International Journal of Scientific Research in Science and Technology
Print ISSN: 2395-6011 | Online ISSN: 2395-602X (www.ijsrst.com)
doi : https://doi.org/10.32628/IJSRST52310275

Diabetic Retinopathy Detection Using Optimized Random Under Sampling (ORUS) Algorithm

Kanna Swarnalatha*[1], Dr. I. Kullayamma[2]

*[1]M. Tech, Department of Electronics and Communication Engineering, Sri Venkateswara University College of Engineering, Tirupati, Andhra Pradesh, India

[2]Professor, Department of Electronics and Communication Engineering, Sri Venkateswara University College of Engineering, Tirupati, Andhra Pradesh, India

ARTICLEINFO

Article History:

Accepted: 01 April 2023
Published: 24 April 2023

Publication Issue
Volume 10, Issue 2
March-April-2023

Page Number
495-502

ABSTRACT

This paper provides the detection of diabetic retinopathy (DR) from the different features, the dataset is a combination of both proliferated and non-proliferated diabetic retinopathy. The people who at the age of above 60 and having the diabetes will face this diabetic retinopathy problem. This is mainly occurred due to sugar builds up on the blood vessels. This is because of pancreas doesn't produce proper level of insulin. The early detection of diabetic retinopathy plays very important role, with this early detection it can be able to reduce the level of severity, after that by taking proper medications and all it can be able to control. In this kind of detection process every feature is very important, with these features it can be able to classify whether the person is diabetic retinopathy prone or not. To detect the diabetic retinopathy many algorithms are used. To increase the sensitivity and to reduce error rate, by using optimized RUS boost algorithm.

Keywords: Diabetic retinopathy, insulin, pancreas, optimized RUS boost algorithm.

I. INTRODUCTION

Diabetic retinopathy (DR) is a medical illness due injury to the blood vessels of the light sensitive at the retina capillary vessel walls which is developed under macula. Depending on the severity of the disease, it can eventually lead to blindness and other visual issues. Despite the fact that medication is accessible, it is predicted that many people getting blind every day [1].

Abnormal blood vessels can develop on the surface of the retina or on the optic nerve. This process is called "neovascularization".

Neovascularization occurs when the normal blood vessels in the retina become very damaged from

diabetes and close off, simulating the growth of new abnormal blood vessels [6].

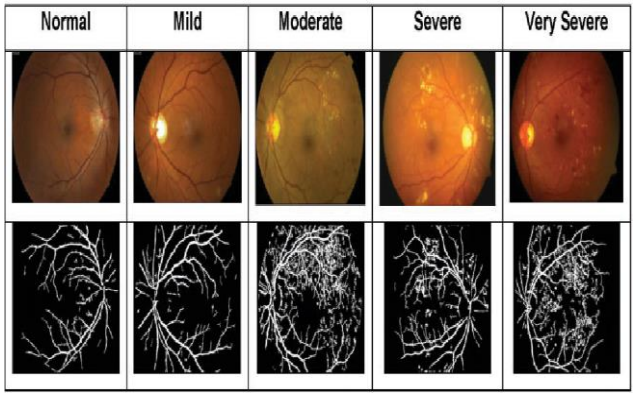

Figure 1(b): different severity levels of diabetic retinopathy

Courtesy: Google

The figure 1(b) shows that normal retina getting turned into diabetic retinopathy and it shows the nerves inside the eye getting damaged as the severity levels get increases, this leads to irreversible blindness

The retina's blood vessels expand as a result of diabetic retinopathy, which causes fluid and blood leakage. It is among those with diabetes mellitus who have microvascular problems the most often. There seems to be two types of diabetic retinopathy: non-proliferative diabetic retinopathy (NDPR) and proliferative diabetic retinopathy (PDR). PDR is distinguished by lesions like microaneurysm's (MA's) and exudates, where NDPR is distinguished by neovascularization of weak blood arteries. DR can cause visual loss and is one of the most common causes of irreversible blind in humans [1]. The difference between proliferative and non-proliferative diabetic retinopathy is shown in figure 1(a).

