

## **CLASSIFICATION OF HEALTHY AND DIABETIC PATIENT BASED UPON EYE'S RETINA HEALTH CONDITIONS USING MACHINE LEARNING**

M. Nadeem<sup>1</sup> and Z. Suhail<sup>1</sup>

<sup>1</sup>Department of Computer Science, FCIT, University of the Punjab Lahore Pakistan

\*Corresponding author: [Muhammad.Nadeem@pucit.edu.pk](mailto:Muhammad.Nadeem@pucit.edu.pk)

**ABSTRACT:** Diabetes has become a global and common disease. There are numerous effects of this disease. Its effect can lead to complete blindness if proper medical precautions are not taken timely. The objective of this research is to classify the patient whether the patient is diabetic or not based upon the training of the model by binary classification of Diabetic Retinopathy (DR). We used available dataset that consists of many eye retina images. A hybrid methodology used in this research that consists of histograms and extraction of first and second order features. We investigated a lot of difference between the histograms of normal retina of eye than the affected eyes. The researchers generated histograms of the eye retina images, fetch its first order (mean, median, variance etc.) and second order features (energy, entropy, homogeneity and contrast etc.). We achieved 0.814 precision, 0.821 accuracy, 0.821 and 0.817 f1 score using decision tree. Hence it can be used effectively for identifying DR and early detection of diabetic patient.

**Key words:** Diabetic Retinopathy, Histogram, Retina, Binary Classification, Medical image Analysis.

(Received 16.04.2023

Accepted 28.06.2023)

### **INTRODUCTION**

It is a common and famous proverb that prevention is better than cure. Therefore, in order to save the eyes of anyone else not only locally, nationally but also at internationally, we conducted research to identify and classify a normal human being as a normal person or diabetic patient based on the health conditions of the retina of eyes using machine learning. The importance of Machine Learning (ML) is increasing day by day with its wide range applications in every field of life especially in medical sciences. The use of ML to analyze different types of infections is highlighted by (Ibrahim and Abdulazeez, 2021). The research conducted (Taylor *et al.*, 1997), indicates that diabetic eyes have been studied and analyzed. The authors also linked the use of insulin with these parameters. They mentioned the outcomes of this research as the developing countries population can be suffer in Blinding Retinopathy (BR) diabetic disease due to adopting westernized food, gain of weights and long lives. As per their findings, 63% patients of the study in United Kingdom found to have decrease vision due to diabetic disease. The importance of early detection and the prevention from diabetic can lead to complete blindness in the research work of (Priya *et al.*, 2013). The researchers concluded that how a diabetic patient can go to complete blindness due to the effect of diabetic disease. They described the types of DR as proliferative-less, pre-proliferative and proliferative DR. According to their outcomes, initially DR indications are unknown but with the passage of time, it may lead to low vision to complete blindness. The only possible

solution to protect for DR and complete blindness is that the patients or individuals should check their sugar level regularly and keep focus about the sight issue if any. According to the research of (Rajalakshmi *et al.*, 2018), they mentioned the importance of using Artificial Intelligence (AI) for automatic detection of DR and Sight-Threatening DR (STDR). They used the smartphone-based device fundus photography with the ophthalmologist rating. The researchers got the images of 301 patients having diabetes II for DR. So they concluded that smart phone based AI detection of diabetic patients can be used as a preliminary tool to check the DR presence in most of the diabetic patients.

The authors (Bhatia *et al.*, 2016) put forward techniques for the development of a system that can detect the DR in diabetic patients automatically. Their main contribution is to find the ways to early detect the DR patient using AI based systems. The research work done (Chetoui *et al.*, 2018.) for DR identification with ML and textual feature was performed to secure complete blindness. As far as their contribution is concerned, they apply different textual features such as Local Energy-based Shape Histogram (LESH) and Local Ternary Pattern (LTP) etc. The researchers also used support vector machine algorithm of AI to classify the obtained histogram. The authors (Ahmed, 2002) conducted research for providing the history of diabetes mellitus. The word mellitus means honey sweet. As per their findings, before 3000 years, the ancient Egyptians defined diabetes mellitus. This term was initially created by Aratus. After that mellitus was enhanced by Aratus of Cappadocia firstly used the label "diabetes" (81-133AD).

