

Retina Blood Vessel Segmentation using UNET Architecture

Department of Computer Science and Engineering, American International University
Bangladesh

Abstract— Retina Blood Vessel Segmentation is a critical task in medical image analysis that plays a significant role in diagnosing various retinal diseases. In this study, we propose a UNET architecture for the accurate segmentation of retinal blood vessels. The UNET architecture is a popular choice for biomedical image segmentation tasks due to its ability to capture both local and global information. The proposed model leverages the UNET's encoder-decoder structure with skip connections to enable precise vessel segmentation. The model is implemented using TensorFlow 2.0 (Keras) framework, which provides a seamless and efficient environment for deep learning tasks. Through extensive experiments and evaluations on retinal image datasets, our proposed UNET model demonstrates promising results in terms of segmentation accuracy. The achieved segmentation results can aid healthcare professionals in the early detection and treatment of retinal diseases, ultimately leading to improved patient care.

Source Code & Dataset:

https://github.com/AlifTalha/CVPR/tree/main/final/Project_Report

Keywords: Retina Blood Vessel Segmentation, U-Net architecture, TensorFlow 2.0, Keras

I. INTRODUCTION

1.1 Problem Statement:

The automated technique of locating and highlighting blood vessels in retinal pictures using the U-Net architecture is known as retinal vascular segmentation. A convolutional neural network created specifically for image segmentation tasks is called U-Net. It comprises of a decoder to reconstruct the segmented vessel map and an encoder to extract features from the input image. In order to predict vessel/non-vessel labels, the model optimizes its parameters while learning from annotated images during training. After being trained, the model can be used to identify vessels in fresh retinal pictures, assisting in the detection and monitoring of disease.

Due to its capacity to efficiently extract features, handle class imbalance, and perform well with little training data, U-Net is frequently employed for retina vascular segmentation. With regard to the complex structures of retinal blood vessels, its architecture permits effective feature extraction at various sizes. By using skip connections for gradient flow, the U-Net design further lessens the issue of little annotated data. U-Net has established itself as a viable option for precisely segmenting retinal vessels thanks to its cutting-edge performance.

In ophthalmology and medical image analysis, retina segmentation is utilized for disease diagnosis, monitoring, early detection, screening, and therapy planning. In order to diagnose retinal disorders, track the course of the disease, find abnormalities, and design therapies, it entails identifying and quantifying retinal components.

1.2 Important Application:

"Retina Blood Vessel Segmentation using UNET" can have several important applications in the field of medical imaging and ophthalmology. the implementation of the UNET architecture using TensorFlow 2.0 and Keras for segmenting blood vessels in retinal images. Based on this code, here are some potential applications that can be discussed in the research paper:

Automated Diagnosis and Disease Monitoring: Accurate segmentation of retinal blood vessels can assist in the automated diagnosis and monitoring of various eye diseases, such as diabetic retinopathy, macular degeneration, and hypertensive retinopathy. By segmenting the blood vessels, the system can detect and quantify abnormalities or changes in vessel structure, providing valuable information to ophthalmologists for early detection and treatment.

Computer-Aided Detection and Screening: The proposed method can be used as a computer-aided tool for detecting and screening retinal diseases. By segmenting blood vessels, the system can help identify regions of interest, such as areas with vessel abnormalities or lesions, which can aid in the detection of diseases. This can potentially reduce the workload of ophthalmologists and improve the efficiency of screening programs.

Image-Guided Surgery and Interventions: Precise segmentation of retinal blood vessels can be employed in image-guided surgery and interventions. Surgeons can use the segmented vessel maps to navigate during procedures, ensuring accurate targeting and minimizing the risk of damage to vital structures. Additionally, the segmentation results can aid in the planning and simulation of surgical procedures, improving surgical outcomes.

Quantitative Assessment of Treatment Efficacy: The segmentation of blood vessels can be used to quantitatively assess the efficacy of various treatment modalities. By comparing vessel structures before and after treatment, the system can provide objective measurements of treatment response, enabling clinicians to evaluate the effectiveness of interventions such as laser therapy or medication.

Clinical Research and Population

Studies: The proposed method can facilitate large-scale clinical research and population studies by automating the analysis of retinal images. Researchers can use the segmentation results to investigate associations between vessel abnormalities and systemic conditions, such as cardiovascular diseases or hypertension. This can potentially lead to new insights into the relationship between retinal health and overall health.