The symptoms of diabetic retinopathy include polyuria, weakness, polydipsia, obesity, sudden weight loss, visual blurring etc. Diabetes is a metabolic disorder that annually results in million fatalities owing to numerous health complications [2].

To provide differentiation from NDPR to PDR some features are selected from the images of diabetic retinopathy. Although it is frequently thought that the data are imbalanced during the data analysis process, there are really many imbalanced data. Imbalanced are those that have a relatively small number of (minority)in one category and large no. of entries (majority) in other area of machine learning research is to how to increase the classification accuracy of unbalanced data. Since it is frequently the case that few data contain more significant information [3].

The different stages of PDR are shown in figure 1(c)

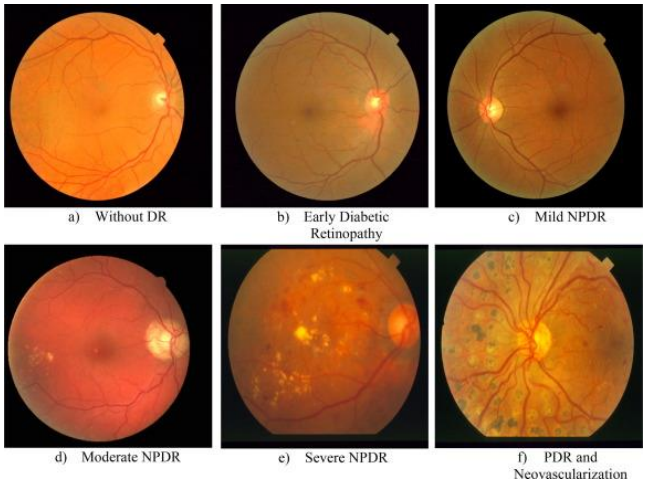

Fig 1(c): Different stages of proliferative diabetic retinopathy

Courtesy-

https://www.sciencedirect.com/science/article/pii/S1319157821001397

This paper provides the differentiation between the two machine learning algorithms called SVM (Support vector machine) and optimized RUS boost algorithm (Random under sampling).

Weiss et al. [4] is used to compare the metrics of oversampling, under sampling and cost sensitive learning while dealing with the data that have both an imbalanced class distribution and error costs.

Chawla et al. [5] gives the overall performance of a sampling approach to minimize misclassification probability and gives detailed evaluation of RUS boost

as a cost sensitive learning method which consists of comparison of different existing methodologies is left for the future work.

Sohini [6] provides the novel methodology that classifies the neovascularisations in the 1-optic disk (OD) diameter (NVD) and (NVS). It gives NVD and NVE detection methods and an important role in automated screening and prioritisation of patients with diabetic retinopathy.

Shailesh [7] demonstrates an enhanced approach for detecting diabetic retinopathy for extracting accurate and the number of microaneurysms from fundus images. It can be detected by using principal component analysis (PCA), contrast limited adaptive histogram (CLAHE), and other morphological process are used.

Sing et al. [8] used histogram equalization technique for low radiance images to clip the pixel values based on the threshold which is calculated by using average median value of the image to improve the normalization results. It provides the discrimination between the diabetic retinopathy and healthy controls (HC) by calculating from the optical coherence tomography angiography (OCTA) images from the 3*3 scans with different machine learning models.

Tang et. Al [9] used KNN for classifying haemorrhages candidate from diabetic retinopathy. For this the retinal images are divided into splat partitions. Splat means that from the retinal colour images are divided into non overlapping segments each segment covers the entire image. That each segment called splat, this contains pixels with same colour and spatial location. This splat will be very useful while defining the texture and shape of the segment information. These features are given to the classifier. This algorithm is having problem with the computational efficiency and generalisability problems.

Akram et.al [10] proposed the hybrid structure of gaussian mixture model (GMM) and SVM for diabetic retinopathy classification. The method is improved by adding the feature with shape, intensity and affected region.