Later, the word mellitus diabetes mellitus was defined by Thomas Willis in 1675 with the discovery of sweetness in urine and blood of the patients. It is due to his efforts that the presence of extra sugar in urine and blood was found in Dobson in 1776 due to this sweetness. Then the use and key role of in liver for cyclogenesis in 1857 in France. After that that the Mering and Minkowski found the use of pancreas to control diabetes in 1889. After all in 1921, these findings become the foundation for the discovery of insulin in clinics by the efforts of and Best. Later on many research based reports given the management of dietary and its chronic complications. As per the research conducted (Li *et al.*, 2019), they suggested early screening of the diabetic patient to avoid blindness due to DR by achieving high specificity and sensitivity. High means greater than 90%. As per the outcomes of this research, the performance of deep learning models was not good because of certain restrictions. One of the main limitation was the availability of the datasets of fundus of eyes from different patients. To assess the deep learning models, images they got 13,673 images after 9,598 different patients. Then these fundus eye images separated into different 6 types with respect to 7 graders as per the level of DR and the quality of the images. Resultantly, 757 images found to be DR and then these interpreted to 4 categories of DR based cuts. The authors applied best deep learning algorithms such as semantic segmentation, object detection and the classification of an image. While, they got 0.8284 for DR classification accuracy for the classification of DR but these have not brought best results for lesion segmentation and its detection. It clearly indicates that detection and segmentation of lesion is a tough task. The research work (Suhail. and Zwiggelhaarb, 2020) suggested afresh way of segmentation about the mammographic images masses. Resultantly, the outcomes of the segmentation highlighted better and enhanced area of the mammogram masses that shows the accurate shapes of irregularities. They finally tested their proposed method on 233 malignant and 233 benign irregularities. The authors (Sarki *et al.*,2020) provided early detection of DR or diabetic eye related material for patients, healthcare professionals and research communities. (Abinaya *et al.*, 2020) also studied a methodical analysis of the published literature using ML techniques. The researchers (Aladawi *et al.*, 2019) studied and evaluated already researched and published methods, specificity, classification and datasets for the automated detection of DR. The recent research studies (Mayya *et al.*, 2021), (Atwany *et al.*, 2022), (Kumar *et al.*, 2022) have been conducted survey studies for the early detection of DR and use of artificial intelligence.

## MATERIAL AND METHODS

**Dataset:** We used already available dataset (Aladawi *et al.*, 2019). This dataset is known as Dataset for Diabetic Retinopathy (DDR). They used this data set for the classification of Diabetic Retinopathy (DR), Lesion Segmentation and Lesion Detection. However, it may be mentioned here that the authors have performed their research using multilevel classification. Following table shows the multilevel classification and number of images for DR grading:

**Table 1: Multilevel Classification of Images.**

S.#	Multilevel Classification	Images
1	0. No DR	1,587
2	1. Mild Nonproliferative DR	1,161
3	2.Moderate Nonproliferative DR	1,452
4	3.Severe Nonproliferative DR	194
5	4. Proliferative	444
6	5. Ungradable.	739
<b>Total images</b>		<b>5,567</b>

**Table 2: Division of Images.**

S.#	Binary Classification	Images
1	0. No Diabetic Patient	1,587
2	1. Diabetic Patient	3,980
<b>Total images</b>		<b>5,567</b>

This dataset consists of DR grading, lesion detection and lesion segmentation. But we used only DR grading for binary classification. It may also necessary to clear that this dataset consist of training, testing and validation images for DR grading, lesion detection and lesion segmentation. But we used only DR grading images for binary classification. The authors of this research study split this DR grading images after extracting first and second order features into training and testing data. The training data is 80% and testing data is 20%.

The researchers developed a training data set as per above multilevel. We have also used same data set for training. But our purpose is different than the research work of (Aladawi *et al.*, 2019). We would like to classify healthy and diabetic patient depending upon the eyes condition of the patient. So, in this research, we used the following labels for binary classification:

- 0. No Diabetic Patient
- 1. Diabetic Patient.

