INTI JOURNAL | Vol.2024:47
eISSN:2600-7320

Diabetic Retinopathy Detection Model using Hybrid of U-Net and Vision Transformer Algorithms

Mudit Khater

SRM Institute of Science and Technology, Kattankulathur, Chennai, India.

Email: mk6740@srmist.edu.in

Abstract

Diabetic retinopathy is one of the leading causes of vision impairment noticed among individuals with prolonged diabetes. Early-stage detection is very crucial for its treatment. Now, we present a hybrid model which is a combination of U-Net algorithm used for image segmentation and Vision Transformer for classification. The total integration offers a robust model which helps in detecting various stages of diabetic retinopathy. We leverage the use of U-Net algorithm in image segmentation process to delineate the regions of interest in retinal images. Further, the outputs which are segmented are passed into Vision Transformer, which is enhanced by Efficient Net, which is used across various severity levels involved in Diabetic Retinopathy. The usage of transformer architecture helps improve feature extraction and classification performance which ensures that our model captures all patterns in retinal images. We have evaluated our model on APTOS Blindness detection dataset in which our model outperforms traditional convolutional neural networks-based models. Hence, the hybrid approach consisting of combination of both the algorithms demonstrates excellent robustness and generalization which offers a promising application for diabetic retinopathy screening, involving the potential to revolutionize early diagnosis in clinical settings.

Keywords

U-Net, Vision Transformer, APTOS, Diabetic Retinopathy, Neural Networks.

Introduction

Precise image segmentation and classification of the retinal lesions (Zhan, Y., 2020) from medical images obtained is crucial for early-stage diagnosis, treatment planning, and preventing vision loss. Traditionally, this task was done manually by healthcare medical professionals, which is indeed time-consuming, inconsistent, and prone to human fatigue. But in order to tackle these restrictions, sophisticated and automated methods have been launched recently, thanks to advancements in deep learning and machine learning approaches. Convolutional neural networks, particularly the (Ronneberger O., 2016; Zhan, Y., 2020) U-Net, have demonstrated a significant and vast influence (Wong, T.Y., 2017) on medical pictures segmentation because of its extensive capacity to extract both fine-grained spatial information and high-level semantic elements. Diabetic retinopathy is a serious eye condition which is

Submission: 12 November 2024; **Acceptance:** 6 December 2024

bridged to diabetes which can result in blindness, if left untreated. The above work evaluates the U-Net model using the APTOS dataset, (Nelson, P., 2016) which consists of various retinal fundus images which are categorized by the severity of diabetic retinopathy like: No_DR, Mild, Moderate, Severe, and Proliferative_DR. The dataset used is a component of the global APTOS Blindness Detection, which has an aim to improve the diagnosis of diabetic retinopathy. Besides the usage of U-Net, this study also investigates the use of Vision Transformers, (Zhou, Y., 2021) which is a recently developed deep learning architecture that uses attention mechanisms to highlight significant portions of an image provided. Moreover, this further allows for improved feature extraction and (Brox, T., 2015) global context understanding, both of which are really crucial for medical image analysis. Hence, to provide the best possible training, preprocessing techniques (Zagoruyko, S., 2020) like intensity normalization and data augmentation are used. Further, the model's performance and working are evaluated using the metrics like Dice coefficient and Jaccard Index, which measure the overlapping between predicted and the actual lesion areas. The final findings from the above demonstrate how well the U-Net segments (D. Dhanasekaran,2022) and classifies lesions and how addition of Vision transformers enables the model to focus on additional important regions. Subsequent enhancements may entail the application of transfer learning to be applied on pre-trained models, hence enhancing performance for smaller datasets or the introduction of multi-task learning for simultaneous segmentation and classification. In clinical contexts where there is a lack of labelled data, these approaches are helpful because automated tools are essential for precise diagnosis.

