MEKATRONIKA
e-ISSN: 2637-0883
VOL. 2, ISSUE 2, 62 – 67
DOI: https://doi.org/10.15282/mekatronika.v2i2.6769

ORIGINAL ARTICLE

The Diagnosis Of Diabetic Retinopathy By Means Of Transfer Learning With Conventional Machine Learning Pipeline

Farhan Nabil Mohd Noor[1], Wan Hasbullah Mohd Isa[1*] and Anwar P.P. Abdul Majeed[1]

[1]Faculty of Manufacturing and Mechatronic Engineering Technology, Universiti Malaysia Pahang, 26600 Pahang, Malaysia.

ABSTRACT – Diabetic Retinopathy is one of the common eye diseases due to the complication of diabetes mellitus. Cotton wool spots, rough exudates, haemorrhages and microaneurysms are the symptoms of the diabetic retinopathy due to the fluid leakage that is caused by the high blood glucose level disorder. Early treatment to prevent a permanent blindness is important as it could save the diabetic retinopathy vision. Hence, in this study, we proposed to employ an automated detection method to diagnose the diabetic retinopathy. The dataset was obtained from the Kaggle Database and been divided for training, testing and validation purposes. Furthermore, Transfer Learning models, namely VGG19 were employed to extract the features before being processed by Machine Learning classifiers which are SVM, kNN and RF to classify the diabetic retinopathy. VGG19-SVM pipeline produced the best accuracy in training, testing and validation processes, achieving 99, 99 and 96 percents respectively.

ARTICLE HISTORY
Received: 14th Nov 2020
Revised: 2nd Dec 2020
Accepted: 16th Dec 2020

KEYWORDS
Diabetic Retinopathy
Transfer Learning
SVM
kNN
RF

INTRODUCTION

Diabetic Retinopathy (DR) is an eye disease caused by the complication of diabetes mellitus (DM). It occurs due to the elevated blood glucose level disorder that deteriorates the blood vessels within the eye and leaks the blood and fluids into surrounding tissues. Fluid leakage causes microaneurysms, rough exudates, cotton wool spots, and haemorrhages [1]. Globally, DR cases would increase from 126.6 million in 2010 to 191.0 million by 2030 if prompt interventions are not taken [2]. In Malaysia, the prevalence rate of DR ranges between 44.1 and 48.6 percent, with 12.3 percent for type 1 DM and 22.3 percent for type 2 DM in the majority of reported DR cases [3].

Early identification and detection of DR are essential since, if untreated, DR may lead to irreversible blindness. [2]. Unfortunately, most DR screening is based on labour-intensive human grading, which requires tidious education, training and examinations [4]. Unfortunately, there are also not enough graders to grade fundus photographs from distant clinics in third world nations, thus impeding the tele-screenings at tertiary care centres [5]. Due to these reasons, Wilkinson et al. found that a reliable detection method is necessary to help diagnose DR [6]. Therefore, Amin et al. (2016) recommended an automated DR screening system to reduce the risk of vision loss and laborious works of ophthalmologists [7]. Thus, the main goal of this research is to enhance such automated systems using Transfer Learning (TL) models and Machine Learning (ML) classifiers.

The remainder of this paper is organized as follows: Section 2 presents a brief overview of diagnosis DR through ML. Section 3 explains the TL and ML models employed in this research, then section 4 discusses the experimental results, and lastly, section 5 concludes the work in this paper.

LITERATURE REVIEW

Chetoui et al. used Local Energy-based Shape Histogram (LESH) and Local Ternary Pattern (LTP) methods to detect DR using imagery datasets acquired from the MESSIDOR database. SVM classifier with three different kernels, such as Linear, Polynomial and Radial Basis Functions (RBF), were used and tested with the 10-fold cross-validation technique to classify the DR images. They found that SVM coupled with RBF kernel and LESH features extraction obtained the highest results for accuracy and Area Under Curve (AUC), 90.4 and 93.1 percent, respectively [8].

Kirange et al. used ML to diagnose and grade DR based on IDRiD dataset. They employed Gabor Filter and Local Binary Pattern (LBP) to extract local features in the images. The dataset was separated such that 413 images for training and 103 images testing purposes, respectively. The identified local features were further processed by Naïve Bayes (NB), k-Nearest Neighbours (kNN), Decision Tree, Support Vector Machine (SVM) and Neural Network classifiers to predict the DR. kNN achieves the highest accuracy of 71.63 percent for images extracted using the Gabor features, while NB outperformed other classifiers with 77.86 percent accuracy for images processed using the LBP method [9].