However, the number of images used in this research study are given in the following table:

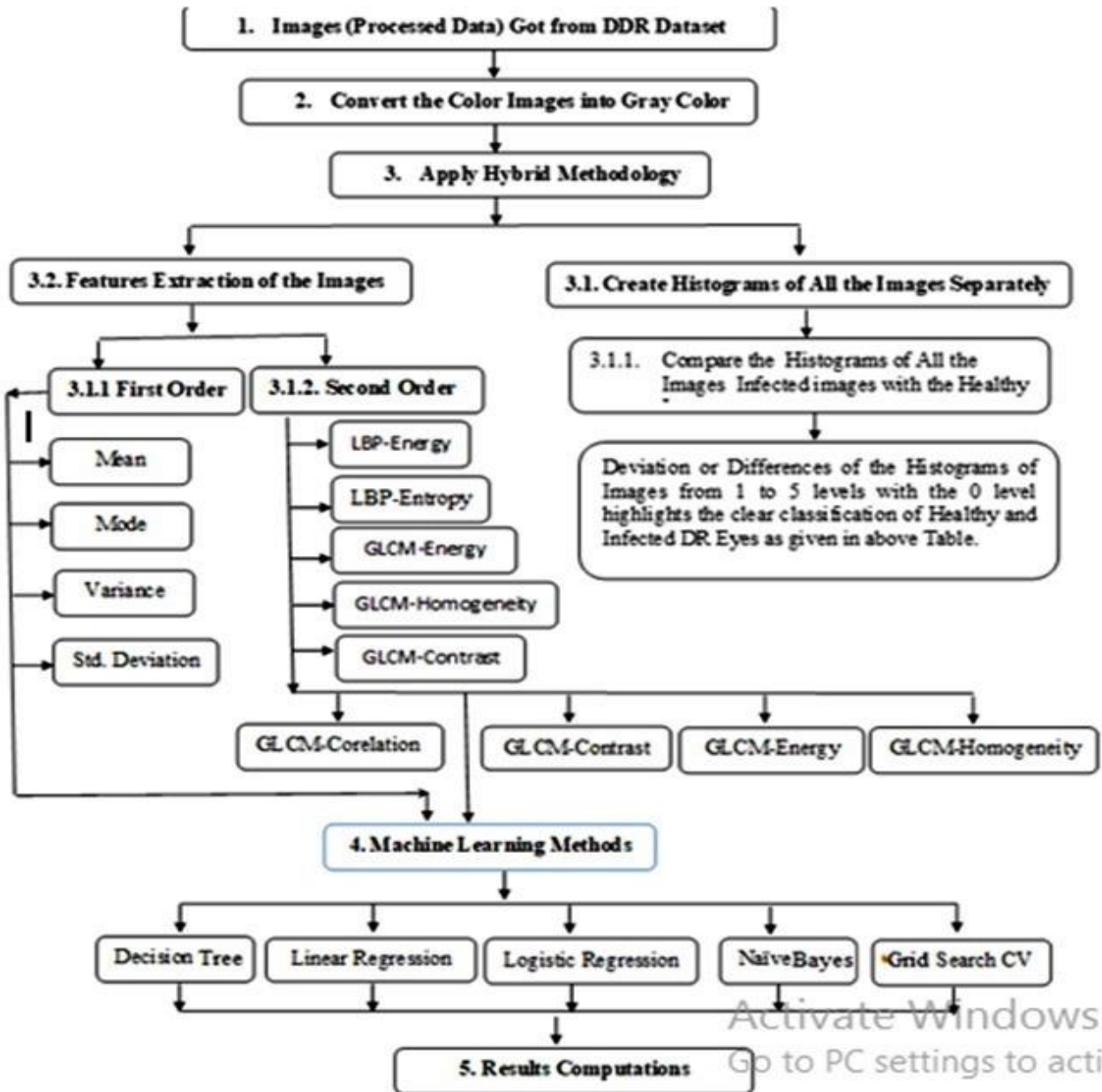
**Table 3: Images Used in Underlying Research Study.**

