

# Automated Measurement of AVR Feature in Fundus Images using Image Processing and Machine Learning

Abhay N. Nadkarni<sup>1</sup>, Ashish A. Phadatare<sup>1</sup>, Aditya B. Shah<sup>1</sup>, Prof. Sushma Kadge<sup>2</sup>

<sup>1</sup>Department of Electronics Engineering, K. J. Somaiya College of Engineering (Autonomous), (Affiliated to University of Mumbai), Mumbai, India.

<sup>2</sup>Assistant Professor, Department of Electronics Engineering, K. J. Somaiya College of Engineering (Autonomous), (Affiliated to University of Mumbai), Mumbai, India.

\*\*\*

**ABSTRACT**— Arteriolar-to-venal ratio (AVR) is an important parameter to predict whether a person may suffer from cardiovascular diseases, retinopathy of prematurity, diabetic retinopathy, cerebral atrophy and stroke among many others. However, adequate measures can be taken to avoid the harmful effects of these diseases or even cure them if changes in AVR can be detected early on. Manual and semi-automatic methods are not only difficult to implement, but they also prove to be inefficient and less accurate while dealing with thousands of images available. This paper explores and examines a proposed fully automatic method of grouping of blood vessels into arteries or veins, computation of arteriolar-to-venular ratio i.e. AVR and all other intermediate steps involved in the process. This paper corresponds to the project our team has undertaken and expounds on the stages completed. The software used in the project are MATLAB and Python. Images have been retrieved from DRIVE database [1].

**INDEX TERMS**—Artery vein ratio, retina, ROI, Vessels, Artery, Vein.

## 1) INTRODUCTION

Cardiovascular diseases (CVDs) are one of the major causes of mortality around the globe. They are largely caused by heart strokes and heart related diseases. Such diseases can severely affect the thickness of retinal vessels, alter their curves, and/or transform their reflectance of light. Ophthalmologists have recognized this fact for years, viz. shrink of the thickness of arteries and expand the thicknesses of veins is possible. However, with recent advancements of technology, it is achievable to obtain accurate and precise measurement of retinal vessel parameters. This led to the possibility of recording subtle changes in the arteriolar-to-venal diameter ratio (AVR). These deviations are directly correlated with rises in the threat for cerebral atrophy, stroke myocardial infarct, and cognitive decline. An early discovery of the abnormality in the ratio can prevent severe damages in the future.

Unfortunately, significant deviations in AVR are too refined to be identified by ophthalmologists through regular clinical inspection. Moreover, manually assessing the AVR for every individual through a digital retinal fundus image is not pragmatic and very cumbersome for clinical practice.

Hence, an automatic process is favorable to calculate the AVR as it will revolutionize medical clinics, which will steer the process to an efficient evaluation of patients all around the globe with such diseases.

Our paper attempts to calculate AVR automatically on any given color fundus image. It involves six core stages: Preprocessing, Vessel Segmentation, ROI detection, Vessel Width Measurement, Classification of vessels, and finally calculation of the ratio. The goal is to make all these stages automated where no input from the user is required.

This paper has been divided into five sections. The first section describes the literature survey done in order to facilitate this project, highlighting the work which was previously done in this field and discussing their significance in solving the problem at hand. In the next section, there are details on the software used for this project, viz. MATLAB and Python wherein its merits and advantages over other frameworks are discussed. In the subsequent section, there is an in-depth discussion on the stages of the project which have been completed. In the final section, we talk about the future scope of this project and how to achieve the targets we have highlighted in the previous sections.

## 2) LITERATURE SURVEY

The primary source of information while undertaking this project was the work by Meindert Niemeijer et al. [2]. This study highlighted various stages involved in actually automating the entire process of calculating the AVR of numerous fundus images, right from preprocessing of the image till classification of the blood vessels into arteries and veins and then

eventually obtaining the parameter of AVR for each image. The paper shed some light on the earlier accomplishments made in this field, which we summarize in the upcoming paragraphs.