Qomariah et al. proposed the coupling between Convolutional Neural Network (CNN) and SVM method for to classify DR. They used 77 and 70 images consisting of normal and severe Non-Proliferative Diabetic Retinopathy (NPDR) (base 12 and 13) from the Messidor database. CNN models including Alexnet, GoogleNet, Resnet50, DenseNet201, VGG16 & VGG19 and InceptionV2 &V3 were employed to extract the images' significant features before being fed to the SVM

*CORRESPONDING AUTHOR | Wan Hasbullah Mohd Isa | wanhasbullah@ump.edu.my

classifier. Their results showed that Resnet50 achieved the highest accuracy of 95.83 percent for base 12 images, while VGG19 and InceptionV3 were most accurate with 95.24 percent for base 13 images [10].

Lahmiri investigated the efficiency of using hybrid deep learning, which combined CNN with different classfier models to predict DR. CNN model was employed to extract the features from the images obtained from STARE database while SVM, Latent Dirichlet Allocation, NB and kNN classifiers are used to classify them. He also employed student t-test, and 10-fold cross-validation is his work. The results showed that CNN-SVM pipeline produced the best performance with 99.11 percent accuracy [11].

Gayathri et al. coupled CNN with distinct classifiers, namely SVM, Multi-Layer Perceptron, Random Forest (RF) and J48, to classify fundus images using the datasets from the IDRiD, Kaggle and Messidor database. Two experiments have been conducted for binary and multi-class classifications, respectively. They found that the J48 classifier outperformed the others with an average accuracy of 99.89 percent for binary classification and 99.59 percent for multi-class classification [12].

METHODOLOGY

In this section, we explain the methodology that we used to obtain the results in our work. First, we explain the details of the dataset that we processed to classify the DR. Then, we discuss the features extraction methods using TL methods, namely CNN and VGGNet. In the following subsection, we elaborate on different classification methods on the extracted features beforehand, including SVM, kNN, and RF. Lastly, we describe the performance matrices that were used to evaluate the classification results.

Dataset

We obtained contains 2000 colourful fundus retina images from the Kaggle Database. They include normal and DR images of various severity levels, as shown in Figure 1. We used them for the evaluation and classification processes. These images were resized into 224×224 pixels dimensions to be compatible with the VGG19 TL model. This set of images was divided into 70:15:15 ratio each for the training, testing and validation purposes.

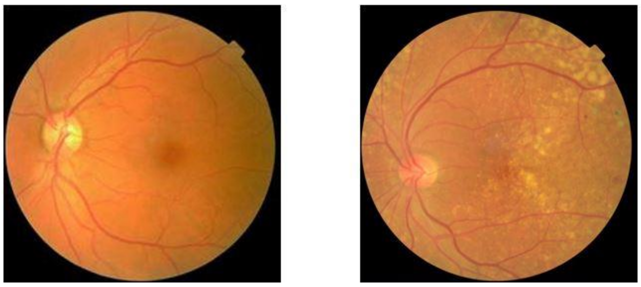


Figure 1. Left is the Normal retinal image, and right is the Proliferative DR image from the dataset.

Features Extraction: Transfer Learning

CNN

We used CNN to extract features from the cancerous images because it is highly effective on big datasets [6], especially in state-of-the-art visual applications, including image classifications [12]. This method is able to retrieve both local and global contextual features in images automatically and utilizes local sub-sampling and convolutional filters to hierarchize increasingly complicated characteristics [11]. CNN's initial convolutional layer functions as a feature extractor from retina images. Image matrix and filter are utilized to perform the mathematical procedure [13]. Besides, CNN does not require as many total parameters since convolution is done using the weights shared by each neuron. [14]. Figure 2 illustrated the architecture of the Convolution Neural Network (CNN).

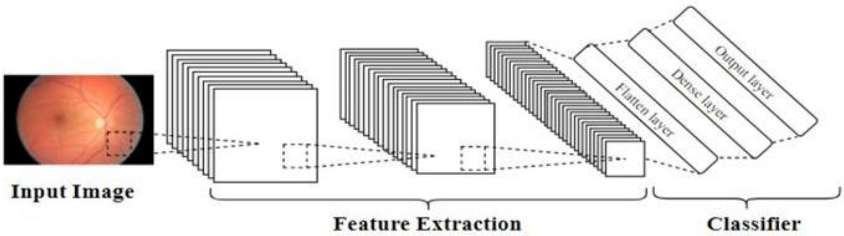


Figure 2. General Convolution Neural Network (CNN) architecture [6].