S.#	Multilevel Classification	Images
1	0. No DR	500
2	1. Mild Nonproliferative DR	500
3	2. Moderate Nonproliferative DR	500
4	3. Severe Nonproliferative DR	194
5	4. Proliferative	444
6	5. Ungradable.	500
<b>Total images</b>		<b>2628</b>

**Methodology:** As we have mentioned in data set section that we are working for binary classification whether a patient is a diabetic patient or not depending upon the DR

condition or overall eye condition of the patient. In this way, we will annotate 0 level of (Aladawi *et al.*, 2019) as 0 and 1 to 5 levels of the (Aladawi *et al.*, 2019) as 1. We further like to mention here that our problem is binary problem but not a multilevel problem. So, as per the data set developed by (Aladawi *et al.*, 2019), we reused this data set for the identification and classification of a patient into non-diabetic patient or diabetic patient. We used hybrid way for the classification and identification of early diabetic patient. This hybrid method consists of the following:

1. Pictorial and Histogram Based Comparison of Images
2. Extraction of First Order and Second Order Features of the Images



**Figure. 1: Research methodology.**

Table 4: Pictorial Comparisons of Multilabel Eyes Images (0 to 5 Levels).

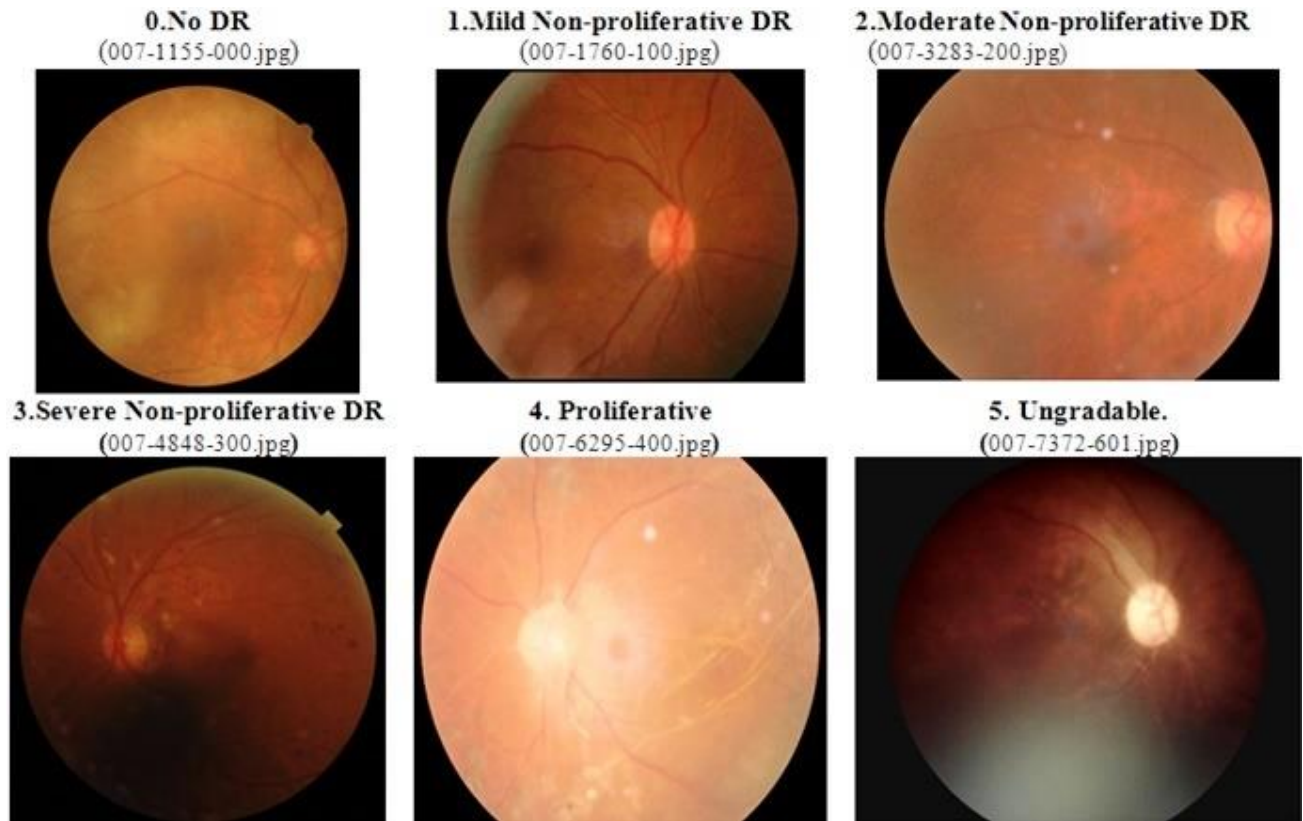
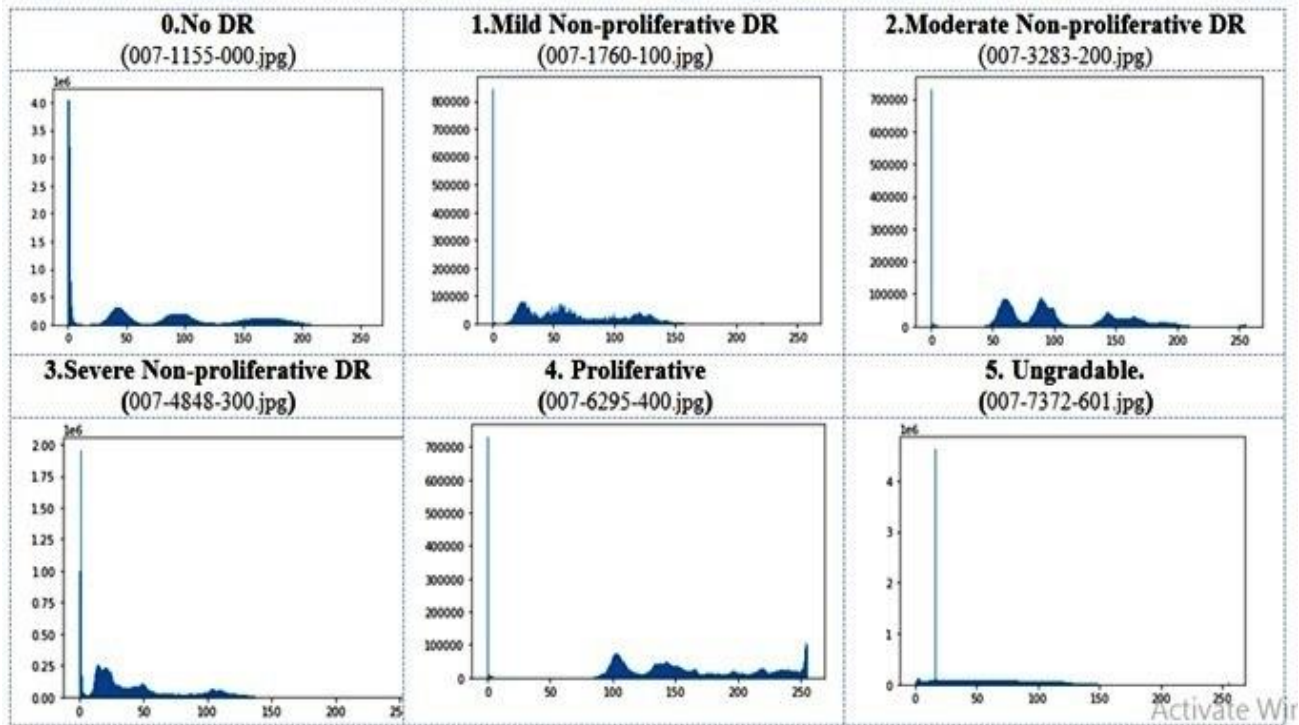


Table 5: Histograms Based Comparisons of Multilabel Eyes Images (0 to 5 Levels).



**Pre-processing:** We used the data set created by (Aladawi *et al.*, 2019). So we didn't need to do pre-processing so much. However, these images were color

images. We only convert each image into gray image before getting features from these images.

