

Hyper parameter tuning based gradient boosting algorithm for detection of diabetic retinopathy: an analytical review

Parul Datta¹, Prasenjit Das², Abhishek Kumar¹

¹Department of Computer Science and Engineering, Chitkara University School of Engineering and Technology, Chitkara University, Himachal Pradesh, India

²Chitkara University School of Computer Applications, Chitkara University, Himachal Pradesh, India

Article Info

Article history:

Received Aug 9, 2021

Revised Dec 7, 2021

Accepted Mar 11, 2022

Keywords:

Diabetic retinopathy

Fundus images

Gradient boosting

Hyper-parameter tuning

Pipeline

ABSTRACT

The pipelines of approaches for classifying diabetic retinopathy were examined in this study. The effort entails developing appropriate transformations and estimators that can be used to automate the process of diabetic retinopathy detection. The segmentation of the blood vessels was done using a hybrid algorithm that uses Otsu and median filter to get the region of interest. Further, ten classifiers were investigated in order to develop an automated pipeline for diabetic retinopathy detection. The ten classifiers were reviewed based on earlier work in a similar setting and on an exploration of new ways for identifying diabetic retinopathy. To overcome the challenge of low volume of dataset, data augmentation was done so that a generic classifier can be configured. Extensive hyper parameter tuning was performed, and it was shown that the gradient boosting approach is the most stable technique for detecting diabetic retinopathy. This was validated using a 10K fold cross validation method on many metrics (accuracy, recall, precision, and v-measure score). Hyper-parameter tuning helped in achieving accuracy of 0.96.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Parul Datta

Department of Computer Science and Engineering, Chitkara University School of Engineering and Technology
Chitkara University

Atal Shiksha Kunj, Pinjore -Nalagarh National Highway (NH-21A) Distt.-Solan, Himachal Pradesh-174103, India

Email: parul.datta@chitkarauniversity.edu.in

1. INTRODUCTION

The world is suffering from health problems due to covid-19 pandemic [1]. India is no exception. Indians are in-fact, quite prone to COVID-19 virus due to multiple medical reasons [2]. One of the many reasons is that a large number of people in India are diabetic. India can be considered as a “Diabetic Capital” of the world [2]-[4]. It is a clinically well-established fact that diabetes creates many severe medical conditions as a comorbid [4], [5]. The co-prevalence of comorbidities increases with increase in biological age as well. These facts are clinically known as a sequel of diabetes [6]. COVID-19 pandemic is posing a lot of challenges to health systems and it urgently needs to be supplemented with emerging proven technologies. And these technologies must work for both healthcare providers and patients. One of the most critical and challenging situations is to manage the co-morbidities of covid-19 patients, including diabetes and associated ailments such as diabetic retinopathy (DR). Due to this, diabetic retinopathy is spreading in India at a considerable speed [7], [8]. For the medical diagnosis and treatment of the diabetic retinopathy, growth of imaging technology is necessary, and at the same time leveraging the latest machine learning algorithm is essential. The landscape of the technologies and algorithms that support detection of diabetic retinopathy begins with the advent of machine learning algorithms [9]-[11]. These days for detection of diabetic

retinopathy, use of fundus photography in diagnoses has become quite reliable. The current literature shows that the most frequently employed algorithms in context of detecting diabetic retinopathy include neural networks, convolutional neural network (CNN) [12]-[14], support vector machine (SVM) [15], [16], K-nearest neighbors (KNN) [15], decision trees [17] and regression based algorithms. In almost all cases, the input dataset is either pre-processed or transformed into a standardized/normalized form. Besides, some authors have used combinational or hybrid approaches for detection of diabetic retinopathy. There is an imperative need to develop frameworks and pipelines of algorithms that can restore the reliability of such existing algorithms. This is because the detection of the diabetic retinopathy involves examination of multiple factors such as exudates [18], [19], optic nerve health [20], blood vessel shape, and observation related to size of other eyeball parts [21], [22]. The imaging helps to clinically corroborate the symptoms of diabetic retinopathy and at the same time helps to ascertain the health of the blood vessels in terms of their shape, formation, and thickness [23], [24]. More importantly, the imaging technology has been in use for a long time in monitoring the brain and eye in terms of various medical conditions and disorders. However, diabetic retinopathy is a condition in which it is of utmost importance to have a tool that can ascertain the condition of the brain along with the eyes. This is because of diabetic retinopathy is not a disease that can be diagnosed without doing the visual acuity test on the eye as well as an ophthalmologist examination. Hence, an imaging tool that can ascertain the condition of the eye and brain at the same time is the most sought-after tool for monitoring the condition of the brain and eye. The imaging technology helps to find if some exudates are blurring the vision of the person suffering from diabetic retinopathy [25], [26]. It helps to spot any abnormalities in the eye with a lot of ease and accuracy. However, at the same time, it should be noted that from digital images only relevant information is required to be extracted and segmented, so that mapping with real life medical conditions could be done. Digital fundus photography enables us to do the diagnosis in an effective way [27]. The standard fundus imaging process includes capturing eye image in posterior pole position at 30 degrees [28], [29]. It can help to examine the optic nerve and macula properly. However, in some cases, it has been observed that this imaging modality cannot verify clinically noticeable differences. Hence, some additional pre-processing of the images may be required to build a clinically relevant solution for the detection of the diabetic retinopathy [30]-[32]. Images may suffer from the effects of angle distortion and artificial artifacts and development of detection will require pre-processing of images. Another method, stereoscopic fundus photography is employed when there is a need to medically examine both eyes and it has been observed in many cases that for tracking the treatment outcomes it is important to examine both the eyes at different stages. The challenge is if we use eye images to have noise, distortion, for building a diabetic detection system, a lot of computation complexity needs to be resolved for building a DR detection system. This problem may be solved by using state of art machine-learning algorithms. It can also be observed that clinical dataset in the context of DR in the public domain suffers from issues such as small in size, unmarked and unannotated, imbalance and non-specific in terms of demographics [33]-[35]. Due to these issues the use of deep learning algorithms that extract image features directly is also in use. At the same time, it can be observed that reliability of the machine learning algorithm is proven beyond doubt in this context of detection DR. In fact, it can be safely stated that with the use of handcrafted features methods, conventional statistics, data fusion, data argumentation, transfer, active learning models the landscape is not complete. There are ample research studies in this context; however, the problems related to imbalance and small size of the dataset remains alive as we undertake this study [36], [37].