Various methods have been previously presented for classifying AV and calculating AVR. A semi-automated process involved separate analysis for arterial and venous branches from the retinal vascular trees in 2002.

Another paper put forward a process which would automatically determine the AVR, though, in that approach manual user input was still needed for separation of vessels in 2005.

More recent work, in 2007, proposed a system to distinguish vessels into arteries or veins by a prevailing vessel division model where a selected few were manually categorized initial divisions. Grisan and Ruggeri were one of the first ones to propose the methodology of automated AV classification. They suggested that the vasculature be divided using 'vessel tracking and analysis' where the vessel centerlines need to be identified. After outlining the ROI (a certain area of significance) around the optic disc and splitting this region into four sections, and eventually classifying into AV by an unsupervised clustering manner. This is done with the aid of features centered on color extracted from the vessel divisions. On 24 images, a total error rate of 12.4% was found using this procedure.

Kondermanns and Yan put forth an approach of classifying AV that uses features obtained from profiles of vessels along with the characteristics based on the intensities of the local image along the vessel centerline. 95.32% accuracy was achieved in this process due to the manual assignment of vessel pixels to the precise class of four images.

In 2010, a method including a combination of clustering and classification for classifying AV was proposed. A comparison of distinctive sets of feature and classification methodologies was made by the authors. An accuracy level of 86.34% was achieved while grouping vessels and this comprised of the effect of the space to the optic disc on their calculations.

The system proposed and evaluated in this project is a combination of approaches already described in the literature along with new, projected systems. The method includes automatic recognition of the optic disc, segmentation of the vasculature, determination of the vessel width, categorization of the vessels and calculation of the parameter called AVR in the region of interest. We will also certify the classification scheme of the vessels since it is an important subcomponent that the AVR depends on directly. The stages involved in the work by Meindert Niemeijer et al. are taken as reference and further work has been done on these stages.

### **3) SOFTWARE USED**

The prime software used in the project is MATLAB, which stands for Matrix Laboratory. MATLAB is a numerical handling software and proprietary coding language established by MathWorks. MATLAB authorizes matrix manipulations, graphical operations and data, algorithms, construction of user interfaces. Also, it is easy to interface with codes composed in other languages. These languages include C, Python, C++, Fortan, C#, and Java. [3].

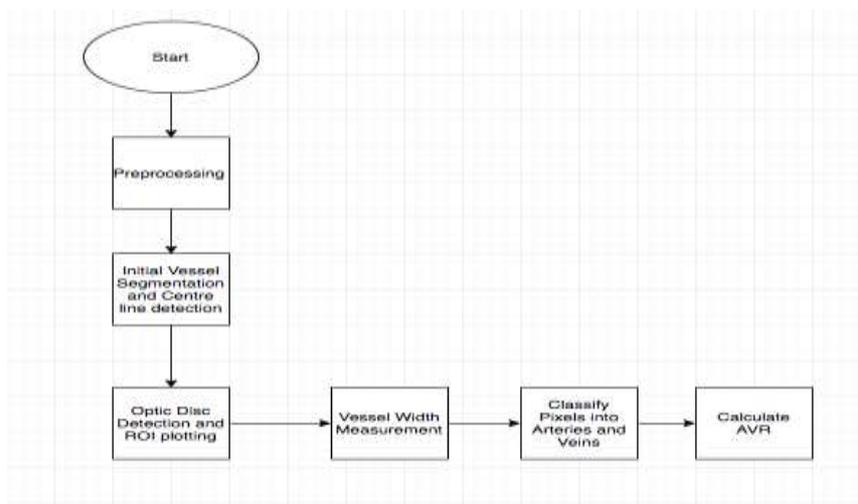
The reason for using MATLAB was related mostly to the functionality offered by this software. The project comprises of image processing and machine learning. MATLAB has toolboxes which not only support these two domains but also increase the ease with which one can write the programs.

