Logical Processor(s) with 32GB RAM.

- Evaluation Metrics

Models were evaluated on both the training and validation datasets at the end of each epoch to monitor performance metrics such as accuracy, IoU score, precision, recall and F1-score given from equations 1-5 below. These metrics provide insights into the model's segmentation accuracy and generalization capability.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \text{ x } 100\% \tag{1}$$

$$\text{Precision} = \frac{TP}{TP+FP} \text{ x } 100\% \tag{2}$$

$$\text{Recall} = \frac{TP}{TP+FN} \text{ x } 100\% \tag{3}$$

$$\text{IoU} = \frac{GT \cap S}{GT \cup S} = \frac{TP}{TP+FPX\ FN} \tag{4}$$

$$F_1\text{Score} = 2 \text{ x } \frac{Recall\ x\ Precision}{Recall + Precision} \text{ x } 100\% \tag{5}$$

IV. U-NET ARCHITECTURE CONFIGURATION

U-Net is a distinct alternative to the standard CNN for disease detection and abnormality localisation in biomedical image segmentation (Baccouch et al. 2023). The U-Net architecture consists of a contracting path followed by an expansive path. Convolutional layers with dropout extract hierarchical features in the contracting path, and max-pooling layers reduce spatial dimensions. The expansive path involves transposed convolutional layers that up-sample the features and concatenate them with corresponding features from the contracting path. The network progressively refines features with convolutional layers and dropout. The final layer produces binary segmentation masks.

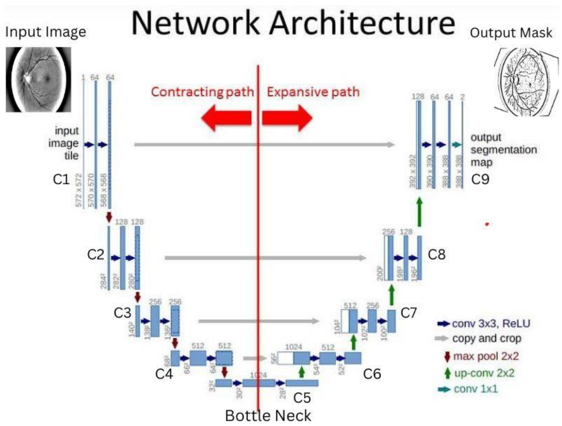

Figure 2: Modified U-Net architecture is depicted, showcasing the contracting path (c1-c4) on the left and the expansive path (c6-c9) on the left. The dense layer is denoted as c5.

- Integration Of Vgg16 as Encoder

The contracting path was modified by incorporating pre-trained models VGG16. VGG16 is a convolutional neural network architecture known for its simplicity and effectiveness in image classification tasks. It consists of multiple convolutional layers followed by max-pooling layers, culminating in fully connected layers for classification (Awf & Muhammet 2022). In this implementation, the encoder part of the U-Net architecture was replaced with the pre-trained VGG16 model. Instead of designing and training a custom encoder from scratch, the weights and architecture of VGG16, which have been learned from extensive training on large datasets were utilized to extract meaningful features from the input images. This integration allows leveraging the rich feature representations learned by VGG16 for image classification tasks. The input image passes through the VGG16 layers during the forward pass, gradually transforming it into a set of high-level feature maps. These feature maps capture hierarchical representations of the input image, ranging from simple edges and textures to complex object features.

- Connection with Decoder

The output of the last convolutional layer in the VGG16 encoder is typically connected to the corresponding layer in the decoder part of the U-Net architecture through skip connections. These skip connections facilitate the fusion of low-level and high-level features, enabling precise localisation of objects in the segmented output. Figure 2 illustrates the adapted U-Net architecture.

V. RESULTS AND DISCUSSION

A standard U-Net model with four convolutional blocks in the contacting path, the bottleneck and 4 convolution decoder paths (Figure 2) was initially used for DR detection before altering the encoder part with transfer learning models VGG16 as the backbone.