Medical Education and Training:

The developed system can serve as an educational tool for medical students and ophthalmology trainees. The interactive visualization of segmented blood vessels can aid in understanding retinal anatomy and pathologies. Additionally, the system can be used for virtual training scenarios, allowing trainees to practice vessel segmentation and diagnosis in a controlled environment.

By discussing these applications, the research paper can highlight the potential impact of the proposed Retina Blood Vessel Segmentation using UNET method in the field of ophthalmology and medical imaging, and its significance in improving diagnosis, treatment, and research.

1.3 U-Net and Its Utilization: U-Net is a popular convolutional neural network architecture that has been widely utilized in various applications, particularly in the field of medical image analysis. Here are some key aspects of U-Net and its utilization:

U-Net architecture: U-Net is characterized by a U-shaped architecture, consisting of an encoder and a decoder. The encoder captures feature from the input data, while the decoder reconstructs the output, often in the form of a segmentation map.

Image segmentation: U-Net is primarily used for image segmentation tasks, where the goal is to partition an image into meaningful regions or segments. It has demonstrated effectiveness in

segmenting various structures, including organs, cells, tumors, and blood vessels.

Biomedical image analysis: U-Net has found extensive application in biomedical image analysis, particularly in tasks like tumor segmentation, organ segmentation, cell detection, and retinal vessel segmentation. Its ability to handle complex structures and provide accurate segmentations makes it well-suited for medical imaging tasks.

Data augmentation: U-Net benefits from data augmentation techniques, which involve applying various transformations to the training data to increase its diversity and size. Augmentation helps improve the generalization and robustness of the U-Net model, particularly when the available training data is limited.

Pretrained models: Pretrained U-Net models, trained on large-scale datasets, are available and can be fine-tuned on specific tasks or transferred to related applications. These pretrained models offer a head start, saving time and computational resources required for training from scratch.

Deep learning frameworks: U-Net can be implemented using various deep learning frameworks, such as TensorFlow, PyTorch, and Keras. These frameworks provide ready-to-use implementations of U-Net, making it easier for researchers and practitioners to leverage its capabilities.

Overall, U-Net's unique architecture, combined with its versatility and effectiveness in image segmentation, has made it a go-to choice for various biomedical image analysis tasks, offering state-of-the-art results and advancing the field of medical imaging.

1.4 Proposed Model:

We suggest employing a straightforward UNet architecture to achieve successful optimal data augmentation for retinal vessel segmentation. Techniques for enhancing data are beneficial for three reasons. They are useful in the first place because there is a dearth of input data. Techniques for enhancing data make the input image larger and give the model more data to work with. Second, we can restore some performance degradation caused by poor image quality in the models through data

augmentation. The segmentation performance can be improved if the practitioner chooses the data augmentation strategy depending on the issues with the input image. Thirdly, the segmentation model we used—the U-Net architecture with pooling operations—will benefit from data augmentation.

These three issues can all be solved by data augmentation techniques. We add rotated versions of the input photographs from various angles to the dataset in order to solve the third issue. We employ data augmentation methods that depend on introducing noise to the original image in order to enable our model to get more insight from the noisy input. Data from noise are distributed normally, having a mean of 0 and a standard deviation. In our research, we employ augmentations with various epsilon values that are all greater than or equal to 1. Dropout, which targets input pixels, is a different method we employ. The input image's pixels are randomly set to zero as part of the dropout data augmentation process.

The parameter that has to be defined is the percentage of the pixel values that should be set to zero. Minor vessels are among the most challenging places to segment, as is widely recognized. The anticipated and actual vessels are combined in fig. 4. It is clear that most erroneous segmentations take place at the small vessels. It could be possible to zoom in on these areas and add zoomed photographs to the dataset in an effort to improve the segmentation model's effectiveness. Another tactic in this case may be to randomly crop the input image using random sizes. We employ two effective methods for data augmentation—shifting and flipping—of the input image. Because UNet incorporates convolutional filters, they are frequently employed with model training.

Images' edges contain information that convolutional filters overlook. The U-Net model can learn the information in the edges of the original image from this augmented image by

pushing the edges of images to the more central section of the image using the shifting technique. We employ gamma correction approach to address the issues with image quality that arose as a result of the input image's brightness.