On a comparative study between MATLAB and other prospective platforms like OpenCV, Tensorflow etc. we found that MATLAB was more user friendly and also had a greater support community and documentation.

We used Python for the machine learning part of our project. This decision was taken because python offers a variety of machine learning based libraries and also has a great support for machine learning tools. This support includes software updates, query resolution on online support communities and plenty of documentation available.

### **4) METHODOLOGY AND RESULTS**

The methodology of this project involves six core stages namely: Preprocessing, Vessel Segmentation, ROI detection, Vessel Width Measurement, Classification of vessels, and finally calculation of the ratio. They can be diagrammatically shown in fig A



(A)

This project has been divided into two parts, first one comprises of image processing techniques while the second part constitutes the machine learning algorithms. Amongst the six stages mentioned above, first four stages are completed using image processing techniques while the last two stages involve the usage of machine learning techniques.

### 1. PREPROCESSING

Preprocessing phase basically involves clearing any unwanted data like noise or blurs from an image to advance the properties of the fundus image and also enhance important, required features. In this stage, we passed our sample images (fig a) through a Gaussian Filter, which basically is a smooth blur filter. It blurs the background in the image thereby increasing focus on the object under study. The parameter used to determine the extent to which blurring occurs is called standard deviation. [4]

The standard deviation for our study was set at 100. Since a Gaussian filter acts as a normalizing filter, we were able to improve focus of the area near the Region of Interest (ROI), whilst removing the slow background variations (fig b). This blurred image was then subtracted from the actual fundus image so as to obtain the enhanced optic disc features (fig c).

Colored images increased the complexity of our calculations and hence we decided to convert colored images into grayscale images before performing our operations (fig d).



(a)



(b)



(c)



(d)

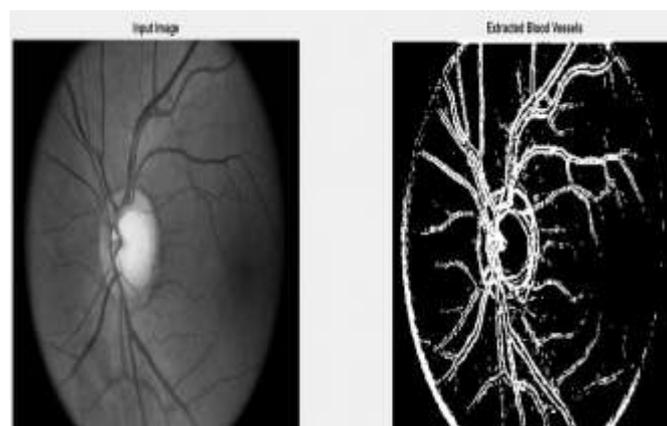
## 2. VESSEL SEGMENTATION

Vessel segmentation is basically applying operations on any fundus image to obtain required set of characteristics or features from the image while removing other redundant data. The image obtained after the preprocessing stage was segmented to obtain the required features, viz. vessel map in this case.

The method used for vessel segmentation in this work was by using Kirsch's templates. The Kirsch operator (Kirsch compass kernel) is a non-linear tool that is able to detect edges and that searches for the highest edge strength in some preset directions. A computer scientist Russell A. Kirsch invented this operator and hence, it's named after him. [5]

The image was inputted to a Kirsch's kernel in eight different orientations and the blood vessels were obtained by the operation. [6]

The results of the operation can be shown in (fig e)



(e)

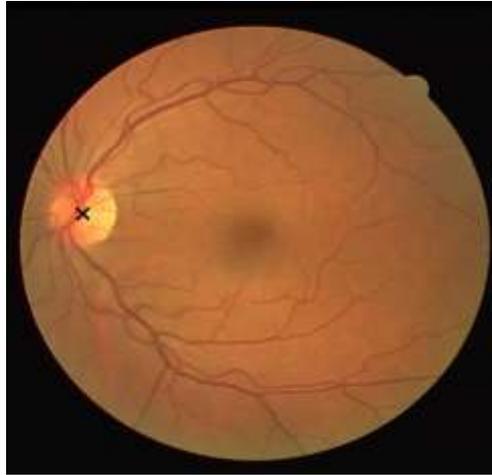
## 3. ROI PLOTTING

ROI signifies the region of interest. This step is necessary because the vessel map of the entire image is very cumbersome and redundant while evaluating AVR and also while classification. This is because there are lot of unnecessary crossovers

and complications in the region farther from the optic disc.