- Model Performance Comparison

Table 1 shows the experimental results of the segmentation models. The segmentation model based on the standard U-Net architecture achieves impressive performance metrics across multiple evaluation criteria. With an accuracy of 99.03%, the model demonstrates a high degree of pixel-wise classification accuracy. Additionally, the precision (99.50%) and recall (99.45%) metrics indicate the model's ability to correctly identify positive samples while minimizing false positives and false negatives. The IoU score stands at 98.54%. The F1-score is calculated at 99.26%, this metric provides a balanced assessment of the model's performance in terms of both precision and recall. On the other hand, the U-Net model augmented with a VGG16-based encoder denoted as U-NetVGG16, exhibits slightly improved performance metrics across the board. Notably, the model achieves a slightly higher accuracy of 99.10%, precision of 99.63%, recall of 99.40%, IoU of 98.62%, and F1-score of 99.30% compared to the standard U-Net model.

Table 1: Comparative performance of U-Net and U-NetVGG16 model

Model	Accuracy (%)	Precision (%)	Recall (%)	IoU (%)	F1-Score (%)
U-Net	99.03	99.50	99.45	98.54	99.26
U-NetVGG16	99.10	99.63	99.40	98.62	99.30

- Discussion

In evaluating the performance of the segmentation models for DR detection, the results indicate their outstanding capabilities in accurately identifying and segmenting regions indicative of DR. Both U-Net and U-NetVGG16 models demonstrated high IoU scores of 98.54 and 98.62 respectively and across other metrics, including accuracy, precision, recall and f1-score. The IoU score indicates that a large proportion of the region segmented by the models overlaps with the actual region in the ground truth [Soulami et al., 2021]. The F1-score represents the harmonic mean of precision and recall and provides a balanced assessment of the models' segmentation performance, considering both false positives and false negatives. The slight improvement observed in the performance metrics of the U-NetVGG16 model compared to the standard U-Net model suggests that leveraging a VGG16-based encoder for feature extraction contributes to enhanced feature representation and discrimination capability. Also, the masks predicted by the model were visualised (Figure 3) to validate the efficiency of the U-Net models.

Figure. 3 illustrates images and masks predicted by the segmentation model (U-NetVGG16). Row 1 shows a retinal image with a predicted mask while rows 2 and 3 showcase images and masks with DR. This visual provides valuable insights into the model's proficiency in identifying DR concentrations in retinal images, supporting robust result validation.

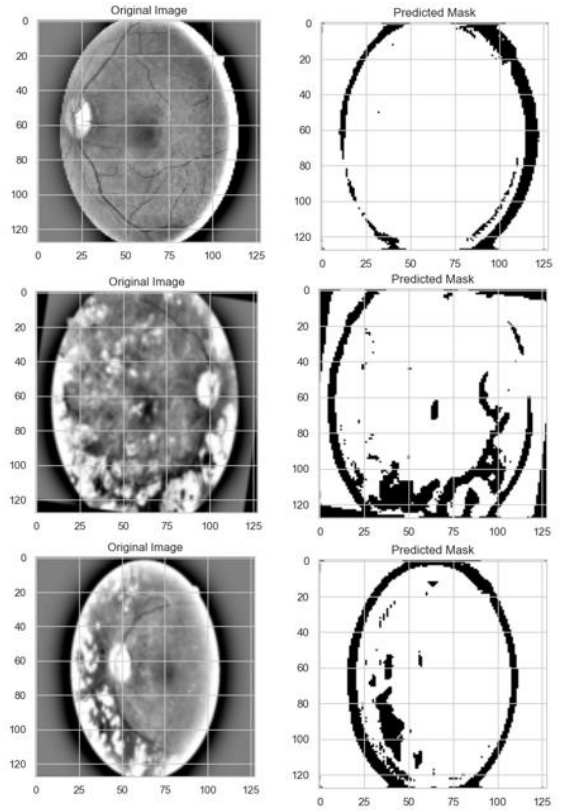

Figure 3: Images predicted by the segmentation model (U-NetVGG16), the arrangement, from left to right, presents original images alongside the model's mask predictions. Row 1 features an image without DR, while rows 2 and 3 show images with DR.

- Performance Comparison with Other Literature

The segmentation model U-NetVGG16 was assessed against comparable works to further validate its efficiency. Table 2 presents the experimental results from different authors. Notably, U-NetVGG16 outperforms Parthasharathi et al. (2022) by 17.4%, Tejpal et al. (2017) by 10.3%, and Anas et al. (2022) by 5.6% in terms of accuracy. The incorporation of transfer learning models into the U-Net architecture (U-NetVGG16), demonstrates superior performance compared to other approaches. This suggests that the approach is highly effective for DR detection.