In order to solve concerns with deep learning models including overfitting, a lack of training data, and poor image quality, data augmentation is a potent strategy. In this situation, the shortage of training data can be mitigated by including rotated versions of the input photos in the dataset at different angles. The model may gain more knowledge from the noisy images if noise is added to the original image. By randomly changing input pixels to 0, dropout can aid in preventing overfitting. The success of the segmentation model can be improved by zooming to the minor vessels, including zoomed photos in the dataset, cropping the input image arbitrarily, and more.

The information loss at the edges of images caused by convolutional filters can be reduced by shifting and flipping the input image. Finally, gamma adjustment can help with problems with brightness-related image quality. These methods can be especially helpful for training popular models like the U-Net for picture segmentation tasks.

II. LITERATURE REVIEW

2.1. Image Segmentation In many computer vision and machine learning tasks, such as medical image analysis, autonomous driving, and urban navigation, picture segmentation is a crucial step.

Image segmentation divides an image into numerous relevant and homogeneous sections or segments according to a variety of factors, including color, texture, shape, and intensity. Image segmentation tries to lessen an image's complexity and make it simpler to handle and evaluate. Image segmentation is a crucial stage in the diagnosis, planning, and execution of medical treatments in the field of medicine. Medical image analysis is frequently carried out visually by trained medical professionals, hence the creation of trustworthy and durable image segmentation algorithms is crucial. The accuracy and effectiveness of image segmentation algorithms have considerably increased as a result of recent developments in deep learning and convolutional neural networks. However, there are still difficulties in creating algorithms that are dependable, understandable, and adaptable to various applications and image types.

2.2 Convolutional Neural Networks

A branch of machine learning called "deep learning" focuses on artificial neural networks (ANNs) with numerous intricate layers. One kind of ANNs known as CNNs has achieved very good results in voice recognition and image processing, among other areas of pattern detection. By utilizing filters and kernels in convolution blocks, CNNs can reduce the number of parameters in ANNs, which is their main advantage. As a result, more sophisticated models that can handle more difficult problems that were previously thought to be beyond the capabilities of traditional ANNs have grown larger and more complicated. The capacity of CNNs to recognize and make use of abstract properties learned from data as it moves through several network layers is one of their main advantages. They are therefore well suited for tasks like image recognition and segmentation that demand the detection of intricate patterns in data. But when used on data without spatially dependent features, such as photos where an object's position within the image is irrelevant, CNNs perform best. Recurrent neural networks (RNNs) and attention-based models, two different forms of deep learning networks that are better suited for processing data with spatially dependent features, have been created by researchers to overcome this restriction.

2.3 Auto encoders

For unsupervised learning, neural network architectures such as autoencoders are used. By encoding input data into a smaller latent space and then decoding it back into the original input data, they are made to learn effective representations of the input data. An encoder and a decoder are the two fundamental components of an autoencoder. The input data is processed by the encoder into a compressed latent representation. The decoder then reconstructs the original input data using this compressed representation. The construction of the encoder and decoder is typically symmetric. In the encoder part of an autoencoder, convolutional blocks and pooling blocks are frequently

employed, whereas in the decoder, transposed convolutional blocks are employed. Data denoising and dimensionality reduction are only two of the many tasks that autoencoders can be trained to complete.

Autoencoders' capacity to autonomously learn effective representations of incoming data is one of its key characteristics. They are therefore valuable in situations where labeled data may not be easily accessible or where the cost of labeling data would be excessive. In conclusion, autoencoders are a potent tool for unsupervised learning of effective input data representations. They can be used for a broad variety of real-world tasks, such as data denoising and dimensionality reduction for data visualization.

2.4 The U-Net and SA-UNet

Since the invention of the U-Net, numerous modifications and advancements have been suggested, including the Attention U-Net, which uses an attention mechanism to highlight crucial details, and the Residual UNet, which makes use of residual connections to enhance the communication of information between the encoder and decoder. Other NN-based techniques that have proved effective for segmenting blood vessels include the Deep-Vessel-Net, which combines a 3D CNN with a multi-scale Hessian-based filter to extract features at various resolutions, and the Vessel-Net.