The technique used for plotting the ROI involved thresholding. [7]

The output of the stage is displayed in fig f.



(f)

#### 4. DETECTION OF OPTIC DISC

The method we have employed for the optic disc detection involves a series of steps starting with morphological operations and followed by Circular Hough Transform and GrowCut algorithm.

Mathematical morphology (MM) is a practice for the evaluation and handling of geometrical configurations, based on set theory, lattice theory, topology, and random functions. Mathematical morphology is usually used on digital images, however its use can be extended on graphs, solids, and several spatial configurations. [8] Circular Hough Transform is used to identify circles in a given image. [9]. This is property used to identify the Optic disc, since optic disc is also circular.

GrowCut is an algorithm that is based on interactive division. It uses Cellular Automaton [10] as an image model. Automata evolution simulates the division process. Each cell of the automata has a particular label (eg. binary segmentation - 'object', 'background' and 'empty'). Through this process, some cells seize their surrounding neighbors, changing their labels. The operator or user of the algorithm loosely draws some strokes inside the object of interest with an object brush, and for the outside of the object, a background brush is required. In basic cases, a few strokes are enough for the process of segmentation. [11]

The algorithm of optic disc detection was referenced from a previously done study in the study [12] and used with minor modifications. Fig g shows the result of this stage.

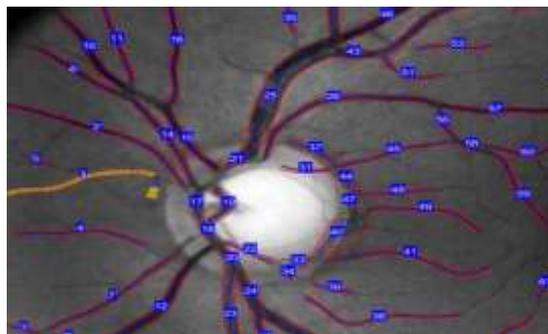


(g)

## 5. VESSEL WIDTH MEASUREMENT

For vessel width measurement, we used Automated Retinal Image Analyzer (ARIA) which is software (open source) devised for automatic recognition and computation of retinal factors and parameters. It was used by making certain modifications to the original code. [13]

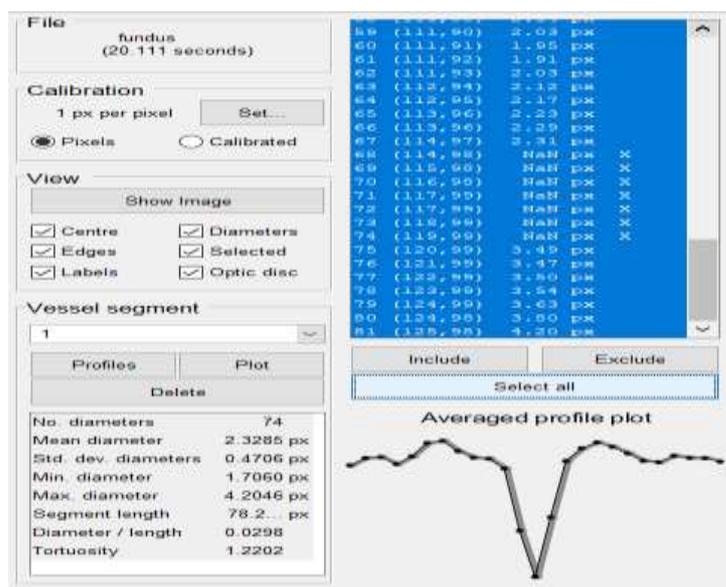
The ARIA software uses an array of steps which basically involve creating a vessel map and then identification of each part of that map (similar to a graph) by giving labels to each and every vessel in the fundus image. Fig h shows the labels which are provided by ARIA



(h)

ARIA works better than a normal width measuring algorithm because ARIA provides more insightful data on the other parameters of the vessel structure like number of vessels, average diameter etc.