II. EXISTING METHOD

The existing methodology uses the SVM classifier to classify the diabetic retinopathy problems.

SVM classifier:

SVM classifier is used to classify the both classification and regression problems. In this it uses the hyperplane to classify the different classes. Not only hyperplane there exists margins also which are parallel to the hyperplane and these two margins are separated with distance called marginal distance. This hyper plane divides the classes in a better way.

If higher marginal distance means that the model is most generalized model. This is all about the linearly separable case. On other hand2, in nonlinear separable case SVM uses SVM kernels. This SVM kernels tries to convert the low dimensional features into high dimensional features. The graph of SVM classifier is shown in figure 2(a), it shows that the data set consists of class 1 and class 2 data, by using hyperplane the classifier can be able to divide the two classes with the help of support vectors.

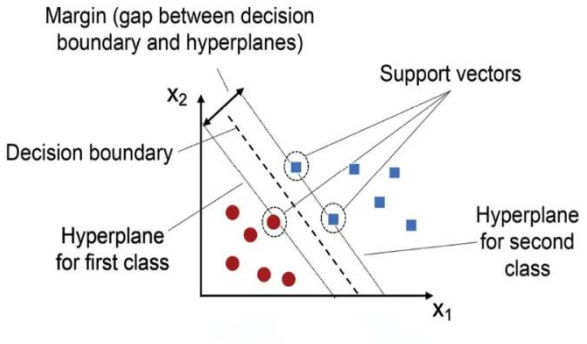

Fig 2 (a): SVM classifier Courtesy: google

In this existing methodology the total data set is divided into training data set and testing data set which is in the ratio of 70:30.

The existing methodology consists of high error rate, and low sensitivity, low F1 score to improve these metrics moving towards another methodology called optimized RUS boost ensemble technique.

III.PROPOSED METHOD

The dataset having different classes to classify the data of different classes we are using many classifiers. There exists the problem of misclassification to reduce the probability of misclassification we are moving towards the ensemble learning. Ensemble learning is the process of different models like SVM, random forest algorithm etc., are created and mixed to solve a specific problem.

Ensemble techniques are mainly used to improve the classification, prediction, decision making etc. There are different techniques like bagging, boosting, averaging, stacking etc. In ensemble learning there will be different classifiers and those are get trained by training data. Each classifier follows the different training data set. To maintain the diversity of the output then the base learners are following different algorithms. This is called heterogenous ensembles.

These base learners are also called weak learners. By combining the output of all weak learners and it builds the strong classifier. The predictive power accuracy of strong classifier is more and error rate, precision is less compared to weak classifier.

The proposed methodology follows optimized RUS boost algorithm which removes the problem of class imbalance. Many techniques have been used to remove the problem of class imbalance which includes data sampling and boosting

RUS boost algorithm combines the data sampling and boosting which gives the best simple and efficient method for classification of data. Along with the RUS boost algorithm there exists another one also called SMOTE (synthetic minority over sampling technique) boost algorithm.

In training data adding instances to minority class is called oversampling and removing of instances from training data is called under sampling.

In SMOTE boost algorithm random oversampling balances, a dataset by adding duplicating instances of minority class until getting required class ratio. Likewise random sampling balances the data set by removing the duplicate instances of minority class until to achieve desired class ratio.

There exist drawbacks regarding the under sampling is loss of information which comes with deleting the instances from training data. It has benefit that it will decrease time required to train the models therefore the size of training data set get reduced.

In case of oversampling leads to overfitting and increase time required to train the models. Then the size of training data gets increased. There exists the disadvantage of RUS boost is loss of information it can be overcome by combining with boosting.

The algorithm of RUS boost as shown in the given below.