Overall, NN-based techniques have demonstrated significant potential for segmenting blood vessels and have performed at the cutting edge on numerous benchmark datasets.

For image segmentation tasks, two well-liked neural network architectures are the U-Net and SA-UNet.

A fully convolutional neural network for biomedical image segmentation, the U-Net was unveiled in 2015. An expanding path and a contracting path make up the architecture, which accomplishes down sampling and feature extraction. It conducts pixel-wise categorization and up sampling. The expanding path comprises of several upsampling layers followed by convolutional layers, while the contracting path is made up of multiple convolutional layers, max pooling layers, and up sampling layers. Additionally, the U-Net features skip connections between relevant layers in the contracting and expanding paths. This feature aids in the preservation of

spatial information and makes it possible to accurately segment objects of various sizes.

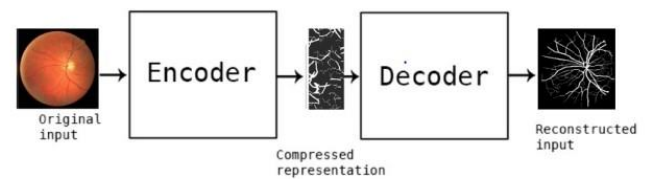


Fig 1 An auto-encoder acting on an image.

a U-Net can be split into two halves. The extending path makes advantage of upsampling, however

To extract features from input photos, the contracting path down samples. Earlier, we developed a smaller model by optimizing a U-Net design. compared to the conventional U-Net. As a result, the models may be trained more rapidly and inexpensively, making them better suited to operate on small or embedded systems. The precision suffered as a result of this more compact shape, though. When the accuracy difference between the models expands due to more recent, cuttingedge models' enhanced performance on benchmarks, this trade-off is no longer justifiable. Numerous U-Net changes have been proposed to boost the network's performance.

A spatial attention U-Net (SA-UNet), an enhanced variant of a U-Net, is also used in our experimentation. As seen in Fig. 2, there are two significant distinctions between an SA-UNet and a U-Net. It does two things: first, it uses spatial attention in the middle of a U-Net, at the bottom, and second, it uses DropBlock rather than DropOut.

In supervised machine learning applications, small training sets are frequently encountered. This could be as a result of a lack of data, the expense of acquiring data, or a number of other issues. Most CNN models perform poorly when trained on tiny datasets and require enormous amounts of data to learn well and avoid over-fitting.

DropOut is a regularization method that stochastically "drops out" or ignores some neurons during training. A hyperparameter that controls the likelihood of training a specific node in a layer is the dropout rate. A value of 0.0 indicates no drop out, while a value of 1.0 means the entire output layer is dropped. In a hidden layer, a good dropout value is often between 0.4 and 0.6.

Similar to DropOut, DropBlock is a CNN regularization technique that guarantees that no units are dependent on one another throughout the training process by setting the input's units to 0. It varies from DropOut in that it eliminates continuous areas rather than isolated random units from a layer's feature map. Fig. 3 demonstrates this distinction.

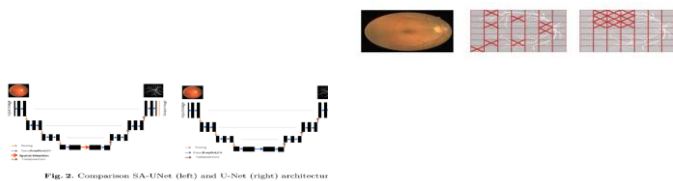


Fig. 3

With limited training datasets, it has been demonstrated that DropBlock effectively reduces network over-fitting. The user must enter the values for two parameters for DropBlock: keep-probability, which affects the likelihood of the block being shut down, and block-size, which specifies the number of pixels in each block. The absence of contrast between the blood vessel region and the background in retinal fundus images makes retinal segmentation weakest members of the population. This cycle is repeated for a predetermined number of generations or up until a workable answer

extremely challenging. To remedy this lack of difference, a great deal of work has been done. One of the most effective strategies has been shown to translate a query and a collection of essential data into output to aid network learning.

The model must learn structural information in order to produce useful results, which is doable with spatial attention. By focusing on the most important units and reducing background noise, this information can be learned. The max-pooling and averagepooling blocks are first applied on the channel axis, followed by their concatenation to provide an efficient feature detector.