The results of the stage are shown in fig i.



(i)

## 6. CLASSIFICATION INTO ARTERY-VEIN

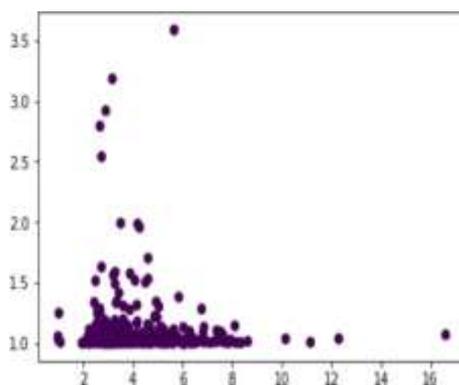
For the stage of classification of vessels, we propose to work on the following methods viz.

- 1) Unsupervised Classification using k-means clustering for this and the subsequent classification approaches, we first tabulated our results of 20 images into a dataset. This dataset included two parameters viz. mean vessel diameter and tortuosity for the unsupervised classification scheme.

We used k-means clustering for the classification because it's one of the simplest clustering algorithms to use.

The fundamental notion of k-means clustering is to assemble/cluster all the data points having similar characteristics together and thus forming clusters of different data points.

The results of our unsupervised dataset are shown below in fig j:



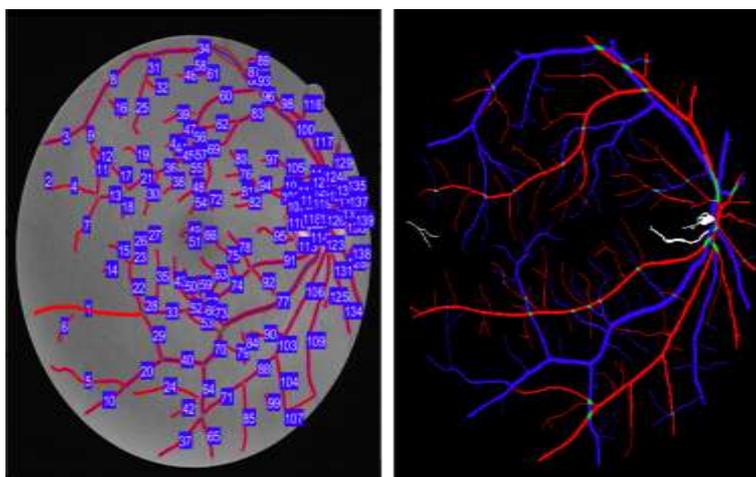
(j)

As is visible, the clustering algorithm yielded very poor results since there wasn't sufficient information to create a proper divide between the two vessel class arteries and veins.

## 2) Supervised Learning Using Naïve Bayes

We decided to opt for a better decision making algorithm to improve the accuracy of our computation and thus decided to use supervised learning. Before starting the process, we had to include "vessel types" into our dataset so that we can train the model using the training set. Thus, we manually labelled all the vessels as arteries and veins in our dataset.