Algorithm RUSBoost

Given: Set S of examples $(x_1, y_1), ..., (x_m, y_m)$ with minority class $y^r \in Y$, $|Y| = 2$

Weak learner, $WeakLearn$

Number of iterations, T

Desired percentage of total instances to be represented by the minority class, N

1 Initialize $D_1(i) = \frac{1}{m}$ for all i.

2 Do for $t = 1, 2, ..., T$

a Create temporary training dataset S'_t with distribution D'_t using random undersampling

b Call $WeakLearn$, providing it with examples S'_t and their weights D'_t.

c Get back a hypothesis $h_t : X \times Y \to [0, 1]$.
d Calculate the pseudo-loss (for S and D_t):

$$\epsilon_t = \sum_{(i,y):y_i \neq y} D_t(i)(1 - h_t(x_i, y_i) + h_t(x_i, y)).$$

e Calculate the weight update parameter:

$$\alpha_t = \frac{\epsilon_t}{1 - \epsilon_t}.$$

f Update D_t:

$$D_{t+1}(i) = D_t(i)\alpha_t^{\frac{1}{2}(1+h_t(x_i,y_i)-h_t(x_i,y:y\neq y_i))}.$$

g Normalize D_{t+1}: Let $Z_t = \sum_i D_{t+1}(i)$.

$$D_{t+1}(i) = \frac{D_{t+1}(i)}{Z_t}.$$

3 Output the final hypothesis:

$$H(x) = \underset{y \in Y}{\operatorname{argmax}} \sum_{t=1}^{T} h_t(x, y) \log \frac{1}{\alpha_t}.$$

IV. SIMULATION RESULTS

In this section it shows that the comparison of simulation results of diabetic retinopathy detection using SVM classifier and RUS boost ensemble technique.

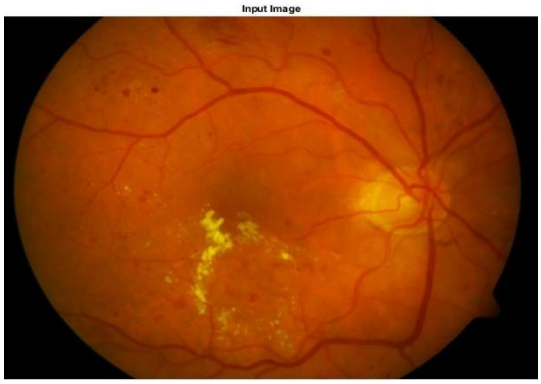

Figure 4(a): PDR as input image

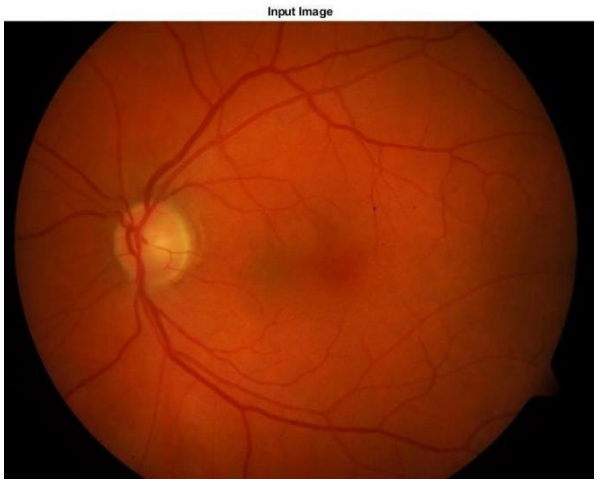

Figure 4(b): NDPR as input image

The input image is taken for the preprocessing stage which involves the illumination, denoising, equalization, and color normalization to make the input images as prepared for the next stage which are shown in figure 4(c) and 4(d).