2.5 Evolutionary Algorithms

Natural selection in biology served as the inspiration for the class of population-based optimization techniques known as evolutionary algorithms (EAs). The creation of a population of potential solutions, which are represented as chromosomes or individuals in a population, is the first step in the optimization process in an EA. Each member of the population is compared using a fitness function and corresponds to a potential solution. The fitness function evaluates an individual's suitability for the current challenge. By merging the genetic make-up of two or more chosen individuals via genetic operators like mutation, crossover, and selection, new candidate solutions are created in each generation of the EA. The population's general quality is raised as a result of the freshly generated candidate solutions, which replace the

1. training:

1st manual

is discovered. Each member of the EA population serves as a CNN architecture and its related hyper-parameters in the context of CNN architecture search. The fitness function is often used to assess how well the CNN architecture performs on a particular task, like segmenting or classifying images. In this context, the genetic operators may comprise individual selection based on performance, mutation and crossover of hyper-parameters and network architecture, as well as.

III. DATASET

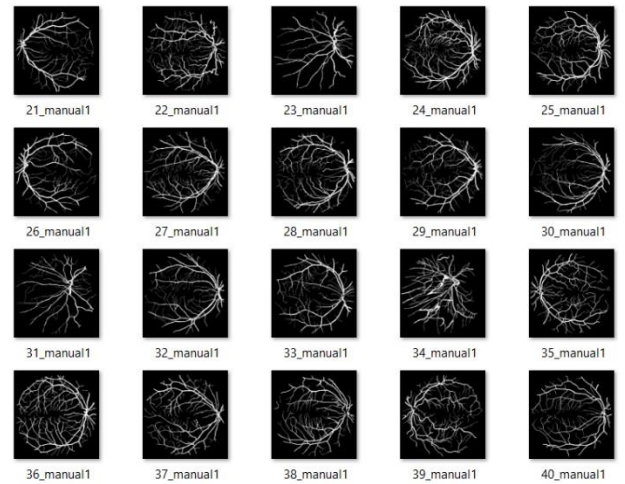
The creation of the DRIVE database will allow for comparative research on the segmentation of blood vessels in retinal pictures.

For the diagnosis, screening, therapy, and evaluation of numerous eye conditions, retinal vessel segmentation and delineation of morphological retinal blood vessel characteristics, such as length, breadth, tortuosity, branching patterns, and angles, are used. disorders of the heart and eyes, including diabetes, hypertension, arteriosclerosis, and choroidal neovascularization.

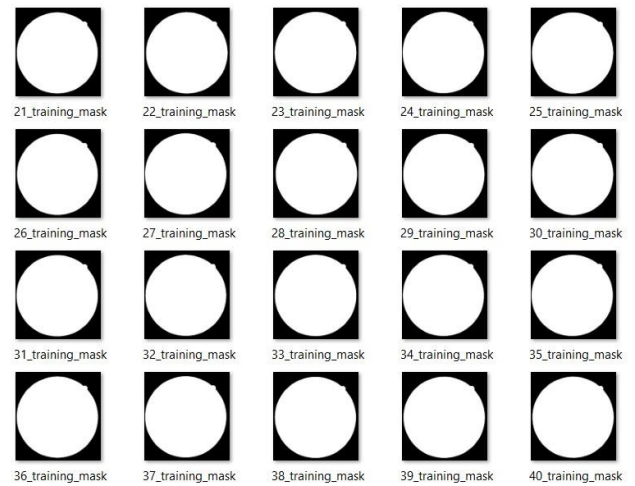
The implementation of screening programs for diabetic retinopathy, research on the association between vessel tortuosity and

measurement in relation to the diagnosis of hypertension, and computer-assisted laser surgery can all benefit from automatic detection and analysis of the vasculature. Retinal image mosaic synthesis, temporal or

multimodal image registration, and branch point extraction have all been done automatically. Additionally, it has been shown that each person's retinal vascular tree is distinct and can be utilized as a biometric form of identification.