Literature Review

Image segmenting and classifying diabetic retinopathy images from retinal fundus images are crucial for early-stage diagnosis, proper treatment and planning and proper patient monitoring. Traditionally, the manual grading done by ophthalmologists is time-consuming, subjective, and prone to variability. However, recent advancements in deep learning techniques, particularly the Vision transformer and convolutional neural networks, have paved the way for automation, efficient, and objective oriented methods. This review explores the use age of the APTOS blindness dataset, which is a large-scale collection of retinal images labelled for different Diabetic retinopathy stages, in combination with the U-Net architecture and Vision transformer for Diabetic retinopathic segmentation and classification criteria. Ensuring the proper identification and investigation of complex retinal lesions such as microaneurysms and exudates which requires efficient and appropriate captures of both high-level semantic information and low-level spatial details, which the encoder-decoder structure of the U-Net with its skip connections provides. By capturing the global contextual relationships, the Vision transformer, which is well-known for its self-attention mechanism, improves the model's capacity to classify images into the proper severity levels of diabetic retinopathy. The main challenges in Diabetic retinopathy segmentation process includes image quality variability, class imbalance due to lesser severe cases available and limited labelled data present. The Techniques like data augmentation, weighted loss functions and transfer learning from the pre-trained models can help in addressing these issues. Future work in Diabetic retinopathy detection may explore and develop explainable AI models to increase trust among clinicians, ensuring that these models are reliable and interpretable in real-world applications usage.

Methodology

The above research project utilizes (Karim N., 2021; Zaman T., 2021) the usage of APTOS blindness dataset alongside with advanced deep learning models which includes U-Net and Vision transformer, to develop a comprehensive and efficient system for segmenting and classifying diabetic retinopathy. The following methodology begins with the process of dataset acquisition and preprocessing techniques, where retinal fundus images are normalized to mitigate the lighting variations and augmented through random rotations, flips and brightness adjustments as per the needs. The above step increases the data diversity, enhancing the model's ability to generalize. For the segmentation tasks, (Zhan Y., 2020) U-Net's encoder-decoder robust architecture with skip connections employed, allows the model to capture high-level features and preserve the intricate spatial details. This architecture is useful and effective in segmenting regions which are affected by diabetic retinopathy, such as haemorrhages and microaneurysms. For the classification, (Ronneberger O., 2016) Vision transformer is used, which treats the images as a sequence of patches and leverages self-attention mechanisms to capture both local and global dependencies which is crucial for assessing the severity of diabetic retinopathy. During the model training, U-Net focuses on the pixel-level segmentation, (Dhanasekaran D., 2022) with classification of severity of the condition, with both models using weighted cross-entropy loss to handle the class imbalances. Additionally, by using the Adam optimizer, (Wong, T.Y., 2017) which improves prediction power, we may adjust the model weights to decrease losses over time. Once the training procedure is performed successfully, the metrics like accuracy, sensitivity, and specificity are used to objectively evaluate the model's performance. Furthermore, the retinal pictures visual overlays of the expected segmentations (Brox T., 2015) can yield qualitative information. Thus, the above-mentioned approach provides a dependable and efficient way of identifying and classifying (Nelson P., 2016) diabetic retinopathy while also adding in the enhancement of diagnostic accuracy by merging Vision transformer (Zagoruyko S., 2020) for classification with the (Zhou Y., 2021) U-Net for segmentation. The figure 1 shows methodology flow chart representation.