The objective of this research work is to construct a numerically stable classifier that can classify DR. This manuscript provides mechanisms for detecting DR using image based features and machine learning algorithms. To accomplish these goals, an explorative journey for building a classifier that can detect DR using eye fundus images was performed. The exploration will be focused on finding an appropriate array of methods that can be classified using a limited number of instances of eye images. The machine learning models that are going to be put under evaluation will be based on the foregoing work done by many authors in the context of building diabetic retinopathy classifiers and the challenges such as small size of dataset. At the end of work, an attempt to overcome the inadequacies of existing algorithms in terms of accuracy is demonstrated. The inadequacies will be overcome by conducting extensive hyper-parameter tuning and research oriented exploration.

2. PROPOSED METHOD

This section gives detailed information on all the steps taken to construct the most stable and accurate machine model for detecting DR. The flow of the machine learning algorithms/classifier can be streamlined with the support of a concept known as “Pipelining”. The process of pipelining in machine learning is typically used for automating the standard functions that are required for construction of an accurate classifier system. In the context of our research work, the objective is to build an automated assortment of functions that can help classify a healthy eye from an unhealthy eye i.e. an eye suffering from

diabetic retinopathy. The original database contains a total of 30 high-resolution fundus images (15 healthy and 15 images of patients suffering from diabetic retinopathy). The reference to the dataset can be found here [38]. The quantity of the dataset is a significant hindrance to the creation of machine learning models. The largest obstacle in this research has been the dataset's short size. As a result, the generalization of the machine learning model becomes more challenging. Simultaneously, it may result in under-fitting. To address this obstacle, data augmentation and 10K cross-validations were used to verify the construction of a highly credible model. Ten distinct augmentation functions (flip, rotation, brightness, and zoom.) were used on the dataset. This way we extended the dataset size to 300 images ((healthy and unhealthy)) for both classes. The second technique for developing a reliable classifier is to keep the number of attributes to a maximum. Figure 1 shows sample images used in this research work after the segmentation process.

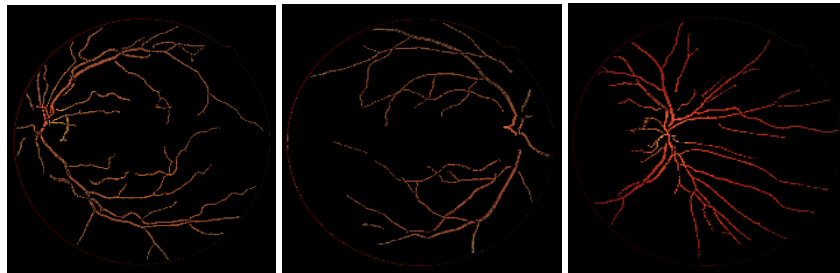


Figure 1. Segmented sample images

2.1. Segmentation of the blood vessels

An exploratory study was conducted in this research work to determine which segmentation method(s) can accurately segment the blood vessels of images of both (left and right) eyes. The segmentation was performed in order to obtain the regions of the eyes that correspond to the blood vessels. This current work employs the same methodologies as in [39], [40]. Explicitly, Otsu global thresholding [41], [42] with the mean filter for extraction of the blood vessels worked best. For extraction of the features and attributes for the machine learning algorithms region props method was used. The region props method treats the negative and zero values pixels as background and other pixels are marked as one. It checks that at least four pixels are connected with each other and it can be considered as a region. Each region is given a numerical label or index and geometric and statistical properties of each region can be extracted. The geometric property such as area, convex area, and can give hints about the state of the eye [43]. The properties include area, filled area, and convex area, centroid, bounding box area, orientation, major axis length, minor axis length, eccentricity, moments normalized, inertia tensor, inertia tensor eigen-values, and label. The selection of these attributes/properties has been done on the premise that when an eye is examined in detail (region wise), changes in the shape, and size can be observed. These changes (morphology) can be observed by measuring the area, volume, eigen vectors, and other statistical properties of the region of interest. In the coming sections, discussion on the next step i.e. model selection in building a pipeline. For doing model selection ten classifiers (logistic regression (LR), SVM, KNN, linear discriminant analysis (LDA), classification and regression trees (CART), extra trees classifier (ETC), Naïve Bayes (NB), random forest (RF), gradient boosting classifier (GBC), and multilayer perceptron (MLP)) [44]-[50] of different types have been selected so that a robust pipeline can be constructed for detection of DR. Secondly; this covers some of the work done by previous authors as well. However, in this part of the paper, a discussion on the impact of scaling on accuracy of classifiers followed by the interpretation is done. For best possible performance, some algorithms require normalization so that all the attribute values have a similar range while some algorithms require data that should be standardized. In this subsection, we evaluate the impact of the different scaling methods on the accuracy of the classifiers using a 10K fold cross validation method.

2.1.1. Interpretation

From the outcome of all these algorithms and after the application of different scaling methods, it is clear that the data contain a large number of extreme values and each attribute has a different scale (Refer Figure 2). This can be attributed to the fact that the accuracy is getting impacted due to change in scales of the data. These two issues lead to degradation of predictive performance [46]. Gradient-based classifiers assume that the data is approximately standardized and other classifiers have some other underlying presumption of the distribution of the data. Quantile transformers [47] and power transformers perform non-linear transformation of the data. Quantile scaling is based on the minimization process between the outliers

and inliers data points. The power transformers map its data to normal distribution so that variance and skewness is minimized. It can be observed from Figure 2 that non-linear transformation [48] (power) is helping to achieve high level accuracy in case of GBC (0.91). The standard deviation is also tight (0.1500). It seems that MinMax and MaxAbs scaling [49], [50] is not working in favor of accuracy. This may be attributed to the fact that in MinMax scaling the data is rescaled in range of (0, 1) and compresses all the inliers into a narrow range of [0, 0.005]. The robust scaling however is performing even better than the power transformer by 2% for the GBC algorithm. However, for other classifiers the accuracy is less as compared to the power transformer scaling. Normalization [51] does not have a good impact on the balanced accuracy of the classifiers. Therefore, it is imperative to use robust or power transformation scaling for doing data preprocessing and preparation of feature data for building a reliable classifier. The evaluation of the impact of training data size shows interesting facts. Since, the number of eye images is just 30; it is difficult to determine the level of generalization that can be achieved by these classifiers. Hence, to make a generalized algorithm and numerically stable we are relying on a large number of attributes as input data for machine learning algorithms. Consequently, we continued with the 139 input features (i.e. geometric properties of first fifteen objects) to find what percentage of training data is enough for building a classifier pipeline. Furthermore, after the data augmentation processes the dataset size increased to 300 samples keeping generalization as an important goal in mind.