In this work, we removed the original classifier, namely Fully Connected Layers, and freeze all the network weight. Then, we used the VGG19 model to automatically extract the retinal images' significance features before being classified. The model was developed by researchers at the University of Oxford in 2014 [6]. The 3×3 small convolution kernels and 2×2 top pooling layers were stacked repeatedly to create 19 deeps layers of neural networks for VGG19 [15]. VGG19 consists of an extra convolution layer compared to the VGG16 model, which is 143,667,240 parameters. It is stated that complicated structures may be learnt and improved via additional layers [6]. The input must be 224×224 pixel images [10].

Classification

We used three types of classifier, namely Support Vector Machine (SVM), k-Nearest Neighbours (kNN), and Random Forest (RF).

Support Vector Machine (SVM)

SVM has proven to be effective classifiers for a wide variety of application applications due to its ability of the framework to generalize, robustness to divergences, capacity to withstand massive data collections, and higher dimensions. [16]. The empirical risk and overfitting were alleviated by SVM, resulting in a compelling performance [17]. The SVM uses a hyperplane to segment data in more dimensions to handle classification constraints. [18]. The hyperplane can be represented as:

$$\vec{w}.\vec{x} - b = 0 \quad (1)$$

Where, w is known as a normalized hyperplane vector while by using the distance from a point to a plane equation, we can calculate the $\frac{b}{\|w\|}$ and the x set of points. Given that the data points are linearly separable, it is feasible to construct two parallel hyperplanes. The bounded region is regarded as a margin, meaning the data points belong to any data class [19].

k-Nearest Neighbours (kNN)

The kNN classifier is the most superficial supervised ML used for regression and classification purposes. It is classified as a 'lazy' algorithm because it identifies a new data point using the whole training set [16]. It loads all the qualifying instances and utilizes their k-neighbours' votes [20]. The k-value may be varied between 2 and 8 to get the most efficient outcome [21]. The selection of the *k*-value can be governed by a basic rule that states that k should be $\sqrt{n}$ where *n* represents the sample's total number for the learning set [22]. In this algorithm, the distance between the data and its closest neighbour is used to categorize the data, with the data being assigned to the most frequent class among its *k* closest neighbours. The term distance is used in the kNN technique to represent each data point in the multidimensional feature space using position vectors [23].

Random Forest (RF)

RF is a supervised ML that combines numerous decision trees in an ensemble to provide a more accurate forecast than a single model. Each tree contains a portion of characteristics of the original data [24] and casts the votes to choose a class for the new item with the majority votes will determine the classification [20]. The random subsamples of the training set, such as random split selection and bagging, will cultivate the tree ensembles. The random attributes are chosen at each node, and the algorithm determines each one's entropy to identify the optimal attribute for the remaining training examples classification process [25].

Performance Metrics

In this work, we utilized classification accuracy and confusion matrix as the performance metrics for evaluating and analyzing the results. Accuracy determines the percentage of the predictions that the classification model precisely produces. A confusion matrix is represented by a table that visualizes the classifier's performance, either the correct prediction or the wrong prediction made by the classification model, as shown in Table 1.

Table 1. Confusion Matrix Table. T means true, F means false, P means positive, and N means negative.

		Predicted	
		Positive	Negative
Actual	Positive	TP	FP
	Negative	FN	TN

RESULTS

This section discusses our work's classification results, where the VGG19 model was used to extract features and coupled with SVM, kNN and RF classifiers to detect RD. Table 2 shows the accuracies of classification using the couplings of the VGG19 model and different classifiers in each training, testing and validation process.

Table 2. Accuracy of classification using VGG19 model and different classifiers in percentage.

