

Dr Miner: An Application of Auto Detecting Diabetic Retinopathy using Auto Colour Correlogram and Bagging

Chew-Wai Yap¹, Kai-Jie Lim², Keng-Hoong Ng³, Kok-Chin Khor⁴

^{1,2,3}Faculty of Computing and Informatics, Multimedia University, Cyberjaya, Malaysia.

⁴Department of Internet Engineering and Computer Science, Lee Kong Chian Faculty of Engineering and Science, Universiti Tunku Abdul Rahman, Sungai Long, Kajang, Malaysia.

Corresponding author: ²kckhor@utar.edu.my

Article History: Received: 10 November 2020; Revised: 12 January 2021; Accepted: 27 January 2021;

Published online: 05 April 2021

Abstract: An application of auto-detecting Diabetic Retinopathy (DR) is indispensable to aid the ophthalmologists in diagnosing patients and also to help relevant organisations in accumulating and analysing data. This project presents DR Miner, an application that can extract data from fundus images, identify the symptoms of DR in retina images by using data science approaches, and collect the ophthalmologist's review to improve the detection model in the future. To form the DR data set with binary classes, Auto Colour Correlogram (ACC) was utilised to extract the features from DR images. Over-sampling was then conducted to balance the class distribution in the data set. To reduce the variance of the single learning algorithms, we evaluated various bagging approaches. The results showed that the bagging approaches gave better results than the single learning algorithms in general. Out of all bagging approaches we evaluated, bagged k-nearest neighbours gave the best result. The sensitivity achieved was 85.1%, which met the requirement set by the UK National Institute for Clinical Excellence.

Keywords: Bagging, Auto Colour Correlogram, Diabetes Retinopathy.

1. Introduction

The eyes are the visual system of the human being. It is a sensitive, complex, and also the weakest part of the human body. Therefore, extreme care is needed for the eyes. Diabetic Retinopathy (DR) is a diabetes hurdle that brings inconvenience to the eyes. It is a metabolic disease caused by the high level of blood sugar, which leads to eye damage over time (Who.Int, 2019). DR will occur in anyone with Diabetic Mellitus (DM), either type 1 or type 2 diabetes (Harnett et al., 2017).

DR is a global common cause of blindness among adults (Cheung et al., 2010). DR has impacted worldwide significantly as the number of people suffered from DR will grow to 191.0 million by the year 2030 (Zheng et al., 2012).

Also, by the year 2030, the World Health Organisation estimates that about 2.48 million Malaysians would suffer from DM (MoH, 2017). The Ministry of Health Malaysia finds that within 20 years of DM diagnosis, nearly two-third of Malaysians with DM are diagnosed with some degrees of DR. DR is the most common cause of visual loss in Malaysia, ranked second after cataract. Patients are asymptomatic in the early stage of DR.

In 2010, the Malaysian Government spent about RM 2.4 billion on healthcare related to diabetic diseases, including DR. The spending would expect to be more than RM 3 billion by the year 2020. DM costs 16% of the Malaysian healthcare budget. Such spending makes the country the top 10 in the world in the percentage of the healthcare budget spent on DM (Zhang et al., 2010).

The impact of DR on a country's economy is significant. Unfortunately, there is no way to cure DR. However, via various laser treatments, preventing vision loss before the deterioration of a patient's retina is possible (Moutray et al., 2016). However, the standard argon laser treatment remains vital for treating DR.

Even though DR can be detected through eye tests, however, the process is manual and laborious in many countries, including Malaysia (Hussein et al., 2016; Hussain et al., 2017; Zaki et al., 2016). Therefore, it is a good alternative to use automated systems to detect DR. Automating DR detection leads to not only a more efficient and cost-effective assessment but also to provide a second opinion for the ophthalmologists.

Therefore, we aimed to develop an application named DR Miner to auto-detecting DR using data science approaches in this study, and also to collect reviews of ophthalmologists for building better detection models in the future. We will provide a review of this study in the Literature Review Section, followed by a

methodological approach on how to develop the application in the Methodology Section. The results of utilising various data science approaches shall be discussed in the Results and Discussion Section, and this study shall be concluded in the Conclusion Section.

2. Literature Review

2.1 The Problem of Clinical Information in Malaysia

The majority of clinics in Malaysia still rely partially on manual documentation, including the tracing of investigational results (Hussein et al., 2016). Further, it is not easy to study and compare the efficiency of various clinical information systems in the country. The National Diabetes Registry (NDR) of the Ministry of Health Malaysia helps to gather medical data and perform data analyses, but it is very dependent on human to input information. Therefore, the NDR needs laborious and tedious updating to reflect the actual burden and performance of a clinic accurately.