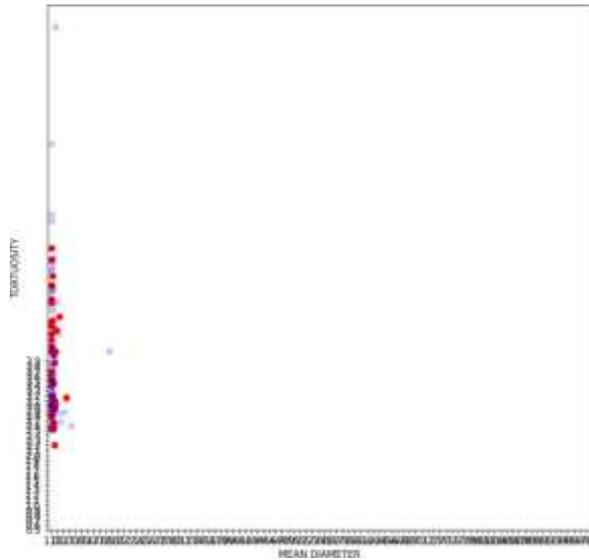
The reference which we used for this labelling procedure was obtained from RITE dataset [14]. The process of manual labelling has been described in fig k:



(k)

After this manual labelling was completed, the next step was preparation of the model. We used the Naïve Bayes algorithm. It is a classification algorithm and consists of essentially a collection of classifiers belonging to the Bayes Theorem family [15]. We used Naive Bayes and SVC (Support Vector Classifier) approach for classification purposes. The decision was taken because a probabilistic approach is better suited in this case as SVM relies on producing decision boundaries. In our case, there is no distinct decision boundary because we lack sufficient information, hence we have chosen a probabilistic model. For scalability, and when more features are available - SVM would be the appropriate choice for the classification model.

We trained the model for 10 images and three parameters viz. mean vessel diameters, tortuosity and vessel type and similarly tested it for 10 images. The average accuracy obtained in the entire process was close to 60% with slight variations in each individual image. One of the result is shown below in fig l:



(l)

Accuracy of the training and testing model for an example are shown in fig m:

```
print(accuracy_score(pred, Y_test))
0.5833333333333334

print(accuracy_score(pred, Y_test))
29
0.6206896551724138
```

(m)

### 7. CALCULATION OF AVR

For this step, we computed the mean width of all of the arteries and all of the veins in a particular fundus image and calculated the ratio. A sample of this stage is shown in fig n:

E	F	G
<b>Mean Diameter of Arteries</b>	<b>Mean Diameter of Veins</b>	<b>AVR</b>
3.516996154	3.530463235	0.996185463
<b>Number of Arteries</b>	<b>Number of Veins</b>	
52	68	

(n)

### 8. CONCLUSIONS

In this project we attempted to use image processing and machine learning technologies to automate and preempt the process of disease detection and curing. Both these technologies were fundamental to this project’s development and deployment.

Image processing techniques are very useful in a variety of applications. The crux of image processing technology is to use images of different varieties to extract relevant and important information from them. This information can be some kind of pattern, some hidden or embedded information and at times even some trends. In this project, it was highlighting the blood vessels in a fundus image to make the further analysis simpler. Without the usage of image processing techniques, the accuracy and efficiency of the project would have significantly dipped and would have made analysis harder.

The other technology which was employed was machine learning. The fundamental idea of machine learning is to use lots of historic data in order to improve decision making of any machine learning model. Machine learning models are used to automate simple and at times very complex tasks in order to reduce the response time and error rate of systems without any need for human interference. In this study, machine learning played a huge role in shaping up the classification process and finding the AVR. Machine learning models take decision by using data fed to them to identify different trends.

The machine learning model in this project used manually marked dataset to find out the nature of arteries & veins and then classify blood vessels into arteries and veins in the testing dataset.

This project is a great step towards the advancement of medical technology. This method of disease detection is a low cost system and significantly faster and better than any other existing system. Secondly, if it's deployed across the country, it can help poor people in getting prompt checkups at very subsidized rates. It can significantly improve the life expectancy of every citizen since disease prediction can be made months or even years before the onslaught of symptoms. Lastly, due to digitization of health records of patients due to this technology, interlinking of different cases and streamlining healthcare systems will become a lot easier.