**Features extraction:**

**Table 6: First Order Features of Images from 0 to 5 Levels.**

1 <sup>st</sup> Order Image Features													
S.#		Images											
		0 Level		1 Level		2 Level		3 Level		4 Level		5 Level	
		Min Val.	Max Val.	Min Val.	Max Val.	Min Val.	Max Val.	Min Val.	Max Val.	Min Val.	Max Val.	Min Val.	Max Val.
1	Mean	15	98	12.4	112.6	9.00	101.3	20.69	162.8	11.11	169.0	4.54	146.84
2	Median	0	145	0	146	0	105	0	196	0	208	0	112
3	Variance	262	5604	300.2	9425.0	162.4	9509.1	454.4	7781.3	143.62	10283	49.5	15179
4	Std. Deviation	16	75	17.3	97.08	12.7	97.51	21.31	88.21	11.98	101.40	7.03	123.20

**Table 7: Second Order Features of Images from 0 to 5 Levels.**

2 <sup>nd</sup> Order Image Features													
S.#		Images											
		0 Level		1 Level		2 Level		3 Level		4 Level		5 Level	
		Min Val.	Max Val.	Min Val.	Max Val.	Min Val.	Max Val.	Min Val.	Max Val.	Min Val.	Max Val.	Min Val.	Max Val.
1	LBPEnergy	0.20	0.57	0.20	0.60	0.20	0.60	0.172	0.588	0.197	0.640	0.198	0.803
2	LBPEntropy	1.5	2.7	1.25	2.7	1.24	2.65	1.37	2.78	1.158	2.676	0.690	2.673
3	GLCM_Energy	0.14	0.62	0.11	0.62	0.127	0.56	0.0789	0.567	0.080	0.620	0.106	0.628
4	GLCM_Homogeneity	2.3	4.6	2.27	4.74	2.37	4.35	2.498	4.496	2.244	4.710	1.550	4.715
5	GLCM_Contrast	0.49	0.9	0.36	0.91	0.516	0.912	0.391	0.859	0.391	0.924	0.420	0.961
6	GLCM_Corre	0.01	0.39	0.013	0.384	0.016	0.32	0.006	0.321	0.006	0.384	0.011	0.395
7	GLCM_Dissimilarity	0.81	1	0.82	53.81	0.603	61.8	0.680	56.09	0.843	59.868	0.194	168.42
8	GLCM_Entropy	0.98	0.99	0.98	0.99	0.98	0.99	0.981	0.999	0.959	0.999	0.933	0.999
9	GLCM_ASM	0.28	5.0	0.227	4.15	0.28	3.88	0.326	2.938	0.216	4.184	0.105	6.710

Here, GLCM stands for Gray Level Co-occurrence matrix, LBP stands for Local Binary Pattern.

**RESULTS AND DISCUSSIONS**

Results of the Experiments using Python Libraries and Weka tool are given below:

According to Tables 4 and 5, pictures and histograms representation of each of the images representing clear-cut difference among the multilabel images of the dataset. It also further elaborated the effective use of these images for binary classification of a normal person into a diabetic or non-diabetic patient. As per Table 6, first order features such as mean, median, variance, standard deviations are extracted. These results clearly show the differences of the images with respect to image level for the clear cut identification of the separation of these images. Similarly, second order features have also been extracted and provided in Table 7. We extracted GLCM and LBP total nine features named as LBPEnergy, LBPEntropy, GLCM\_Energy, LCM\_Homogeneity, GLCM\_Contrast, GLCM\_Corre, GLCM\_Dissimilarity, LCM\_Entropy and GLCM\_ASM. The extracted values of second order features clearly

show the differences of different level eye images used in dataset. We used five binary classification techniques named as decision tree classifier, linear regression, logistic regression, naïve Bayes and grid search CV. We used four model evaluation measures such as precision, accuracy, recall and f1 score. According to Table 8 results, best results are indicated by bold face text. This tables depicts that decision tree classifier performed very well for the classification of diabetic patients or non-diabetic patients. Its reason is that maximum precision is 81%, accuracy is 82%, recall is also 82% and f1 score is 81%. So 81% precision indicates that our trained model returned or classified 81% correct results for the binary classification of diabetic and non-diabetic patients. However 82% accuracy shows that it is very much close results out of 100%. Similarly, 82% result of recall represents that this model returned 82% relevant instances of the images. In the same way, 81% f1 score indicates the harmonic mean of precision and recall that also shows excellent performance of the model evaluations and measurements. Resultantly, most of the



evaluations scores are more than or equal to 70% using all five classification techniques.