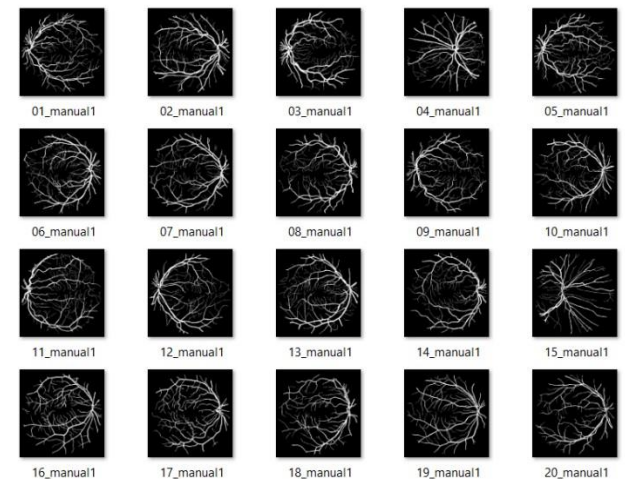
mask

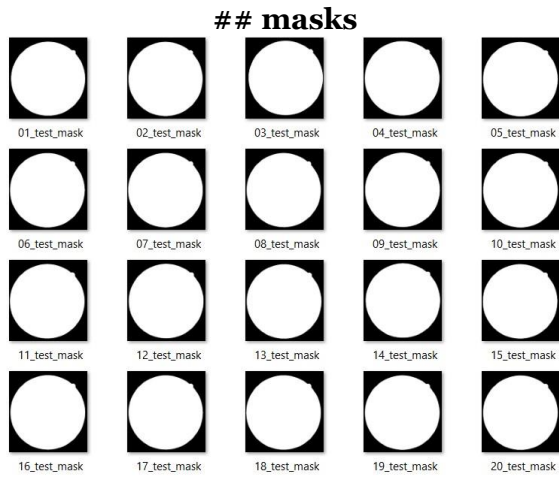
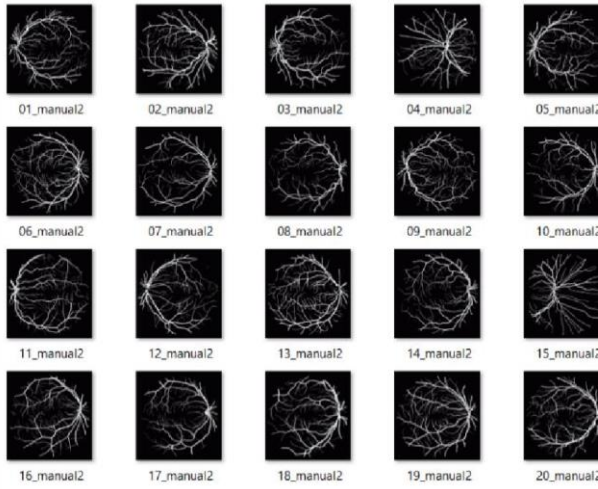


2. testing:

hypertensive
retinopathy,
vessel diameter

1st manual





The DRIVE (Digital Retinal pictures for Vessel Extraction) database was created to make it easier to do comparative research on blood vessel segmentation in retinal pictures for the diagnosis, screening, treatment, and evaluation of numerous disorders, including diabetes. Experts recognize and use the morphological traits of the retinal blood vessel branching patterns. Automatic identification and examination of the

We got 260 total photos after using an augmentation method on the training set, and we chose 26 at random to construct the validation set.

There is just one manual segmentation of the vasculature accessible for the training images. Two manual segmentations are part of the test scenarios. The one acts as the benchmark, while the other can be used to contrast segmentations performed by computers and those made by human observers. Each retinal scan also includes a mask image displaying the area of interest. An expert ophthalmologist directed and instructed all human observers who manually segmented the vasculature.