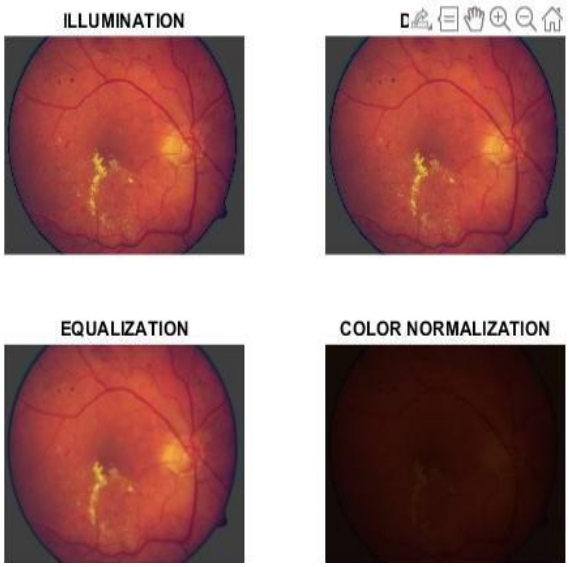

Figure 4(c): PDR preprocessed image

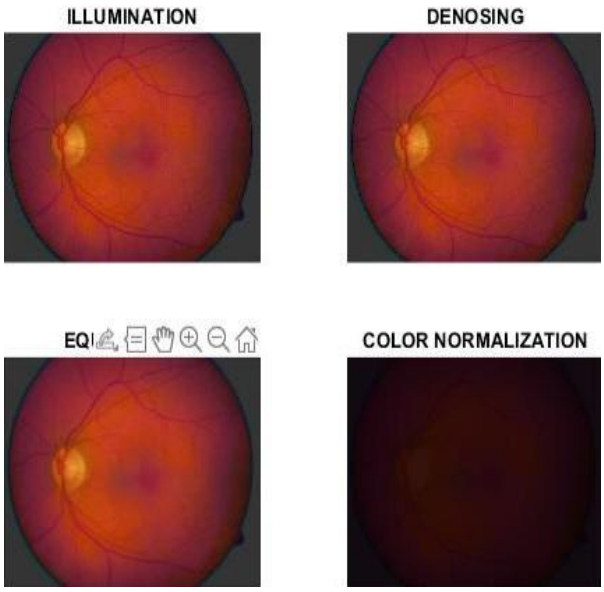

Figure 4(d): NDPR preprocessed input image

After preprocessing and all the next stage is to find the lesions like microaneurysms, hemorrhages, exudates, cotton wools. The non-proliferated finding candidates highlighted only the optical disk because it is not having any lesions on it but in the case of proliferated diabetic retinopathy case it is not highlighted the optical disk part but also the lesions are appeared around it. The figures 4(e) and 4(f) show the candidates which are get highlighted.

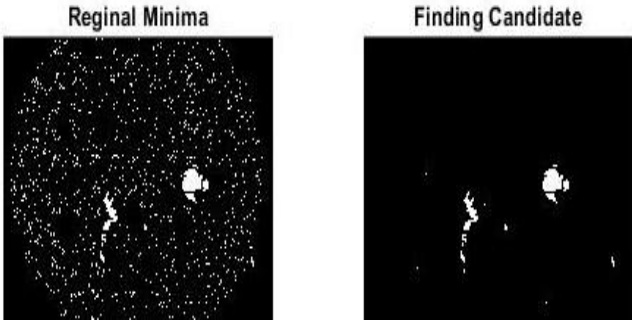

Figure 4(e): proliferated DR finding the candidates

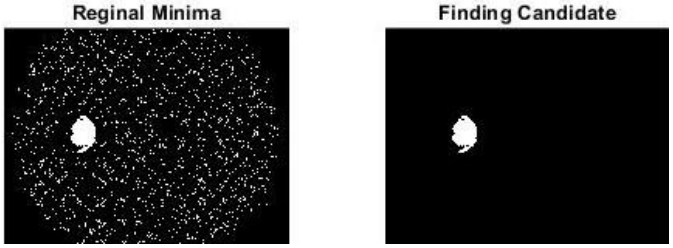

Figure 4(f): Non-proliferated DR finding the candidates in normal image

After the optimized RUS boost algorithm, the classifier founds the areas which are affected by diabetic retinopathy is as shown in the figure below.