It can be observed from Figure 3(a) and Figure 3(b) (in Appendix) that when the testing ratio is 15% the classifier's accuracy remains stable. This is an indication 85% of data is required for maintaining a good level of balanced accuracy of almost all the algorithms. In this next section, we evaluate the other performance parameters such as recall, precision, and so that final selection of the machine-learning model can be made and construction of the automated pipeline can be made.

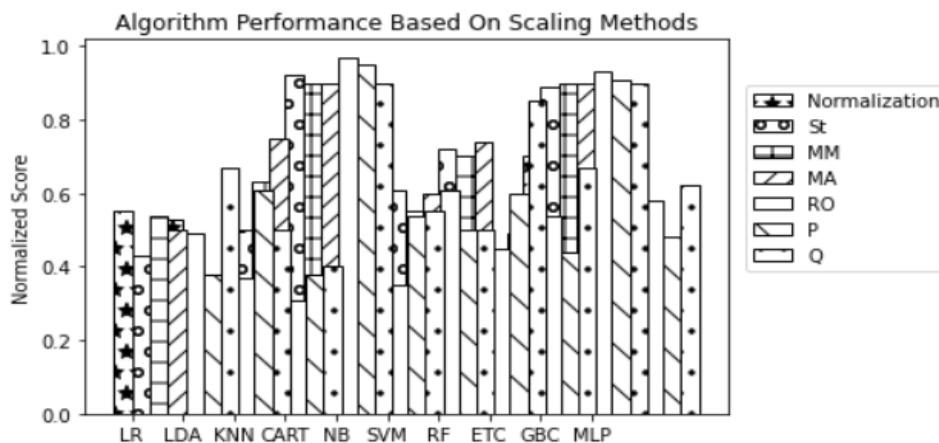


Figure 2. Impact of scaling on accuracy

3. METHOD

In this section, we conduct an evaluation of important metrics for final selection of the machine-learning model. Figure 4(a) and Figure 4(b) shows the performance analysis of the classifiers. The interpretation is followed thereafter.

The metric R2 [52] helps to find the goodness of the machine learning model. This metric helps to find out how good the model is in terms of fitting (fit, over or under) the data. It can be observed that the GBC and CART algorithm has positive R2 value (0.5) with tight standard deviations. The value in case LDA, KNN, and ETC is negative showing that the machine learning model is not choosing the correct trend for identifying class boundaries. Since, this work is medical related, the best scenario will be “high recall and high precision”, because the cost of not diagnosing a person with DR may lead to worsening of the medical condition of the person if he is actually suffering from DR. At the same time, diagnosing a person with DR, who is not suffering from DR may lead to giving him or her false treatment and unnecessary medical complications. Hence, in the context of our research work it is important to have a good trade-off between recall and precision. Moreover, it is always desired that algorithms should be able accurately classify DR and non-DR cases for the full dataset and with small samples. This can be observed in the case of GBC and CART algorithms. The standard deviation of GBC is smaller than the CART algorithm, showing that GBC is a bit better in terms of stability. Furthermore, it can be observed in case of precision [53], that in the case of

GBC that it is able to acquire the highest level of precision. From this, it can be inferred that true positives are highest in the case of GBC and it has the lowest false alarm rate. It is evident that GBC will have the highest accuracy level followed by CART, as it will have similar values as those of recall and precision in context of V-measure metric. It is always desired that a class should have homogenous members and at the same time be different from the other class members while being assigned labels by the clustering or classification algorithm-measure is a metric that gives the harmonic mean of the homogeneity (all data members belong to respective class) and completeness (all data points are covered and assigned a class). It can be observed that the GBC algorithm has value 0.8, which is closest to the ideal value of one, followed by CART. Again, here also, it can be observed that GBC is offering better results as it has a tighter level of standard deviation. It can also be observed that SVM has the lowest level of performance. MLP, RF, and NB are slightly better than SVM. From this, it can be inferred that not all types of regression, probability and tree based classifiers are suitable for this task. There is a need for either to construct a new algorithm or improvise the existing algorithms such as GBC for better level of accuracy and precision. But before making a decision for making a new algorithm, a deeper level of investigation can be done on the best performance so that a valid reason for making a new algorithm can be taken in due course of this research work. In the next subsection, finalization of the classifier pipeline is done. This involves a hyper-parameter tuning of the GBC algorithm along with validation using 10 K-fold. Technically it includes: selection of the appropriate parameters, their ranges and running a grid search for finding the best possible combination of parameter values. The grid search process involves running a GBC algorithm with various permutations and combinations of parameter values to find a point at which the accuracy, recall, and other performance metrics do not further yield to higher levels of performance. This way a new machine learning model is constructed which is optimized, numerically stable and consistent with its performance as it is also validated using a 10K fold method.