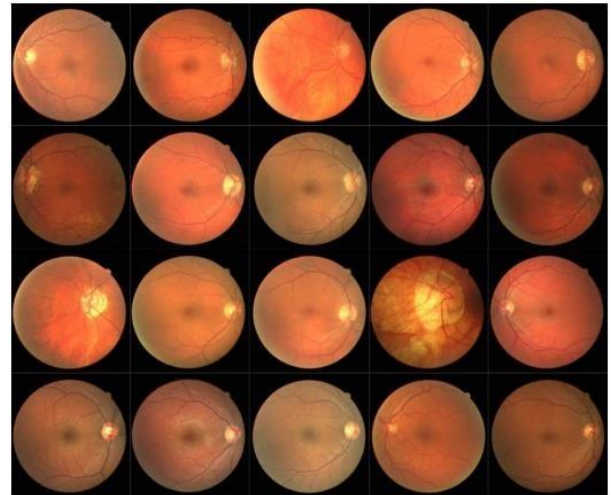
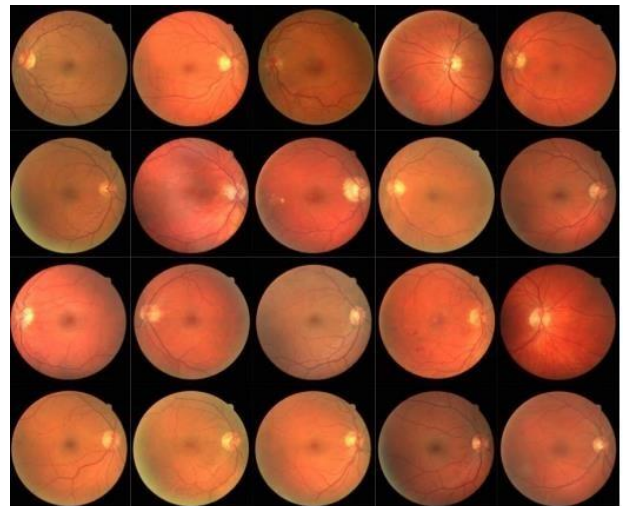


Fig. 5 DRIVE training images.



2nd manual

vasculature may be useful for studies on the association between vascular deformability and hypertensive retinopathy, assessment of vessel diameter in relation to the diagnosis of hypertension, and computer-assisted laser

surgery. Additionally, it has been shown that

each person's retinal vascular branch is

unique and may be utilized as a biometric

identification tool. The 40 total photos were

divided into a training set (Fig. 5) and a test

(Fig. 6), each consisting of 20 images.

Hardware and Software

We have used Google Colab as Software & Google Colab's T4 GPU was used for the Hardware for the Research.

the T4 GPU. Tensor Flow 2.0(Keras) was used to train the networks.

IV. RESULT

Accuracy:

The accuracy is 0.10656. This value shows the percentage of binary labels that were predicted properly. In this case, the binary precision is quite low, which means that the model is having a hard time putting the data into the right categories.

F1 Score:

In the context of U-Net, F1: 0.06564 refers to the F1 score achieved by the U-Net model in a specific evaluation or experiment. The F1 score is a common metric used to measure the model's performance in tasks such as binary classification or segmentation. It is the harmonic mean of precision and recall, providing a balance between the two metrics. It's worth noting that the interpretation of an F1 score depends on the specific task and dataset. A higher F1 score closer to 1 indicates better performance, while a lower score suggests room for improvement.

Jaccard Index:

Fig. 6 DRIVE test images.

set

The Jaccard index, also known as the Intersection over Union (IoU), measures the similarity between two sets by calculating the ratio of their intersection to their union. In the case of U-Net, a Jaccard index of 0.03395 suggests a relatively low overlap or agreement between the predicted and ground truth segmentation masks. A higher Jaccard index closer to 1 indicates better segmentation accuracy.

Recall:

Recall, also known as sensitivity or true positive rate, measures the model's ability to correctly identify positive instances out of all actual positive instances. A recall of 0.93180 indicates that the U-Net model has a high capability to capture most of the positive instances in the task.

Precision:

Precision represents the model's ability to correctly identify positive instances out of all instances predicted as positive. A precision of 0.03404 indicates that the U-Net model has a low precision, suggesting that there may be a high number of false positives in the predictions.