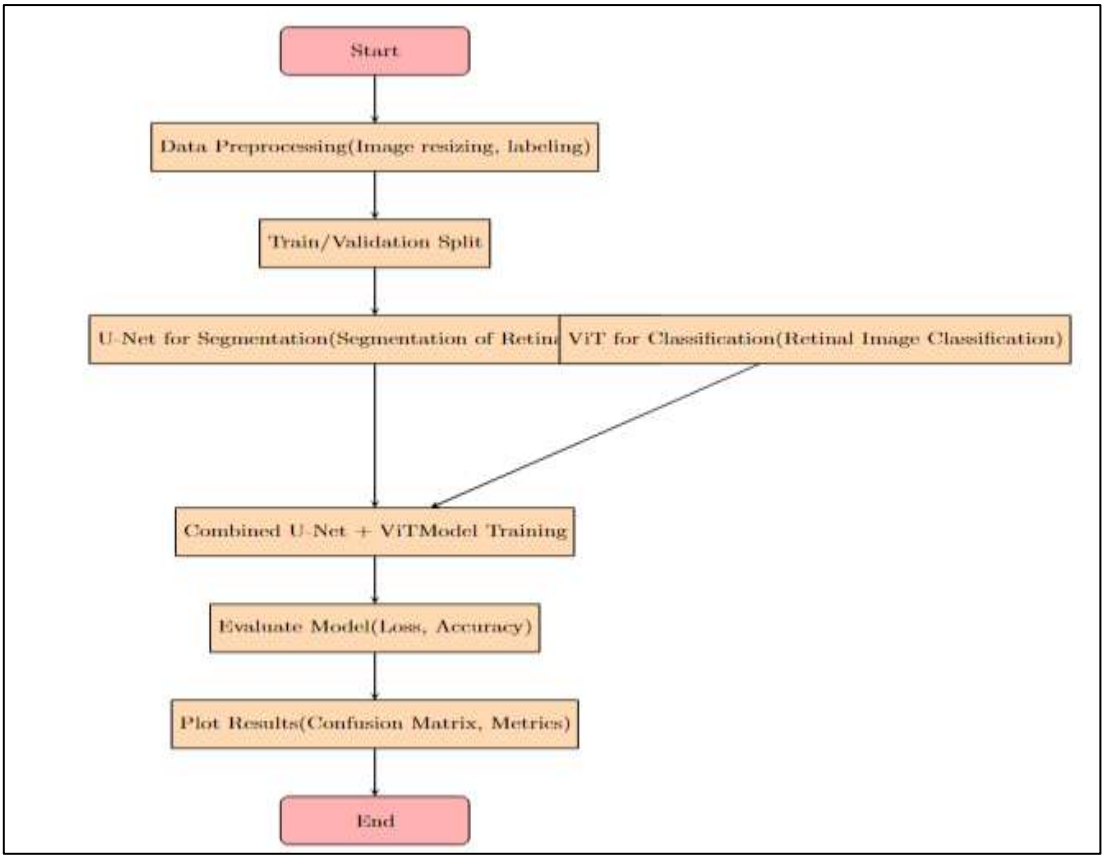

Figure 1: Methodology flow chart representation

Results and Discussions

Our hybrid model, which is a combination of Vision Transformer and U-Net performs well in recognizing and classifying diabetic retinopathy. Using the U-Net for accurate segmentation of afflicted regions, the model uses the architecture to classify retinal pictures into five different stages of diabetic retinopathy. The segmentation masks which the U-Net model generates demonstrate its potential to accurately identify disease-affected regions while segregating pertinent data that are crucial for diagnosis. According to the quantitative study, the model produces good performance indicators the Vision Transformer's self-attention method allows it to detect minute changes in retinal pictures, hence improving feature extraction and classification. This is especially useful for differentiating between mild and moderate stages, which can be difficult for conventional convolutional models to do. The classification highlights the model's exceptional accuracy at all severity levels and highlights its potential as an early diagnosis and treatment tool. Qualitative insights are provided by visual overlays of anticipated segmentations, which demonstrate the model's capacity to pinpoint important disease-affected areas.

The model's ability to manage class imbalances guarantees consistent performance across all classes in addition to the previously mentioned findings. Model’s capacity to generalize is further improved by the introduction of transfer learning, particularly when training on with small datasets. These visual and analytical outcomes describe how the integrated model can assist in the diagnosis and treatment of diabetic retinopathy, especially in

situations where automated analysis could improve many clinical practices. The amount of time required for ophthalmologists to perform manual analysis is greatly reduced by the combination of U-Net and Vision Transformer model, which offers a good framework for early identification and categorization. In addition to increasing diagnostic precision, the automated method can also be used in clinical settings, particularly in underdeveloped regions where there is limited access to specialized treatment. Finally, in order to improve the model's interpretability and build confidence among the medical practitioners, future research may investigate the incorporation of explainable AI mechanisms.