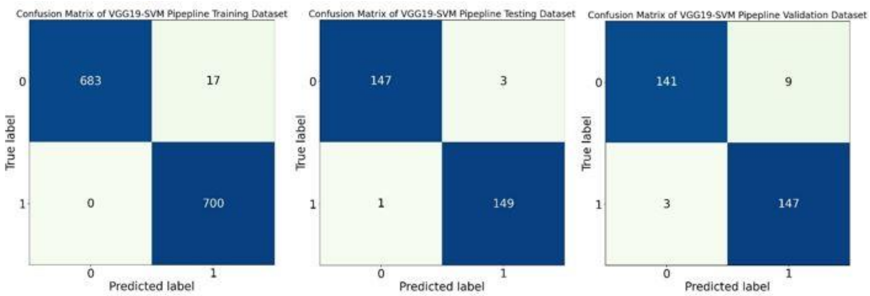
Performance Metric	SVM	kNN	RF
Training	99	94	100
Testing	99	93	98
Validation	96	93	95

As shown in Table 2, the VGG19-RF combination obtained a perfect accuracy using the training dataset. On the other hand, the VGG19-SVM pipeline achieved an almost similar result in accuracy, only a 1 percent difference. Even though VGG19-kNN coupling performed least accurately at 94 percent, the result is still considered very good and acceptable.

As for testing and validation datasets to classify DR cancerous features from images, VGG19-SVM outperformed the other pipelines. VGG19-RF coupling produced the second best accuracy, with only 1 percent difference with the VGG19-SVM combination for both datasets. VGG19-kNN classified least accurately using these datasets, both with 93 percent accuracy.

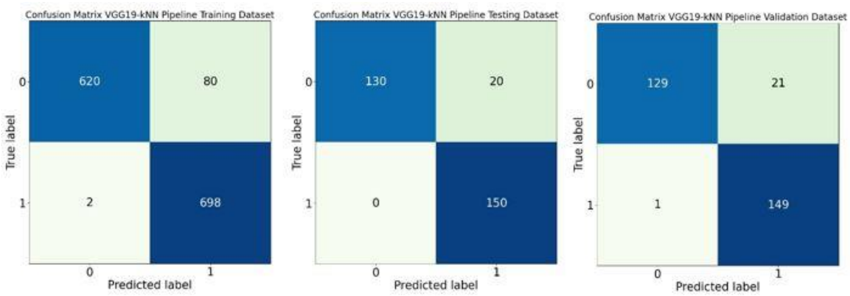
Figure 3-5 illustrate the confusion matrices for all VGG19-classifier pipelines using distinct training, testing and validation datasets. Table 2 refers to the DR case, while 1 to Normal, i.e. non-DR case, respectively. Generally, all pipelines have excellent performances in the classification for the Normal case images. VGG19-SVM pipeline produced no misclassified Normal case images using the training dataset, followed by only one and three labels using testing and validation datasets. Even though VGG19-kNN pipeline classifies DR images least accurately compared to other pipelines, it produce very good results in confusion matrix, non misclassified Normal images using testing dataset, but only two and one Normal images using training and validation datasets. Meanwhile, the VGG19-RF pipeline produced a precise prediction using a training dataset with no misclassified images for DR and Normal images. It only misclassified six and 10 DR images using testing and validation datasets. On the other hand, VGG19-kNN produced the most misclassified DR images during training, testing, and validation works. VGG19-SVM pipeline has misclassified 17 DR images in training, and the number decreased to nine and three images using validation and testing datasets.

From these results, the SVM classifier was most efficient when coupled with the VGG19 model because it obtained the best classification accuracy on average for the classification of DR images compared to kNN and FR classifiers.



(a) (b) (c)

Figure 3. Confusion Matrix of the VGG19-SVM pipeline for training, testing and validation.



(a) (b) (c)

Figure 4. Confusion Matrix of the VGG19-kNN pipeline for training, testing and validation.

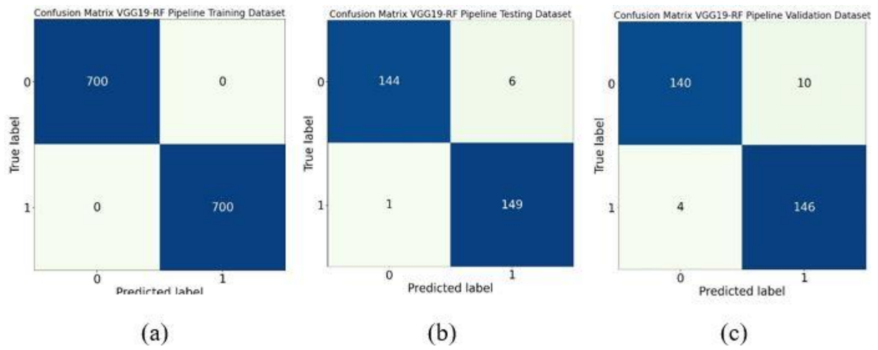


Figure 5. Confusion Matrix of the VGG19-RF pipeline for training, testing and validation.