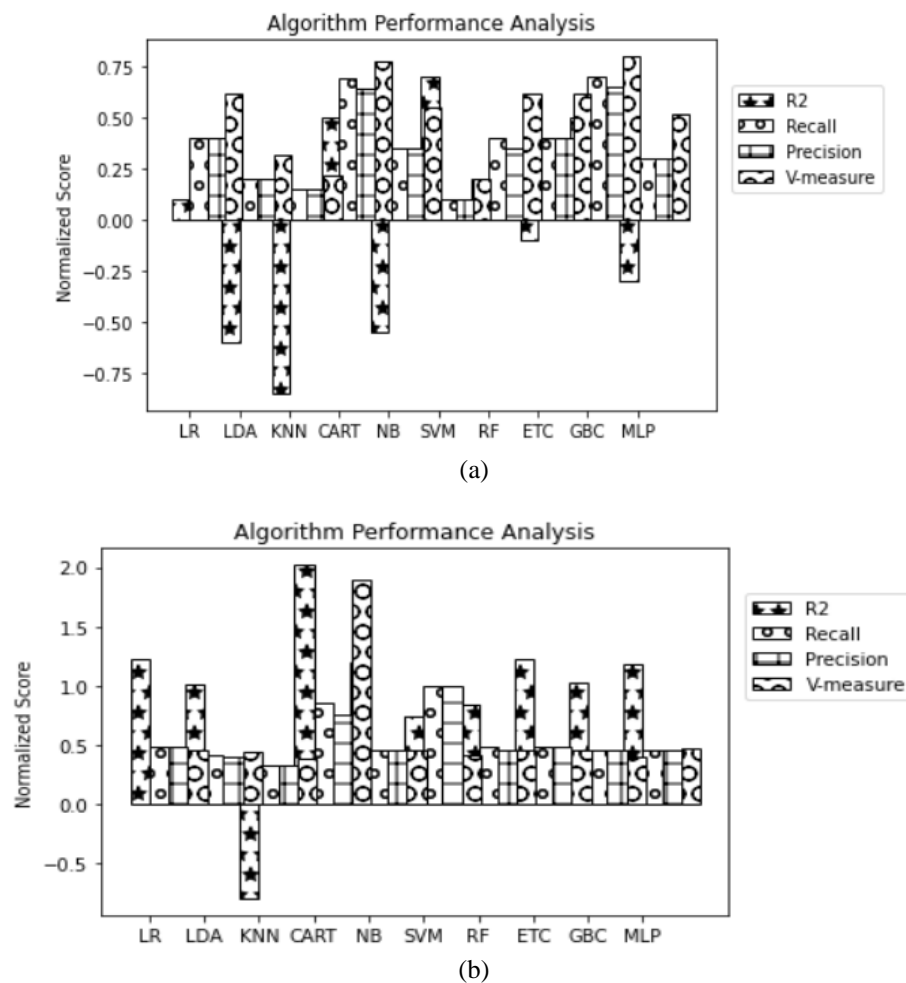


Figure 4. Impact of R2, recall, precision and V-measure (a) performance analysis of classifiers and (b) performance analysis of classifiers (with standard deviation)

3.1. Finalization of pipeline

It is apparent from the above hyper parameter tuning process using grid search method [54] and analysis that GBC is the best performer in terms of recall, precision, accuracy, and v-measure scores. At each step, we are using a 10K fold cross validation method to arrive at a decision. The 10 K-fold validation provides average values of the metrics using the inter quartile range method. The advantage is unbiased average calculation based on ten rounds of evaluation. From the scaling impact analysis, it has become apparent that normalization and standard scaling does not help, in-fact robust scaling or non-linear transformation (power) helped to improve the accuracy level of the classifiers. It was also found that train-test data split ratio also influences the accuracy of the classifiers. Again, through the process of 10 K-fold cross validation process it was found that a minimum of 0.85% available data is required for building a reliable classifier. Finally, the recall, precision, accuracy and v-measure score were computed using a 10 K-fold cross validation process to check if the GBC algorithm is best among these algorithms. In this section, prediction quality of the model is checked using different hyper-parameters. There are three kinds of hyper parameter variables that impact the working of gradient boosting performance. The first category is tree specific parameters that affect the performance of each individual tree and the second one impact the performance of operations of boating trees. Last but not the least, is miscellaneous parameters that support the overall performance of the GB algorithm. For fine tuning the algorithm specific range or value are given input to the algorithm and grid search method is involved based on specific goals. In the context of this optimization of GB algorithms the goal is to find optimized values of parameters which would yield the highest possible level of accuracy in classification. The Table 1 (in Appendix) gives details on all the hyper parameter options available and the optimized value found by the grid search method of fine tuning algorithms.

It can be observed from Table 1 that the lower learning rate (0.15) with minimum number of features (sqrt) is the best possible combination to get the maximum accuracy. The hyper-tuning process showed that the exponential loss function provides better moderation and accuracy than the 'deviance' function. It can also be observed that the tree does not have to be too deep for getting maximum level of accuracy, as the optimized value is three. Since, GBC is a combination algorithm that takes feedback from multiple classifier models, the determination of the number of models required to get maximum accuracy is also important. Higher number of estimators means high level of overload and higher level of prediction latency. Nevertheless, it can be observed that it took 10 estimators to get a 0.96 level of accuracy. Adding more estimators did not yield higher order of accuracy. It implies that GBC may require some kind of modification or improvisation further to improve performance from this benchmark.

4. RESULTS AND DISCUSSION

The purpose of this research work was to build a pipeline of methods that directly help to detect DR. We started with the exploration and evaluation of 10 algorithms for this purpose. All these algorithms were given training with the fundus image's image geometric features. Each eye image was analyzed at small region levels. Using the region prop method, the geometric properties of each region were extracted and compiled to form a feature dataset from the machine learning modeling. The preliminary evaluation of the entire algorithm set showed that gradient boosting algorithm is the best performer in terms of recall, precision, and accuracy. Hence, further exploration was done so as to ascertain the numerical stability and reproducibility of the GBC algorithm by using a 10K fold cross validation method. It can be observed from Figure 5 that the accuracy of the algorithm is improved by approximately 6%.

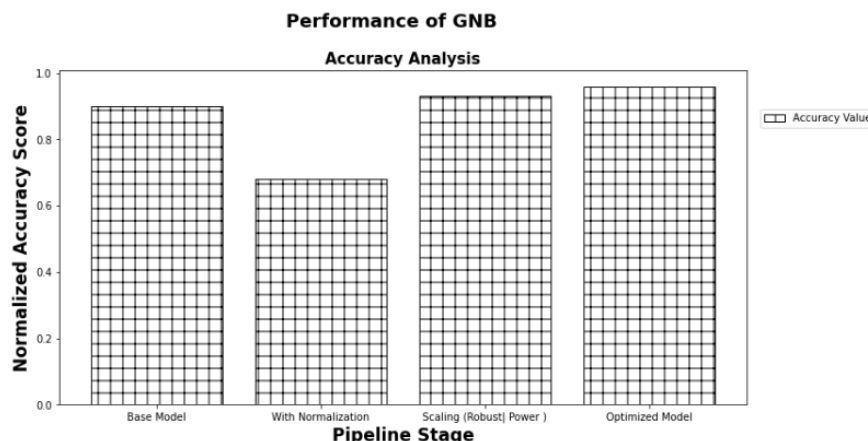


Figure 5. Accuracy levels with respect to pipeline (base model=0.901, with normalization=0.68, scaling (robust|power)=0.93, optimized model=0.96)