Conclusion and Recommendations

The aforementioned study has established a solid basis for the diagnosis and categorization of diabetic retinopathy disease utilizing the U-Net architecture in conjunction with the Vision transformer, which is based on Efficient Net and uses deep learning techniques to analyse retinal images. With the effective classification of diabetic retinopathy into five severity levels and the automation of the segmentation of complex retinal regions, the suggested model provides a helpful tool to support doctors in the diagnosis and treatment of this vision-threatening disorder. Although our findings show that the model is effective for both segmentation and classification, there are still a number of areas that might be improved for even greater practical usefulness of this technology. The future work, can be considered as follows:

- **Addressing Real-World Data Variability:** The concepts like domain adaptation and data augmentation can further enhance the model's robustness, making it more adaptable and reliable to variations in retinal images from different acquisition devices and patient populations.
- **Beyond Classification: Towards Disease Progression Prediction:** Further future research could be focused on predicting long-term disease outcomes and its treatment responses by correlating it with radiomic features from segmented retinal images with the patient data.
- **Deployment and Clinical Translation:** Rigorous validation and optimization techniques are needed for transitioning the above model into clinical practices, which ensures that it meets the regulatory standards and integrates seamlessly into existing clinical workflows.

References:

Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., & Zagoruyko, S. (2020). End-to-end object detection with transformers. In *European Conference on Computer Vision* (ECCV), Springer, 2020, pages 213–229. https://doi.org/10.1007/978-3-030-58452-8_13

Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Lu, L., Yuille, A.L., & Zhou, Y. (2021). TransUNet: Transformers make strong encoders for medical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention* (MICCAI), 2021. Pages 1–10. https://doi.org/10.48550/arXiv.2102.04306

Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O. (2016). 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. In: Ourselin, S., Joskowicz, L., Sabuncu, M., Unal, G., Wells, W. (eds) Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016. MICCAI 2016. Lecture Notes in Computer Science, vol 9901. Springer, Cham. p.424-432 https://doi.org/10.1007/978-3-319-4723-8_49

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations (ICLR)*, 2021. Pages 1-18. https://doi.org/10.48550/arXiv.2010.11929

Du, G., Cao, X., Liang, J., Chen, X., & Zhan, Y. (2020). Medical image segmentation based on U-Net: A review. Journal of Imaging Science and Technology, 64(1), 1-12. https://doi.org/10.48550/arXiv.2211.14830

Gulshan, V., Peng, L., Coram, M., Stumpe, M.C., Wu, D., Narayanaswamy, A., Venugopalan, S., Widner, K., Madams, T., Cuadros, J., Kim, R., & Nelson, P. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. In *JAMA*, 316(22), pages 2402-2410. https://doi.org/10.1001/jama.2016.17216

Karim, N., Zaeemzadeh, A., and Rahnavard, N. (2019). "RL-NCS: Reinforcement learning based data-driven approach for nonuniform compressed sensing." In 2019 IEEE 29th International Workshop on Machine Learning for Signal Processing (MLSP). The reference citation is from the IEEE publication in 2019, spanning pages 1 to 6. https://doi.org/ https://doi.org/10.1109/MLSP.2019.8918768

Lam, C., Yu, C., Huang, L., & Rubin, D. (2018). Retinal lesion detection with deep learning using image patches. In *IEEE Transactions on Medical Imaging*, 37(4), pages 1018-1028. https://doi.org/10.1167/iovs.17-22721

M. Gomathi, D. Dhanasekaran. (2022). Glioma Detection and Segmentation Using Deep Learning Architectures. *Mathematical Statistician and Engineering Applications*, *71*(4), 452–461. https://doi.org/10.17762/msea.v71i4.523

M, Nasim, Al, Dhali, A., Afrin, F., Zaman, N. T., & Karim, N. (2021). The Prominence of Artificial Intelligence in COVID-19. *ArXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2111.09537

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention* (MICCAI), Springer, 2015, pages 234–241. https://doi.org/10.1007/978-3-319-24574-4_28

Tan, M., & Le, Q.V. (2019). EfficientNet: Rethinking model scaling for convolutional neural networks. In *Proceedings of the 36th International Conference on Machine Learning* (ICML), 2019. Pages 6105–6114. https://doi.org/10.48550/arXiv.1905.11946

Ting, D.S.W., Cheung, C.Y., Lim, G., Tan, G.S.W., Quang, N.D., Gan, A., Hamzah, H., Garcia-Franco, R., San Yeo, I.Y., Lee, S.Y., & Wong, T.Y. (2017). Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multi-ethnic populations with diabetes. In *JAMA*, 318(22), pages 2211–2223. https://doi.org/10.1001/jama.2017.18152