An automated DR detecting system, as proposed in this study, can help to accumulate useful DR data. Subsequently, the collected can help to enable automatic DR detection using data science approaches.

2.2 The Problem of Decision Support in Malaysia

Screening of DR is necessary to identify the group of patients at risk of visual loss. However, there is a general lack of doctors and specialists in Malaysia to run practical training consistently and this leads to inadequate decision support in DR screening (Hussein et al., 2016). Even though the decision support is possibly available, the quick replacement rate of medical personnel inevitably causes unpredictability and instability to the standardisation of diabetes care.

DR Miner proposed in this study helps in automating decision making to overcome the shortage of workforce.

2.3 Reducing DR Grading Costs

Countries, i.e., Scotland and England, have their established DR detection procedure. Generally, the procedure involves three graders before sending a patient to consult an ophthalmologist. According to the study by Fleming et al. (2008), using automated grading systems can save almost 50% of the costs in grading DR. The research by Tufail et al. (2017) also showed that the automated DR grading systems achieve good specificity and they are cost-effective alternatives to manual grading. Further, the systems also gain acceptable sensitivity for referable retinopathy against the graders.

In this study, we also aimed at developing DR Miner to help to reduce the DR grading costs.

2.4 Detecting DR using Data Science Approaches

Due to the technology innovation, screening of eye diseases, including DR, can be conducted automatically and safely to replace the manual screening (Fleming et al., 2011). Since the '90s, researchers have attempted various approaches for detecting DR automatically, such as mathematical approaches, AI and machine learning.

Early work of researchers used single classifiers. A study by Gardner et al. (1996) used a neural network with backpropagation in detecting DR. The authors used 32 normal and 147 diabetic images for training the neural network. In the study, the authors focused on recognition of diabetic features, i.e., exudates, haemorrhages and vessel from the fundus images.

Sopharak et al. (2008) focused on the detection of DR exudate. However, the authors did a lot of pre-processing before extracting values from the images such as converting the RGB space of the images to HIS, applying a median filter for noise reduction and enhancing the contrast of small regions using histogram equalisation. The detection of exudates was done using mathematical morphology on the fundus images of non-dilated pupils. The method used by the authors was able to reduce the ophthalmologists' workload by detecting the symptoms of DR.

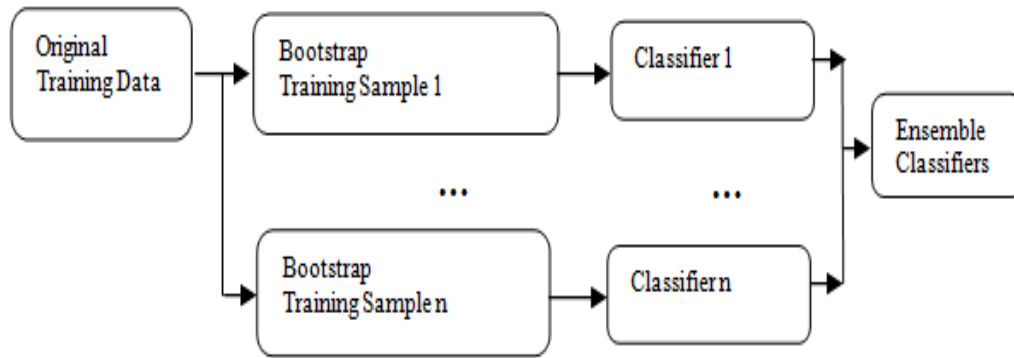


Figure 1. Ensemble classification utilising classifiers that are built based on the bootstrap training data.

To improve the performance of detecting DR, researchers started to consider ensemble classifiers. Ensemble classifiers are a combination of more than one single classifier for accurate classification (Fernández et al., 2018). Using Ensemble classifiers, the classifiers involved are the alternative forms of the same classifier. These base classifiers will classify the same new data sample, and their decisions will be combined or aggregated to produce a final decision. One popular example of Ensemble classifiers is Bagging.

Bagging, or Bootstrap Aggregating, was introduced by Breiman (1996) to reduce variance and to give good stability in classification. The Ensemble method trains classifiers of the same type using new training data that are created using random sampling with replacement from the original training data, as shown in Figure 1. The classifiers are arranged in parallel, and their decisions are aggregated using voting to produce a final decision.

An example of work that used Ensemble classifiers to detect DR was the work by Somasundaram and Alli (2017). In their work, the features such as blood vessels, neural tissue, optic disc size, etc. were extracted from fundus images using t-distributed Stochastic Neighbour Embedding. Then, Bagging was used to detect DR. Bagging gave better results as compared with single classifiers.