This research demonstrates that the best method for building an accurate DR classifier depends upon the type of scaling done on the dataset and the level of optimization we can do on the gradient boosting algorithm. Normalization does not help; in fact, it leads to deterioration of the performance of almost all the classifier algorithms. In the case of the gradient boosting classifier, the accuracy fell from 0.90 to 0.68 due to normalization. Nevertheless, the use of robust or power scaling improved the accuracy of the gradient boosting classifier, up to 0.93. The size of the dataset was small, due to which our strategy was to add more features or keep the same number of features so that maximum information can be provided to the machine learning models for getting trained. Hence, 139 features were used for evaluation of all the ten algorithms. Keeping this in mind, the task of optimization of GBC was done. The gradient boosting algorithm gives nine parameters for optimizing the performance of the base model. Full grid search method was used for finding most optimized hyper-parameters. The diagram in Figure 6 summarizes the pipeline of building DR classifiers.

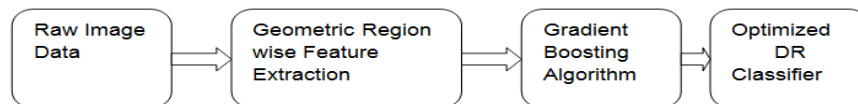
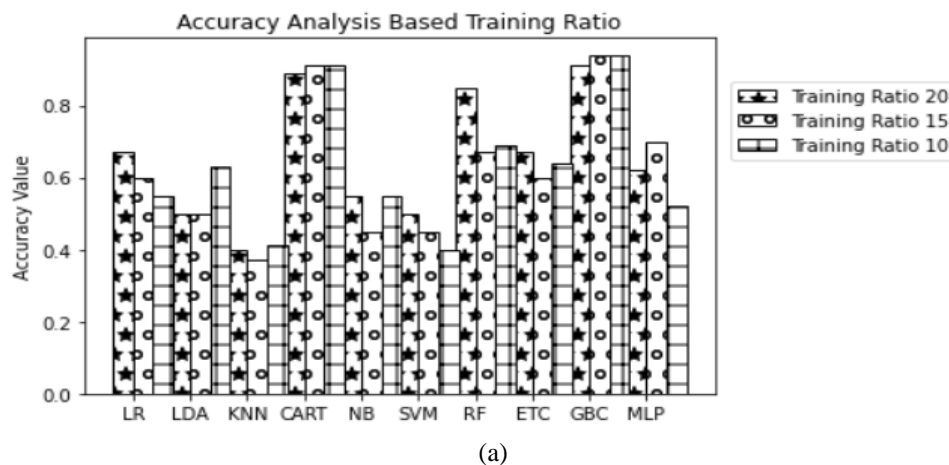


Figure 6. Final pipeline for building DR classifier

5. CONCLUSION

An exploratory experimentation regime was followed and it was found the GBC algorithm is the most appropriate machine-learning model for building a classifier that can detect DR. For selection, the best performing and numerically stable model, a deeper investigation on the internal working of GBC models was done and it was found that some level of hyper-parameter tuning is required. The hyper-parameter tuning helped to achieve an accuracy level of 0.96. This work demonstrates that by following the appropriate pipeline of methods a highly accurate algorithm can be constructed. This work also showed that image geometric features can be effectively used for building a DR classifier, especially, when the dataset is small. For future directions, it is suggested that by understanding the nature of the decision rules, the GB algorithm can further be customized to suit the diabetic retinopathy classification. It is suggested that the pipeline of classifiers built in this context, be renewed with larger volume and size of labeled dataset.

APPENDIX



(a)

Figure 3. Impact of training data size on accuracy (a) accuracy analysis based training ratio

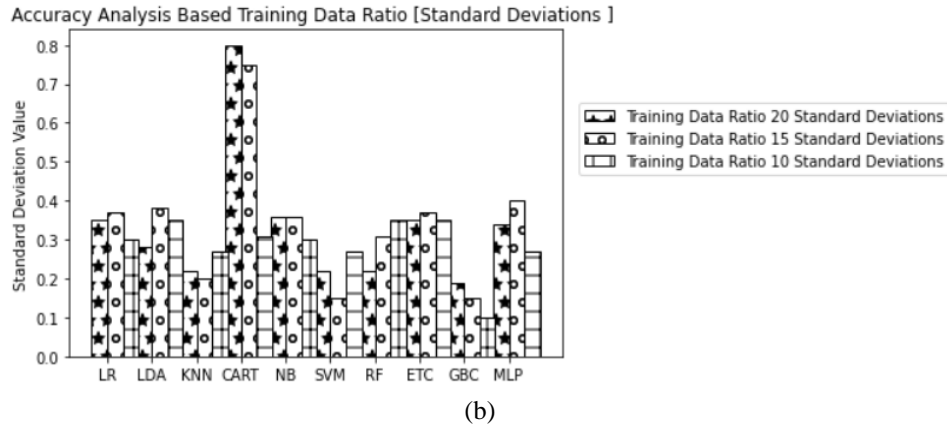


Figure 3. Impact of training data size on accuracy (b) accuracy analysis based training ratio (standart deviations) (*continue*)