Table 2: Performance comparison with related works

Author	Model	Accuracy (%)	IOU (%)
Parthasharathi et al. (2022)	VGG16	84.41	-
Tejpal et al. (2017)	U-Net	89.88	-
Anas et al. (2022)	U-Net	93.52	-
This Study	U-NetVGG16	99.10	98.62

CONCLUSION

Diabetic Retinopathy (DR) is a progressive eye disease linked to diabetes, leading to potential vision loss. Early detection is crucial. The traditional diagnostic method is a comprehensive dilated eye exam, where an eye specialist uses an ophthalmoscope to identify abnormalities, such as blood vessel damage. Specialized imaging devices capture fundus images, aiding specialists in manual assessment. However, manual diagnosis is subjective, prone to errors, and causes delays (Shah et al., 2022). The utilization of segmentation models particularly the U-Net architecture holds significant promise in advancing the early detection and management of DR. The findings of this study underscore the effectiveness of U-Net-based segmentation models in automating the identification of DR-related lesions from fundus images. This study applied two segmentation models: U-Net and U-NetVGG16, U-NetVGG16 excelled with an IOU score of 98.62% and an accuracy of 99.10%. The robust performance of the segmentation models highlights their potential for diverse medical imaging applications. Overall, the reported results underscore the potential of deep learning-based segmentation models in facilitating accurate and efficient analysis of image data, with promising implications for various applications in healthcare and beyond.

REFERENCES

[1] Dutta, A., Agarwal, P., Mittal, A. and Khandelwal, S. (2021). Detecting grades of diabetic retinopathy by extraction of retinal lesions using digital fundus images. *Research on Biomedical Engineering*, 37(4), pp.641–

656. doi:https://doi.org/10.1007/s42600-021-00177-w.

[2] Who.int. (2020). *World Diabetes Day 2020: Introducing the Global Diabetes Compact*. [online] Available at: https://www.who.int/news-room/events/detail/2020/11/14/default-calendar/world-diabetes-day-2020-introducing-the-global-diabetes-compact.

[3] Goh, J.K.H., Cheung, C.Y., Sim, S.S., Tan, P.C., Tan, G.S.W. and Wong, T.Y. (2016). Retinal Imaging Techniques for Diabetic Retinopathy Screening. *Journal of Diabetes Science and Technology*, [online] 10(2), pp.282–294. doi:https://doi.org/10.1177/1932296816629491.

[4] Athanasios Valavanidis(2023) 'Artificial Intelligence in Medical Diagnostics and Imaging. Applications that will revolutionize the fields in biomedical research and healthcare'. Available at: https://www.researchgate.net/publication/375714812_Artificial_Intelligence_in_Medical_Diagnostics_and_Imaging_Applications_that_will_revolutionize_the_fields_in_biomedical_research_and_healthcare (Accessed: [January, 2024]).

[5] Xiao-Xia Yin, Le Sun, Yuhan Fu, Ruiliang Lu & Yanchun Zhang. (2022) 'U-Net-Based Medical Image Segmentation'. Available at: https://www.researchgate.net/publication/359998197_U-Net-Based_Medical_Image_Segmentation (Accesed:[January,2024]).

[6] Anas Bilal, Liucun Zhu, Anan Deng, Huihui Lu, Ning Wu. (2022). AI-Based Automatic Detection and Classification of Diabetic Retinopathy Using U-Net and Deep Learning. Journal of Computational and Theoretical Nanoscience, 14(7), 1427-1437.

[7] Revathy, R., Nithya, B. S., Reshma, J. J., Ragendhu, S. S., & Sumithra, M. D. (2020) 'Diabetic Retinopathy Detection using Machine Learning'. Available at: https://www.researchgate.net/publication/342120641_Diabetic_Retinopathy_Detection_using_Machine Learning (Accessed: [August,2023]).

[8] Parthasharathi, G. U., Kumar, K. V., Nivas, R. P., & Kj, J. (2022) 'Diabetic Retinopathy Detection Using Machine Learning'. Available at: https://www.researchgate.net/publication/360649393_Diabetic_Retinopathy_Detection_Using_Machine_Learning (Accessed: [August 2023]).