To sum it up, this project is a step ahead in streamlining the healthcare industry.

## REFERENCES

- [1] J.J. Staal, M.D. Abramoff, M. Niemeijer, M.A. Viergever, B. van Ginneken, "Ridge based vessel segmentation in color images of the retina", IEEE Transactions on Medical Imaging, 2004, vol. 23, pp. 501-509.
- [2] Meindert Niemeijer\*, Xiayu Xu, Alina V. Dumitrescu, Priya Gupta, Bram van Ginneken, Member, IEEE, James C. Folk, and Michael D. Abrámoff, Senior Member, IEEE "Automated Measurement of the Arteriolar-to-Venular Width Ratio in Digital Color Fundus Photographs" IEEE TRANSACTIONS ON MEDICAL IMAGING, VOL. 30, NO. 11, NOVEMBER 2011.
- [3] Jana Pócsová, Andrea Mojžišová, Matej Mikulšzky, "Matlab in engineering education", 2018 19th International Carpathian Control Conference (ICCC), 2018, pp. 532 - 535.
- [4] E. S. Gedraite and M. Hadad, "Investigation on the effect of a gaussian blur in image filtering and segmentation," in ELMAR Proceedings, 2011.
- [5] G. Ping, S. Xiangju, W. Wei, "Moving object detection based on kirsch operator combined with Optical Flow", Image Analysis and Signal Processing (IASP) 2010 International Conference, pp. 620-624, 2010.
- [6] Athi (2011). Retinal Blood Vessel Extraction (<https://in.mathworks.com/matlabcentral/fileexchange/24990-retinal-blood-vessel-extraction>), MATLAB Central File Exchange. Retrieved August 18, 2018.
- [7] Saied Aboul-Dahab (2015) Retina region of interest. (<https://in.mathworks.com/matlabcentral/answers/179507-retina-region-of-interest>) MATLAB Central File Exchange. Retrieved September, 2018.
- [8] Dong Wei, Wang Tong, "Research on Edge Detection based on Mathematical Morphology Algorithm", International Conference on Optoelectronics and Image Processing, vol. 2, pp. 211-213, 2010.
- [9] M. Nosrati, R. Karimi, "Detection of circular shapes from impulse noisy images using median and laplacian filter and circular hough transform", Electrical Engineering Computing Science and Automatic Control (CCE) 2011 8th International Conference on, pp. 1-5, 2011.
- [10] Akane Kawaharada, "Fractal patterns created by Ulam's cellular automaton", Proceedings of International Workshop on Applications and Fundamentals of Cellular Automata, pp. 484-486, 2014.
- [11] M. Abdullah, M. M. Fraz, "Application of grow cut algorithm for localization and extraction of optic disc

- in retinal images", Proc. 12th Int. Conf. High-Capacity Opt. Netw. Enabling/Emerg. Technol. (HONET), pp. 1-5, Dec. 2015.
- [12]M. Abdullah, M. M. Fraz, S. A. Barman, "Localization and segmentation of optic disc in retinal images using circular Hough transform and grow-cut algorithm", PeerJ, vol. 4, pp. e2003, May 2016.
- [13]Bankhead P, Scholfield CN, McGeown JG, Curtis TM (2012) Fast Retinal Vessel Detection and Measurement Using Wavelets and Edge Location Refinement. PLoS ONE 7(3): e32435. doi:10.1371/journal.pone.0032435
- [14]Hu Q, Abramoff MD, Garvin MK. Automated separation of binary overlapping trees in low-contrast color retinal images. Med Image Comput Comput Assist Interv. 2013;16(Pt 2):436-43. PubMed PMID: 24579170
- [15]X. Fang, "Inference -Based Naïve Bayes: Turning Naïve Bayes Cost-Sensitive[J]", Knowledge & Data Engineering IEEE Transactions on, vol. 25, no. 10, pp. 2302-2313, 2013.