Table 1. Gradient-boosting parameters

Sr. No.	Parameter	How it helps to tune	Input	Optimized value	Interpretation
1	Criteria	Check the quality of the split. Mean squared error, mean absolute error and frieman are supported values by GBN.	Friedman Mae, Mae	friedman _mse	This function is most appropriate for finding optimal solutions
2	Learning rate	Determines the impact of each tree on the overall performance of GBM. Lower value means higher degree of robustness and vice versa, secondly it determines the trade-off between learning rate and no of estimators.	{start, end=0.15 -2}	0.15	The algorithm found that maximum of 10 estimators at a level of 0.15 learning rate is best value for getting maximum accuracy
3	Loss function	The loss function to be optimized (price to be paid for get maximum possible accuracy), there are two possible options in GNB, deviance (optimize using logistic regression with probabilistic output) and exponential (a convex method that grows in such manner it makes algorithm more sensitive to extreme values)	{multino mial deviance loss function, exponenti al loss function}	exponen tial	Minimization of the exponential loss function yields a optimized solution
4	Max depth	Maximum depth of the tree. Deeper depth means that higher levels of overhead	{values= 3,5,8}	3	This implies that 3 is the maximum limit for creating nodes. 5 or 8 are not viable
5	Max features	It is the minimum number of features that should be considered while making splits. Lower number implies less overhead.	As thumb rule square root of the total features works well	sqrt	This implies that
6	Minimum Sample leaf: ceil(min_samples_ leaf * n_samples)	Minimum number of observations required in terminal node/leaf. If the data is not imbalanced, higher values are expected.	{0.1,0.2, 0.3}	0.20	This implies for leaf nodes to be created a minimum of 20% samples are required.
7	Minimum sample split ceil(min_samples_ split * n_samples)	It is the minimum number of observations that are required by node for further data splitting. It helps to control over/under-fitting.	{start, end= 0.10, 0.20}	0.13	The value is rounded off with the ceil function and then a number of observations are computed.
8	Number of Estimators	Number of sequential trees. Higher values means large overhead. It works on the principle of ensemble i.e combines a set of weak learners for delivering prediction accuracy but at the same time we do not want too much overhead. 5 is the minimum number of estimators and 15 is the maximum number.	{start, end= 5, 10}	10	Value of 10 means that a GB classifier requires at least 10 estimators for getting maximum possible level of accuracy.
9	Subsample	Fraction of observations to be selected for each tree. Optimal values are required so that accuracy is not compromised.	0.9,0.5,0. 33	0.9cc	This means at least 90% samples are required for building a tree.




REFERENCES

- [1] S. Roy and A. Ramadoss, "Updated insight into COVID-19 disease and health management to combat the pandemic," 2021, pp. 3–39, doi: 10.1016/B978-0-323-85780-2.00017-2.
- [2] V. Rao *et al.*, "COVID-19: An insight into the developments in diagnostics and therapeutics in India," *Indian Journal of Medical Sciences*, vol. 72, no. 2, pp. 77–82, August 2020, doi: 10.25259/IJMS_152_2020.
- [3] A. K. Singh and A. Misra, "Impact of COVID-19 and comorbidities on health and economics: Focus on developing countries and India," *Diabetes & Metabolic Syndrome Clinical Research & Reviews*, vol. 14, no. 6, pp. 1625–1630, November–December 2020, doi: 10.1016/j.dsx.2020.08.032.
- [4] R. Unnikrishnan and A. Misra, "Infections and diabetes: Risks and mitigation with reference to India," *Diabetes & Metabolic Syndrome Clinical Research & Reviews*, vol. 14, no. 6, pp. 1889–1894, November–December 2020, doi: 10.1016/j.dsx.2020.09.022.
- [5] A. Mithal *et al.*, "High prevalence of diabetes and other comorbidities in hospitalized patients with COVID-19 in Delhi, India, and their association with outcomes," *Diabetes & Metabolic Syndrome Clinical Research & Reviews*, vol. 15, no. 1, pp. 169–175, January–February 2021, doi: 10.1016/j.dsx.2020.12.029.
- [6] S. Mor, P. Saini, S. K. Wangnoo, and T. Bawa, "Worldwide spread of COVID-19 Pandemic and risk factors among Co-morbid conditions especially Diabetes Mellitus in India," *Research Journal of Pharmacy and Technology*, vol. 13, no. 5, pp. 2530–2532, 2020, doi: 10.5958/0974-360X.2020.00450.3.
- [7] D. Agarwal, A. Kumar, and A. Kumar, "Commentary: Understanding diabetic retinopathy trends in India: Lessons learnt and future implications," *Indian Journal of Ophthalmology*, vol. 69, no. 11, pp. 3101–3102, Nov. 2021, doi: 10.4103/ijo.IJO_757_21.
- [8] Early Treatment Diabetic Retinopathy Study Research Group, "Early Photocoagulation for Diabetic Retinopathy: Early Treatment Diabetic Retinopathy Study (ETDRS) Report Number 9," *Ophthalmology*, vol. 98, no. 5, pp. 766–785, May 1991, doi: 10.1016/S0161-6420(13)38011-7.
- [9] S. M. Saleem, L. R. Pasquale, P. A. Sidoti, and J. C. Tsai, "Virtual Ophthalmology: Telemedicine in a COVID-19 Era," *American Journal of Ophthalmology*, vol. 216, pp. 237–242, August 2020, doi: 10.1016/j.ajo.2020.04.029.
- [10] D. M. Zulman and A. Verghese, "Virtual Care, Telemedicine Visits, and Real Connection in the Era of COVID-19: Unforeseen Opportunity in the Face of Adversity," *Jama*, vol. 325, no. 5, pp. 437–438, Feb 2021, doi: 10.1001/jama.2020.27304.
- [11] M. Adeli and W. R. Bloom, "Implementing Telemedicine Visits in an Underserved Ophthalmology Clinic in the COVID-19 Era," *Journal of Primary Care & Community Health*, vol. 12, pp. 1–6, Jan. 2021, doi: 10.1177/2150132721996278.
- [12] A. Salma, A. Bustamam, and D. Sarwinda, "Diabetic Retinopathy Detection Using GoogleNet Architecture of Convolutional Neural Network Through Fundus Images," pp. 1–6, 2021, doi: 10.11594/nstp.2021.0701.
- [13] A. Valizadeh, S. J. Ghouschi, R. Ranjbarzadeh, and Y. Pourasad, "Presentation of a Segmentation Method for a Diabetic Retinopathy Patient's Fundus Region Detection Using a Convolutional Neural Network," *Computational Intelligence and Neuroscience*, vol. 2021, Jan 2021, doi: 10.1155/2021/7714351.
- [14] A. Bilal, G. Sun, and S. Mazhar, "Survey on recent developments in automatic detection of diabetic retinopathy," *Journal Français D Ophthalmologie*, vol. 44, no. 3, pp. 420–440, March 2021, doi: 10.1016/j.jfo.2020.08.009.
- [15] V. S and V. R, "A Survey on Diabetic Retinopathy Disease Detection and Classification using Deep Learning Techniques," *2021 Seventh International conference on Bio Signals, Images, and Instrumentation (ICBSII)*, 2021, pp. 1-4, doi: 10.1109/ICBSII51839.2021.9445163.
- [16] "A Comprehensive Survey On Deep Learning Based Automated Diabetic Retinopathy Detection," *Journal of Xidian University*, vol. 14, no. 7, pp. 1995–2006, 2020, doi: 10.37896/jxu14.7/231.
- [17] Q. H. Nguyen, "Diabetic Retinopathy Detection using Deep Learning," *ICMLSC 2020: Proceedings of the 4th International Conference on Machine Learning and Soft Computing*, 2020, pp. 103–107, doi: 10.1145/3380688.3380709.
- [18] A. K. Dash and P. Mohapatra, "A Survey on Prematurity Detection of Diabetic Retinopathy Based on Fundus Images Using Deep Learning Techniques," 2021, pp. 140–155, doi: 10.4018/978-1-7998-5071-7.ch006.
- [19] S. Basu, S. Mukherjee, A. Bhattacharya, and A. Sen, "Segmentation of Blood Vessels, Optic Disc Localization, Detection of Exudates, and Diabetic Retinopathy Diagnosis from Digital Fundus Images," vol. 1355, pp. 173–184, 2021, doi: 10.1007/978-981-16-1543-6_16.
- [20] P. Nage and S. Shitole, "A Survey on Automatic Diabetic Retinopathy Screening," *SN Computer Science*, vol. 2, no. 6, pp. 1–12, Aug. 2021, doi: 10.1007/s42979-021-00833-z.
- [21] S. B. Koca, M. Akdogan, and S. Koca, "Evaluation of early retinal vascular changes by optical coherence tomography angiography in children with type 1 diabetes mellitus without diabetic retinopathy," *International Ophthalmology*, Oct. 2021, pp. 1–11, doi: 10.1007/s10792-021-02059-7.
- [22] Y.-J. Chen, A. S. Khouri, M. A. Zarbin, and B. C. Szirth, "Early retinal microvascular abnormalities in young adults with type 1 diabetes mellitus without clinically evident diabetic retinopathy," *Retina*, vol. 41, no. 7, pp. 1478–1486, July 2021, doi: 10.1097/IAE.0000000000003047.
- [23] Y. Hatanaka, "Retinopathy Analysis Based on Deep Convolution Neural Network," *Advances in Experimental Medicine and Biology*, vol. 1213, pp. 107–120, Jan. 2020, doi: 10.1007/978-3-030-33128-3_7.
- [24] B. S. Tchinda, D. Tchiotso, M. Noubom, V. Louis-Dorr, and D. Wolf, "Retinal blood vessels segmentation using classical edge detection filters and the neural network," *Informatics in Medicine Unlocked*, vol. 23, p. 100521, 2021, doi: 10.1016/j.imu.2021.100521.
- [25] E. Abdelmaksoud, S. El-Sappagh, S. Barakat, T. Abuhmed, and M. Elmogy, "Automatic Diabetic Retinopathy Grading System Based on Detecting Multiple Retinal Lesions," in *IEEE Access*, vol. 9, pp. 15939–15960, 2021, doi: 10.1109/ACCESS.2021.3052870.
- [26] R. K. Revindran and M. N. Giriprasad, "A Review on Automatic Detection of Retinal Lesions in Fundus Images for Diabetic Retinopathy," 2020, pp. 177–202, doi: 10.1007/978-981-15-6141-2_10.
- [27] A. Pradhan, B. Sarma, R. K. Nath, A. Das, and A. Chakraborty, "Diabetic Retinopathy Detection on Retinal Fundus Images Using Convolutional Neural Network," *Machine Learning, Network Security and Data Sciences. MIND 2020. Communications in Computer and Information Science*, Springer, Singapore, 2020, pp. 254–266, doi: 10.1007/978-981-15-6315-7_21.
- [28] S. Cai and T. Y. A. Liu, "The Role of Ultra-Widefield Fundus Imaging and Fluorescein Angiography in Diagnosis and Treatment of Diabetic Retinopathy," *Current Diabetes Reports*, vol. 21, no. 9, pp. 1–6, Aug. 2021, doi: 10.1007/s11892-021-01398-0.
- [29] D. A. Salz and A. J. Witkan, "Imaging in diabetic retinopathy," *Middle East African Journal of Ophthalmology*, vol. 22, no. 2, pp. 145–150, doi: 10.4103/0974-9233.151887.