**Table 8: Experimental Results.**

S.#	Technique	Measure	Score
1	Decision Tree Classifier	Precision	<b>0.814</b>
		Accuracy	<b>0.821</b>
		Recall	<b>0.821</b>
		F1 Score	<b>0.817</b>
2	Linear Regression	Precision	0.706
		Accuracy	<b>0.797</b>
		Recall	<b>0.797</b>
		F1 Score	0.712
3	Logistic Regression	Precision	0.706
		Accuracy	<b>0.797</b>
		Recall	<b>0.797</b>
		F1 Score	0.712
4	Naïve Bayes	Precision	<b>0.714</b>
		Accuracy	0.656
		Recall	0.656
		F1 Score	0.680
5	Grid Search CV	fit_time	0.058824
		score_time	0.000900
		test_score	0.821088
		train_score	0.821657

**Results of the Experiments using Weka 3.8.5:**

**Table 9: Canopy clustering Results.**

Number of canopies (cluster centers) found: 8

S.#	Cluster Centroid	Instances and %age	Binary Clustering with %age	
1	0	43 ( 45%)	0	45%
2	1	2 ( 2%)	1	52%
3	2	17 ( 18%)		
4	3	12 ( 13%)		
5	4	2 ( 2%)		
6	5	3 ( 3%)		
7	6	7 ( 7%)		
8	7	9 ( 9%)		

**K-Means:** Number of iterations: 11

Within cluster sum of squared errors: 169.74243265990776

Initial starting points (random):

**Cluster 0:** 29,007-1978-100.jpg,51.460312,0.2520.584057,50.205419,4665559073,108091000000000,0,0,0.349056,2.143444,0.453696,2.946492,0.728186,0.20584,2.843438,0.999436,0.801352,1

**Cluster 1:** 1,007-1257-000.jpg,46.706973,1,1231.874154,35.098065,-293240402.1,24919900000000,0,0,0.224552,2.587861,0.175759,3.827302,0.488743,0.030891,9.775557,0.996031,2.023168,0

Missing values globally replaced with mean/mode

**Table 10: Final cluster centroids.**

S.#	Attribute	Cluster #	
1	Full Data	0	1
2	100%	<b>40%</b>	<b>60%</b>

**DISCUSSIONS**

As per Table 9 results using canopy clustering technique with Weka, the results shows that 97% performance found successful using Weka tool. Out of 97%, 45% results indicates non-diabetic patient and 52% results shows diabetic patients with different instances and percentages. According to Table 10, finally two clusters found with 40% non-diabetic patients in one cluster and 60% diabetic patients in second cluster. These results also show the best performance of the research study conducted for the classification of a normal patient eye into diabetic and non-diabetic patients.

**Conclusion:** In this study, we proposed the idea of the early identification, detection and classification of diabetic and non-diabetic patient using hybrid approach. We created and used histograms of 0-label images (Non-diabetic) along with all other classes or levels of images from 1 to 5 as well. The above tables 4 and 5 shows the pictorial differences of the diabetic and non-diabetic patient. The images and their respective histograms both are highlighting the clear classification of diabetic and non-diabetic patient based upon the DR of eyes retina. While, the experimental results performed after extracted features from these images. The results have been given in table 8 using machine learning methods. Similarly the results generated using Weka are given in table 9 and table 10. As per the results created using decision tree, precision is 0.814, accuracy is 0.821, recall is 0.821 and f1 score is 0.817. Hence it can be used effectively for identifying DR and early detection of diabetic patient. However, further experiments also performed with Weka for further analyze the outcomes. As per the clustering results of table 10, Weka identified 40% diabetic patients and 60% non-diabetic patients. It means we can also use clustering technique efficiently for the binary classification of the images.