CONCLUSION

This paper employs an automated detection method to diagnose diabetic retinopathy (DR) using the Transfer Learning method. VGG19 model was used to extract features from images, and three classifiers, namely Support Vector Machine (SVM), k-Nearest Neighbours (kNN), and Random Forest (RF) was utilized to classify the DR images for detection purposes. The results showed that the combinations of the VGG19 model and these classifiers could extract the significant features with above 90 percent accuracy using all training, testing and validation datasets. Although VGG19-kNN pipeline performed slightly less than the other pipelines, it still produced a decent performance with slight overfitting. Generally, the VGG19-SVM pipeline was demonstrated to be the best coupling to classify DR images as it produced the best averaged performances, although the VGG19-RF pipeline was not much outperformed by it. This study indicates a promising potential for automated ML detection in the classification of DR.

REFERENCES

[1] E. v. Carrera, A. Gonzalez, and R. Carrera, "Automated detection of diabetic retinopathy using SVM," *Proceedings of the 2017 IEEE 24th International Congress on Electronics, Electrical Engineering and Computing, INTERCON 2017*, pp. 6–9, 2017, doi: 10.1109/INTERCON.2017.8079692.

[2] M. W. Stewart, *Diabetic retinopathy, Current Pharmacologic Treatment and Emerging Strategies*, vol. 95, no. 3. Singapore: Springer Nature, 1991. doi: 9789811035098.

[3] M. H. Mohd Ali, N. Draman, W. M. I. W. Mohamed, A. Yaakub, and Z. Embong, "Predictors of proliferative diabetic retinopathy among patients with type 2 diabetes mellitus in Malaysia as detected by fundus photography," *Journal of Taibah University Medical Sciences*, vol. 11, no. 4, pp. 353–358, 2016, doi: 10.1016/j.jtumed.2016.03.002.

[4] A. A. Salman, I. M. Khairuddin, A. P. P. A. Majeed, and M. A. M. Razman, "The Diagnosis Of Diabetic Retinopathy By Means Of Transfer Larning And Fine-Tuned Dense Layer Pipeline," *MEKATRONIKA*, vol. 2, no. 1, pp. 68–72, Jun. 2020.

[5] V. Gulshan *et al.*, "Performance of a Deep-Learning Algorithm vs Manual Grading for Detecting Diabetic Retinopathy in India," *JAMA Ophthalmology*, vol. 137, no. 9, pp. 987–993, 2019, doi: 10.1001/jamaophthalmol.2019.2004.

[6] I. Kandel and M. Castelli, "Transfer learning with convolutional neural networks for diabetic retinopathy image classification. A review," *Applied Sciences (Switzerland)*, vol. 10, no. 6, 2020, doi: 10.3390/app10062021.

[7] J. Amin, M. Sharif, and M. Yasmin, "A Review on Recent Developments for Detection of Diabetic Retinopathy," *International Journal of Scientific and Technology Research*, vol. 9, no. 2, p. 20, 2016, doi: 10.1155/2016/6838976.

[8] M. Chetoui, M. A. Akhloufi, and M. Kardouchi, "Diabetic Retinopathy Detection Using Machine Learning and Texture Features," *Canadian Conference on Electrical and Computer Engineering*, vol. 2018-May, pp. 2018–2021, 2018, doi: 10.1109/CCECE.2018.8447809.

[9] D. K. Kirange, J. P. Chaudhari, K. P. Rane, K. S. Bhagat, and N. Chaudhri, "Diabetic retinopathy detection and grading using machine learning," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 8, no. 6, pp. 3570–3576, 2019, doi: 10.30534/ijatcse/2019/139862019.

[10] D. U. N. Qomariah, H. Tjandrasa, and C. Fatichah, "Classification of diabetic retinopathy and normal retinal images using CNN and SVM," *Proceedings of 2019 International Conference on Information and Communication Technology and Systems, ICTS 2019*, pp. 152–157, 2019, doi: 10.1109/ICTS.2019.8850940.

[11] S. Lahmiri, "Hybrid deep learning convolutional neural networks and optimal nonlinear support vector machine to detect presence of hemorrhage in retina," *Biomedical Signal Processing and Control*, vol. 60, p. 101978, 2020, doi: 10.1016/j.bspc.2020.101978.