- [30] A. Pujari *et al.*, "Clinical Role of Smartphone Fundus Imaging in Diabetic Retinopathy and Other Neuro-retinal Diseases," *Current Eye Research*, vol. 46, no. 11, pp. 1605–1613, July 2021, doi: 10.1080/02713683.2021.1958347.
- [31] Y. K. Gan and A. Samsudin, "Verification of 3D-printed universal smartphone retinal imaging adapter against conventional fundus camera imaging for diabetic retinopathy screening," *Malaysian Journal of Ophthalmology*, vol. 2, no. 3, pp. 190–202, Sep. 2020, doi: 10.35119/myjo.v2i3.124.
- [32] A. Markan, A. Agarwal, A. Arora, K. Bazgain, V. Rana, and V. Gupta, "Novel imaging biomarkers in diabetic retinopathy and diabetic macular edema," *Therapeutic Advances in Ophthalmology*, vol. 12, pp. 1–16, 2020, doi: 10.1177/2515841420950513.
- [33] L. Dai *et al.*, "A deep learning system for detecting diabetic retinopathy across the disease spectrum," *Nature Communications*, vol. 12, no. 1, p. 3242, May 2021, doi: 10.1038/s41467-021-23458-5.
- [34] M. Al-Smadi, M. Hammad, Q. B. Baker, and S. A. Al-Zboon, "A transfer learning with deep neural network approach for diabetic retinopathy classification," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 11, no. 4, pp. 3492–3501, August 2021, doi: 10.11591/ijece.v11i4.pp3492-3501.
- [35] A. Erciyas and N. Barışçı, "An Effective Method for Detecting and Classifying Diabetic Retinopathy Lesions Based on Deep Learning," *Computational and Mathematical Methods in Medicine*, vol. 2021, 2021, doi: 10.1155/2021/9928899.
- [36] A. Galdran, G. Carneiro, and M. A. G. Ballester, "Balanced-MixUp for Highly Imbalanced Medical Image Classification," vol. 12905, pp. 323–333, 2021, doi: 10.1007/978-3-030-87240-3_31.
- [37] L. Gao, L. Yang, D. Arefan, S. Wu, T. M. Deserno, and P.-H. Chen, "One-class classification for highly imbalanced medical image data," *Medical Imaging 2020: Imaging Informatics for Healthcare, Research, and Applications*, vol. 11318, p. 113181C, 2020, doi: 10.1117/12.2551389.
- [38] P. Datta, "Classification_features_DR_dataset," *Mendeley Data*, August 2020, doi: 10.17632/77wffjyxdc.2.
- [39] P. Datta, S. Rani, and D. Koundal, "Detection of Eye Ailments Using Segmentation of Blood Vessels from Eye Fundus Image," vol. 597, pp. 515–531, 2019, doi: 10.1007/978-3-030-29407-6_37.
- [40] G. Balkys and D. Dzemida, "Segmenting the eye fundus images for identification of blood vessels," *Mathematical Modelling and Analysis*, vol. 17, no. 1, pp. 21–30, Feb. 2012, doi: 10.3846/13926292.2012.644046.
- [41] J. Dash and N. Bhoi, "Retinal blood vessel segmentation using Otsu thresholding with principal component analysis," *2018 2nd International Conference on Inventive Systems and Control (ICISC)*, 2018, pp. 933–937, doi: 10.1109/ICISC.2018.8398938.
- [42] X. Yan and L. Jiao, "Non-local Three-dimensional Otsu Image Thresholding Segmentation Based on Anisotropic Adaptive Gaussian Weighted Window," *Journal Of Electronics Information Technology*, vol. 34, no. 11, pp. 2672–2679, July 2012, doi: 10.3724/SP.J.1146.2012.00859.
- [43] D. Dobkin and D. Gunopulos, "Geometric problems in machine learning," *Workshop on Applied Computational Geometry*, Springer, Berlin, Heidelberg, vol. 1148, pp. 121–132, 1996, doi: 10.1007/BFb0014490.
- [44] M. Manjramkar, "Survey of Diabetic Retinopathy Screening Methods," *2018 2nd International Conference on Trends in Electronics and Informatics (ICOEI)*, 2018, pp. 1–6, doi: 10.1109/ICOEI.2018.8553843.
- [45] R. Shalini and S. Sasikala, "A Survey on Detection of Diabetic Retinopathy," *2018 2nd International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC) I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 2018 2nd International Conference on*, 2018, pp. 626–630, doi: 10.1109/I-SMAC.2018.8653694.
- [46] H. H. Dewey and D. R. DeVries, "Predictive Performance of Machine Learning Algorithms Trained with Sparse Data," *2021 IEEE Aerospace Conference (50100)*, 2021, pp. 1–7, doi: 10.1109/AERO50100.2021.9438373.
- [47] G. E. P. Box and D. R. Cox, "An Analysis of Transformations Revisited, Rebutted," *Journal of the American Statistical Association*, vol. 77, no. 377, pp. 209–210, March 1982.
- [48] I.-K. Yeo, "A new family of power transformations to improve normality or symmetry," *Biometrika*, vol. 87, no. 4, pp. 954–959, December 2000, doi: 10.1093/biomet/87.4.954.
- [49] H. Alibrahim and S. A. Ludwig, "Hyperparameter Optimization: Comparing Genetic Algorithm against Grid Search and Bayesian Optimization," *2021 IEEE Congress on Evolutionary Computation (CEC)*, 2021, pp. 1551–1559, doi: 10.1109/CEC45853.2021.9504761.
- [50] L. Hertel, P. Baldi, and D. L. Gillen, "Reproducible Hyperparameter Optimization," *Journal of Computational and Graphical Statistics*, pp. 1–16, 2021, doi: 10.1080/10618600.2021.1950004.
- [51] S. Arndt, T. Cizadlo, D. O'Leary, S. Gold, and N. C. Andreasen, "Normalizing counts and cerebral blood flow intensity in functional imaging studies of the human brain," *NeuroImage*, vol. 3, no. 3, pp. 175–84, June 1996, doi: 10.1006/nimg.1996.0019.
- [52] X. Giam and J. D. Olden, "A new R2-based metric to shed greater insight on variable importance in artificial neural networks," *Ecological Modelling*, vol. 313, pp. 307–313, Oct. 2015, doi: 10.1016/j.ecolmodel.2015.06.034.
- [53] H. R. Sofaer, J. A. Hoeting, C. S. Jarnevich, and J. McPherson, "The area under the precision-recall curve as a performance metric for rare binary events," *Methods in Ecology and Evolution*, vol. 10, no. 4, pp. 565–577, 2019, doi: 10.1111/2041-210X.13140.
- [54] L. Yang and A. Shami, "On hyperparameter optimization of machine learning algorithms: Theory and practice," *Neurocomputing*, vol. 415, pp. 295–316, Nov. 2020, doi: 10.1016/j.neucom.2020.07.061.