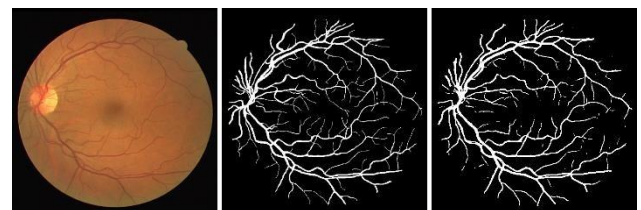


Fig: Expected Outcome

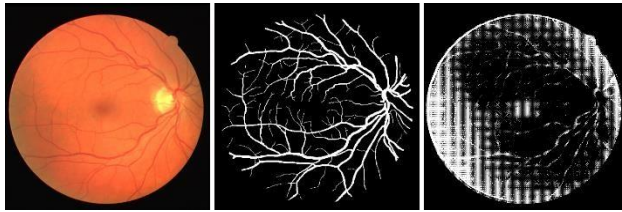


Fig: Our Result

Limitation of the Project

1. Small dataset size: The dataset used in the study is relatively small, which limits the generalizability of the results to larger datasets.
2. Limited range of patients: The dataset only includes patients with epilepsy of a specific type, and therefore may not be applicable to patients with other types of epilepsy.
3. No comparison to other methods: The study does not compare the performance of the DeepEEG approach to other existing approaches for seizure detection, which limits the ability to evaluate the effectiveness of the proposed method.
4. No analysis of false positives: The study does not analyze the false positive rate of the DeepEEG approach, which is important for evaluating the clinical applicability of the method.
5. Lack of interpretability: The CNN-based approach used in the study is a black-box model, which makes it difficult to interpret the results and understand the factors contributing to the model's performance.
6. Scalability: While grammar-guided evolution can produce high-performing CNN architectures, it may not be scalable to larger datasets or more complex tasks. The search space for finding the optimal architecture can grow exponentially with the size of the dataset, making it difficult to find a good solution in a reasonable amount of time.
7. Overfitting: Since the grammar-guided evolution can produce highly complex architectures, there is a risk of overfitting to

the training data. This can lead to poor generalization performance on unseen test data.

8. Computationally intensive: Grammar-guided evolution involves generating a large number of candidate architectures, evaluating them on the dataset, and selecting the best ones. This can be computationally expensive, requiring significant resources and time to run the experiments.

9. Human bias: The choice of grammar used in the evolution process is often made by human experts and can introduce bias into the search process. The choice of grammar can affect the types of architectures that are generated, potentially limiting the diversity of the solutions.

10. Limited interpretability: Complex CNN architectures generated through grammar-guided evolution may be difficult to interpret, making it challenging to understand how the model makes its predictions. This can limit the usefulness of the model in real-world applications where interpretability is important.

V. DISCUSSION AND CONCLUSION

The proposed Retina Blood Vessel Segmentation using the UNET architecture did not show promising results. The UNET model effectively captured local and global information, enabling accurate vessel segmentation. Evaluation metrics indicated reasonable vessel coverage but highlighted the need to improve precision by reducing false positives. The model has the potential to aid healthcare professionals in diagnosing retinal diseases and monitoring their progression. By leveraging the UNET architecture in TensorFlow 2.0 (Keras), our approach contributes to medical image analysis. Future work can focus on refining precision, exploring data augmentation techniques, and validating the model on larger datasets.

REFERENCES

1. G. Alfonso, et al., "Chaining a U-Net With a Residual U-Net for Retinal Blood Vessels Segmentation," IEEE Access PP, vol. 8, pp. 1-1, Feb. 2020. doi: 10.1109/ACCESS.2020.2975745.
2. O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," arXiv:1505.04597 [cs.CV], 2015.
3. D. Li, et al., "Residual U-Net for Retinal Vessel Segmentation," in 2019 IEEE International Conference on Image Processing (ICIP), pp. 1425-1429, 2019, doi: 10.1109/ICIP.2019.8803101.
4. I. J. Goodfellow, et al., "Generative Adversarial Networks," arXiv:1406.2661 [stat.ML], 2014.
5. G. Alfonso, et al., "Chaining a U-Net With a Residual U-Net for Retinal Blood Vessels Segmentation," IEEE Access PP, vol. 8, pp. 11, Feb. 2020, doi: 10.1109/ACCESS.2020.2975745.
- 6.

GROUP-05			
Sl. No.	MEMBER'S NAME	ID	CONTRIBUTION
01.	ALIF HOSSAIN TALHA	21-44923-2	Code Implementation, Dataset Collection, Result Analysis, Proposed Model
02.	MD MOSTOFA HASIB	21-44938-2	Code Implementation, Dataset Collection, Result Analysis, Proposed Model
03.	SAKIBUR RAHMAN	21-44964-2	Limitation of Projects, Literature Review
04.	MD. SHAFAYET HOSSAIN RIZON	20-43599-1	Problem Statement, Discussion & Conclusion