[9] Gunasekaran, K., Pitchai, R., Chaitanya, G.K., Selvaraj, D., Annie Sheryl, S., Almoallim, H.S., Alharbi, S.A., Raghavan, S.S. and Tesemma, B.G. (2022). A Deep Learning Framework for Earlier Prediction of Diabetic Retinopathy from Fundus Photographs. *BioMed Research International*, 2022, pp.1–15. doi:https://doi.org/10.1155/2022/3163496.

[10] Rakhlin, A. (2018). Diabetic Retinopathy detection through integration of Deep Learning classification framework. doi:https://doi.org/10.1101/225508.

[11] Mujeeb Rahman K K, Mohamed Nasor, and Ahmed Imran (2022). Automatic Screening of Diabetic Retinopathy Using Fundus Images and Machine Learning Algorithms, doi:https://doi.org/10.3390/diagnostics12092262.

[12] Nathan Zhang (2022). Predicting Diabetic Retinopathy Using Machine Learning. *Journal of Student Research*, 11(4). doi:https://doi.org/10.47611/jsrhs.v11i4.3179.

[13] Thippa Reddy Gadekallu, Neelu Khare, Sweta Bhattacharya and Saurabh Singh (2020). Early Detection of Diabetic Retinopathy Using PCA-Firefly Based Deep Learning Model. *Electronics*, 9(2), p.274. doi:https://doi.org/10.3390/electronics9020274.

[14] Mohamed Mahmoud, Salman Alamery, Hassan Fouad, Amir Altinawi, and Ahmed Youssef, (2021). An automatic detection system of diabetic retinopathy using a hybrid inductive machine learning algorithm. *Personal and Ubiquitous Computing*. doi:https://doi.org/10.1007/s00779-020-01519-8.

[15] Penikalapati Pragathi and Agastyaraju Nagaraja Rao (2022) An effective integrated machine learning approach for detecting diabetic retinopathy. Available at: https://www.researchgate.net/publication/359162008_An_effective_integrated_machine_learning_approach_for_detecting_diabetic_retinopathy

[16] Diabetes Retinopathy Dataset, 2021. Available at: Diabetic_Retinopathy_Balanced | Kaggle (Accessed: [August 2023]).

[17] Baccouch, W., Oueslati, S., Solaiman, B. and Labidi, S. (2023). A comparative study of CNN and U-Net performance for automatic segmentation of medical images: application to cardiac MRI. *Procedia Computer Science*, 219, pp.1089–1096. doi:https://doi.org/10.1016/j.procs.2023.01.388.

[18] Awf Abd & Muhammet Baykara (2022) 'A Novel Approach to Detect COVID-19: Enhanced Deep Learning Models with Convolutional Neural Networks'. Available at: https://www.researchgate.net/publication/363656696_A_Novel_Approach_to_Detect_COVID-19_Enhanced_Deep_Learning_Models_with_Convolutional_Neural_Networks (Accessed at [Febuary 2024]).

[19] TensorFlow. (n.d.). Module: tf.keras | TensorFlow Core v2.4.1. [online] Available at: https://www.tensorflow.org/api_docs/python/tf/keras [Accessed 7 Dec. 2023].

[20] Soulami, K.B., Kaabouch, N., Saidi, M.N. and Tamtaoui, A. (2021). Breast cancer: One-stage automated detection, segmentation, and classification of digital mammograms using UNet model based-semantic segmentation. *Biomedical Signal Processing and Control*, 66, p.102481. doi:https://doi.org/10.1016/j.bspc.2021.102481.

[21] Tejpal Virdl, John T. Guibas, Peter S. Li (2017), Synthetic Medical Images from Dual Generative Adversarial Networks. (PDF) Synthetic Medical Images from Dual Generative Adversarial Networks (researchgate.net). Available at https://www.researchgate.net/publication/319524751_Synthetic_Medical_Images_from_Dual_Generative_Adversarial_Networks

[22] Shah, A. D., Jain, S. M., & Patel, M. P. Automatic Screening of Diabetic Retinopathy Using Fundus Images and Machine Learning Algorithms. Available at: https://www.researchgate.net/publication/363674594_Automatic_Screening_of_Diabetic_Retinopathy_Using_Fundus_Images_and_Machine_Learning_Algorithms