BIOGRAPHIES OF AUTHORS







Parul Datta    is pursuing Doctorate from Chitkara University, Himachal Pradesh and has done M.E in Computer Science and Engineering from Panjab University, Chandigarh, India. She is currently Assistant Professor in Computer Science and Engineering department at Chitkara University, Himachal Pradesh, India and has academic teaching experience of more than 5 years with 7 publications in reputed books and conferences. Her areas of research include image processing and machine learning. She has filed 3 patents till date. She can be contacted at email: parul.datta@chitkarauniversity.edu.in.



Prasenjit Das     has a doctorate in Computer Science. He is currently Head and Associate Professor in Computer Application Department in Chitkara University, Himachal Pradesh. He has more than 15 years of experience in industry as well as Academics. He has filed a number of patents and has publications in a number of journals. His areas of research are data mining, machine learning and image processing. He can be contacted at email: prasenjit.das@chitkarauniversity.edu.in.



Abhishek Kumar     is Doctorate in Computer Science from University of Madras and has done M.Tech in Computer Science and Engineering from Government engineering college Ajmer, Rajasthan Technical University, Kota, India. He has total Academic teaching experience of more than 7 years with more than 80 publications in reputed, peer reviewed National and International Journals, books and conferences. He can be contacted at email: abhishek.kumar@chitkara.edu.in.