[12] S. Gayathri, V. P. Gopi, and P. Palanisamy, "A lightweight CNN for Diabetic Retinopathy classification from fundus images," *Biomedical Signal Processing and Control*, vol. 62, no. July, 2020, doi: 10.1016/j.bspc.2020.102115.

[13] A. U. Naik and R. K. Kulkarni, "Artificial neural network-based detection of diabetes and its effects on vision - A survey," *Proceedings of the 5th International Conference on Communication and Electronics Systems, ICCES 2020*, no. Icces, pp. 1113–1118, 2020, doi: 10.1109/ICCES48766.2020.09138057.

[14] C. Mahiba and A. Jayachandran, "Severity analysis of diabetic retinopathy in retinal images using hybrid structure descriptor and modified CNNs," *Measurement: Journal of the International Measurement Confederation*, vol. 135, pp. 762–767, 2019, doi: 10.1016/j.measurement.2018.12.032.

[15] S. Wan, Y. Liang, and Y. Zhang, "Deep convolutional neural networks for diabetic retinopathy detection by image classification," *ACM International Conference Proceeding Series*, vol. 72, pp. 68–72, 2018, doi: 10.1145/3239576.3239589.

[16] D. Karanjgaokar, B. K. Prasanna, and R. K. Chaurasiya, "Comparison of classification methodologies for predicting the stages of diabetic retinopathy," *Proceedings of the International Conference on Intelligent Sustainable Systems, ICISS 2017*, no. ICISS, pp. 700–705, 2018, doi: 10.1109/ISS1.2017.8389263.
[17] G. Mahendran and R. Dhanasekaran, "Investigation of the severity level of diabetic retinopathy using supervised classifier algorithms," *Computers and Electrical Engineering*, vol. 45, pp. 312–323, 2015, doi: 10.1016/j.compeleceng.2015.01.013.
[18] R. Santhakumar, M. Tandur, E. R. Rajkumar, K. S. Geetha, G. Haritz, and K. T. Rajamani, "Machine learning algorithm for retinal image analysis," *IEEE Region 10 Annual International Conference, Proceedings/TENCON*, pp. 1236–1240, 2017, doi: 10.1109/TENCON.2016.7848208.
[19] K. K. Mohanty, P. K. Barik, R. C. Barik, and K. C. Bhuyan, "An efficient prediction of diabetic from retinopathy using machine learning and signal processing approach," *Proceedings - 2019 International Conference on Information Technology, ICIT 2019*, pp. 103–108, 2019, doi: 10.1109/ICIT48102.2019.00025.
[20] K. S. Lotlekar and S. D. Desai, "Multilevel Classification Model for Diabetic Retinopathy," *Proceedings of the International Conference on Computational Techniques, Electronics and Mechanical Systems, CTEMS 2018*, pp. 326–331, 2018, doi: 10.1109/CTEMS.2018.8769207.
[21] V. Deepa, C. S. Kumar, and S. S. Andrews, "Automated grading of diabetic retinopathy using local-spatial descriptors," *2020 IEEE International Conference on Computing, Power and Communication Technologies, GUCON 2020*, pp. 660–664, 2020, doi: 10.1109/GUCON48875.2020.9231221.
[22] B. M. Ege *et al.*, "Screening for diabetic retinopathy using computer based image analysis and statistical classification," *Computer Methods and Programs in Biomedicine*, vol. 62, no. 3, pp. 165–175, 2000, doi: 10.1016/S0169-2607(00)00065-1.
[23] P. Nijalingappa and B. Sandeep, "Machine learning approach for the identification of diabetes retinopathy and its stages," *Proceedings of the 2015 International Conference on Applied and Theoretical Computing and Communication Technology, iCATccT 2015*, vol. 1, no. c, pp. 653–658, 2016, doi: 10.1109/ICATCCT.2015.7456965.
[24] S. S. Prem and A. C. Umesh, "Classification of Exudates for Diabetic Retinopathy Prediction using Machine Learning," *2020 IEEE 5th International Conference on Computing Communication and Automation, ICCCA 2020*, pp. 357–362, 2020, doi: 10.1109/ICCCA49541.2020.9250858.
[25] S. Stolte and R. Fang, "A survey on medical image analysis in diabetic retinopathy," *Medical Image Analysis*, vol. 64, 2020, doi: 10.1016/j.media.2020.101742.