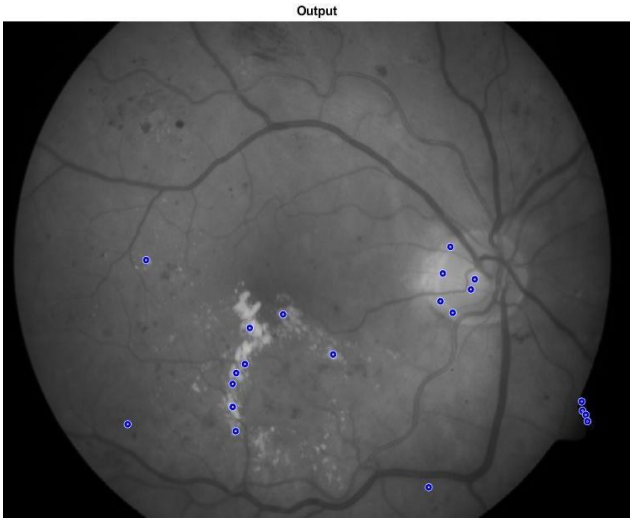

Figure 4(g): PDR output image

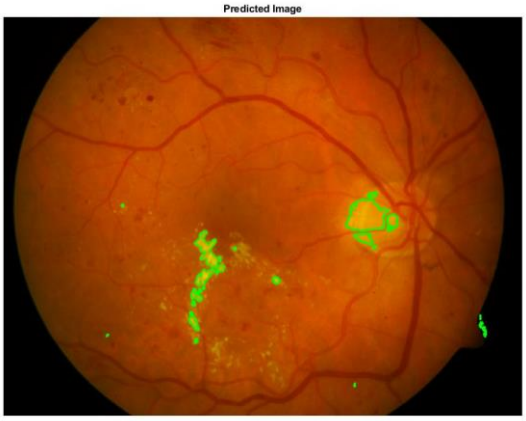

Figure 4(h): Affected region

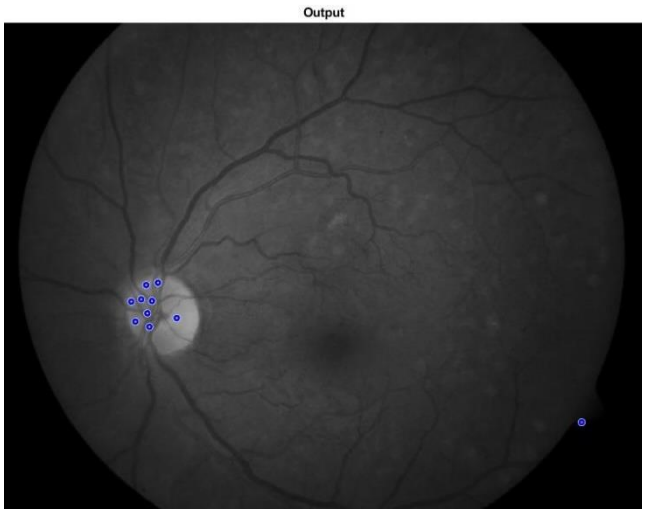

Figure 4 (i) : NDPR output image

After performing all the operations, the classifier shows a popup message as like that which it belongs to that is proliferated or non-proliferated diabetic retinopathy.

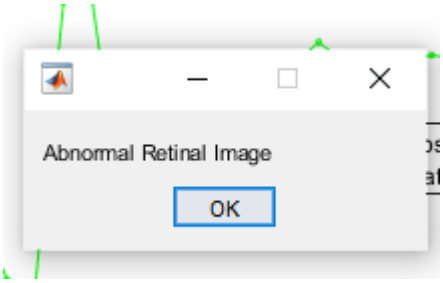

Figure 4(j): popping up of PDR (abnormal) result

The figure 4(g) shows that graph between minimum objective and number of function evaluations.