**Future work and limitations:** The authors would like to create new dataset of the retina images of the diabetic patients along with the images of non-diabetic patients for the further improvements of the early identification and classification of diabetic and non-diabetic patients. One of the main limitation of this study is to collect a large dataset of the retina images of diabetic and non-diabetic patients in time.

## REFERENCES

- Abinaya K., P.S. Meenalaxmi, T.A. Vignesh, D. Kuralamuthan and R. Poonguzhali, (2020). Detection and Classification Of Diabetic Retinopathy Using Machine Learning – A Survey. *Int. J. of analytical and experimental modal analysis*. XII(VI): 2128-2134.
- Ahmed A. M., (2002). History of Diabetes Mellitus. *Saudi Med J*. 23(4): 373-378.
- Aladawi W., C. Jayakumari, E. P. Sumesh and Vidhyalavanya, (2019). Recent Innovations in Automated Detection and Classification of Diabetic Retinopathy. *Int. J. of Innovative Technology and Exploring Engi*. 8(10): 1997-2004.
- Atwany M. Z., A. H. Sahyoun, and M. Yaqub, (2022). Deep Learning Techniques for Diabetic Retinopathy Classification: A Survey”. *IEEE-Access*. 10: 28642-28655.
- Bhatia K., S. Arora and R. Tomar, (2016). Diagnosis of Diabetic Retinopathy Using Machine Learning Classification Algorithm. *2nd Int. Conf. on Next Gen. Computing Tech. Dehradun, India, IEEE*. 347-351.
- Chetoui M., M. A. Akhloufi, M. Kardouchi, (2018). Diabetic Retinopathy Detection Using Machine Learning and Texture Features. *IEEE Canadian Conf. on Electrical & Computer Eng.* 1-4.
- Ibrahim M. I., A. M. Abdulazeez, (2021). The Role of Machine Learning Algorithms for Diagnosing Diseases. *J. of Applied Sci. and Tech. Trends*. 02 (01): 10 – 19.
- Kumar Y , A. Koul, R. Singla and M. F.Ijaz, (2022). Artificial intelligence in Disease Diagnosis: A Systematic Literature Review, Synthesizing Framework And Future Research Agenda. *J. of Ambient Intelligence and Humanized Computing*. Springer. 1-28.
- Li T., Y. Gao, K. Wang, S. Guo, H. Liu and H. Kang, (2019). Diagnostic Assessment of Deep Learning Algorithms for Diabetic Retinopathy Screening. *Information Sci., Elsevier*. 1-25.
- Mayya V., S. S.Kamath and U. Kulkarni, (2021). Automated microaneurysms detection for early diagnosis of retinopathy : A comprehensive Review. *J. of Computer Methods and Programs in Biomedicine Update, Elsevier*. 1: 1-15.
- Priya P.V., A. S. Rao, J. V. C. Sharma, (2013). Diabetic Retinopathy - Can Lead To Complete Blindness. *Int. J. of Sci. Inventions Today*. 2(4): 254-265.
- Rajalakshmi R., R. Subashini, R. M. Anjana And V. Mohan, (2018). Automated diabetic retinopathy detection in smartphone-based fundus photography using artificial intelligence. *Eye, The Royal College of ophthalmologists*. 32: 1138–1144.
- Sarki R., K. Ahmed, H. Wang and Y. Zhang, (2020). Automatic Detection of Diabetic Eye Disease Through Deep Learning Using Fundus Images: A Survey. *IEEE*. 8: 151133-151149.
- Suhail Z. and R. Zwigglelaarb, (2020). Histogram-Based Approach for Mass Segmentation in Mammograms. *Fifteen Int. Workshop on Breast Imaging*. 1-8.
- Taylor R. H., H. S. Jones, P. M. Dodson, A. P. Hamilton and E. E. Kritzinger, (1997). Diabetic Eye Disease: A Natural History Study. *Eye, Royal College of Ophthalmologists*. 11: 547-553.