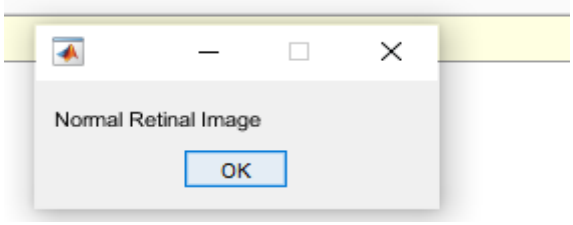

Figure 4(k): popping up of NPDR (normal) result

TABLE- I

METHOD	ACCURACY	SENSITIVITY	F1-SCORE
ORUS	0.988	0.93	0.967
SVM	0.957	0.87	0.933

V. CONCLUSION

Optimized RUS boost algorithm successfully can be able to validate the detection of diabetic retinopathy as compared with the SVM classifier. Compared with the SVM classifier this RUS boost algorithm can be able reduce error rate, increases the accuracy and increases the specificity. With these experimental results it is recommended that RUS boost algorithm is one best method for improving the classification performance of learners where ever skewed data presented.

VI. REFERENCES

[1]. U. Ahmed et al., "Prediction of Diabetes Empowered With Fused Machine Learning," in IEEE Access, vol. 10, pp. 8529-8538, 2022, doi: 10.1109/ACCESS.2022.3142097.

[2]. L. Qiao, Y. Zhu and H. Zhou, "Diabetic Retinopathy Detection Using Prognosis of Microaneurysm and Early Diagnosis System for Non-Proliferative Diabetic Retinopathy Based on Deep Learning Algorithms," in IEEE Access, vol. 8, pp. 104292-104302, 2020, doi: 10.1109/ACCESS.2020.2993937.

[3]. S. Roychowdhury, D. D. Koozekanani and K. K. Parhi, "Automated detection of neovascularization for proliferative diabetic retinopathy screening," 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Orlando, FL, USA, 2016, pp. 1300-1303, doi: 10.1109/EMBC.2016.7590945.

[4]. Weiss, Gary M. et al. "Cost-Sensitive Learning vs. Sampling: Which is Best for Handling Unbalanced Classes with Unequal Error Costs?" International Conference on Data Mining (2007).

[5]. N. V. Chawla, D. A. Cieslak, L. O. Hall, and A. Joshi, "Automatically countering imbalance and its empirical relationship to cost," Data Mining

Knowl. Discov., vol. 17, no. 2, pp. 225–252, Oct. 2008.

[6]. Roychowdhury S, Koozekanani DD, Parhi KK. Automated detection of neovascularization for proliferative diabetic retinopathy screening. Annu Int Conf IEEE Eng Med Biol Soc. 2016 Aug;2016:1300-1303. doi: 10.1109/EMBC.2016.7590945. PMID: 28268564.

[7]. S. Kumar and B. Kumar, "Diabetic Retinopathy Detection by Extracting Area and Number of Microaneurysm from Colour Fundus Image," 2018 5th International Conference on Signal Processing and Integrated Networks (SPIN), Noida, India, 2018, pp. 359-364, doi: 10.1109/SPIN.2018.8474264.

[8]. N. Singh, L. Kaur, and K. Singh, ‘‘Histogram equalization techniques for enhancement of low radiance retinal images for early detection of diabetic retinopathy,’’ Eng. Sci. Technol., Int. J., vol. 22, no. 3, pp. 736–745, Jun. 2019

[9].] L. Tang, M. Niemeijer, J. M. Reinhardt, M. K. Garvin, and M. D. Abramoff, ‘‘Splat feature classification with application to retinal hemorrhage detection in fundus images,’’ IEEE Trans. Med. Imag., vol. 32, no. 2, pp. 364–375, Feb. 2012.

[10]. Akram, M.U., et al. (2012) Automated detection of exu-dates in colored retinal images for diagnosis of diabetic retinopathy. Applied Optics, 51, 4858-4866.

Cite this article as :

Kanna Swarnalatha, Dr. I. Kullayamma , "Diabetic Retinopathy Detection Using Optimized Random Under Sampling (ORUS) Algorithm ", International Journal of Scientific Research in Science and Technology (IJSRST), Online ISSN : 2395-602X, Print ISSN : 2395-6011, Volume 10 Issue 2, pp. 495-502, March-April 2023. Available at doi : https://doi.org/10.32628/IJSRST52310275

Journal URL : https://ijsrst.com/IJSRST52310275