Antal and Hajdu (2014) used a few image processing methods to obtain image-level, anatomical components and lesion-specific features. Multiple classifiers were used in the research. Multiple classifiers are different from Ensemble classifiers as multiple classifiers combine or aggregate different type of classifiers instead of the same type (Fernández et al., 2018). The research involved alternating decision tree, k-nearest neighbours, AdaBoost, Multilayer Perceptron, Naïve Bayes, Random Forest, Support Vector Machine, and Pattern Classifier in overcoming the average detection rate of the early research work.

Recently, researchers also attempted deep learning for effective detection of DR. Deep learning build models with multiple layers of networks to learn complex concepts of data (LeCun et al., 2015; Hussain et al., 2016; Goodfellow et al., 2016). UzZaman et al. (2016) used a deep learning method called convolutional neural network with forward propagation, alias CNN to detect multiple severity stages of DR. However, training an effective deep learning model requires a large amount of image data. The authors used 35,126 images to train the model. For evaluation, 53,576 images had been used.

In this study, we used ensemble classifiers for detecting DR as the reviewed work shows that ensemble classifiers are stable and able to reduce variance well as compared with single classifiers.

3. Methodology

As shown in Figure 2, the DR data set was prepared using the Messidor database that facilitates the DR research using computers (Decencièrre et al., 2014). A hundred of high-quality fundus images provided by a hospital in France was downloaded from the database for data set preparation. The images were then converted to JPEG format.

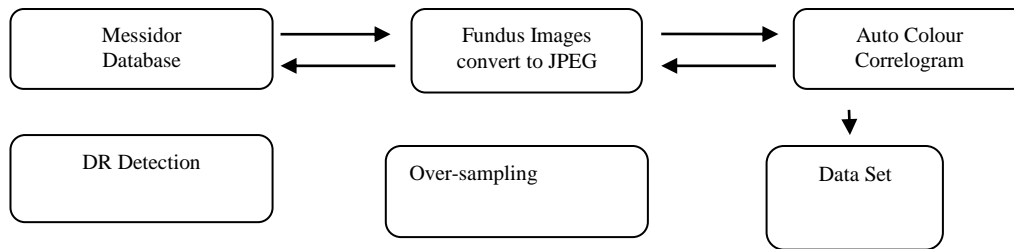


Figure 2.The methodological approach for DR detection.

In this study, a Content-based Image Retrieval (CBIR) technique was used to extract features from the fundus images in JPEG format. CBIR retrieves images that are relevant from databases using pictorial content such as colour, shape, texture, etc. (Chaum et al., 2008). The extracted features from the images are then used for storage, search, and retrieval of images. The analyses of images encompass feature description models, perceptual organisation and spatial relationships for extracting useful information.

The effective and inexpensive CBIR technique used in this study is Auto Colour Correlogram (ACC) (Huang et al., 1997; Huang et al., 1999). The image feature that the technique extract is called colour correlogram; it expresses the spatial correlation of colour pairs corresponding to their distance changes in images. ACC captures spatial correlations between similar colours only. Further, small distances are used because, in an image, global correlations are less significant than local correlations. The size of the features is small and therefore, easy to be computed. ACC is robust against large appearance changes as well as shape changes caused by shifting viewing positions. All these characteristics make ACC a better alternative to the traditional colour histogram approach.

The data set was ready after the ACC stage. However, its class distribution was slightly unbalanced, as shown in Table 1. To enable the learning algorithms to learn DR easily, we grouped all three DR grades into one class called class 1. We then over-sampled the class 0 (normal) 100% using SMOTE (Chawla et al., 2002) so that its size is about equal to class 1 (with DR). With SMOTE, synthetic data were created based on the nearest neighbours of the data points in the data set. We then built detection models using the single learning algorithms and bagging approaches on the data set. The results shall be explained in the next section.

Table 1.The Class Distribution of the Data Set

Class (DR Grades)	Class Distribution (%)
0	33
1	13
2	11
3	43
Total	100

An application called DR Miner was then developed for both ophthalmologists and data scientists. The application is as illustrated in Figure 3 (a) and (b). Using (a), the ophthalmologists uploads a fundus image in the application, and the detection model we embedded in the application will detect for possible DR. An uploaded fundus image can be enlarged by the ophthalmologists for detailed examination, i.e., images of patients with mild DR. Apart from accepting the detection outputs from the application, the ophthalmologists can also be the active learners – the human annotators to label fundus images. Remarks can also be given to indicate the seriousness of DR by stating grades. The more inputs given by the ophthalmologists, the more accurate the detection model. Using (b), data scientists can view the existing image data set, including those appended by the ophthalmologists during active learning. Various data science approaches can be applied using this interface, including building models using the single learning algorithms or bagging approaches evaluated in this study.

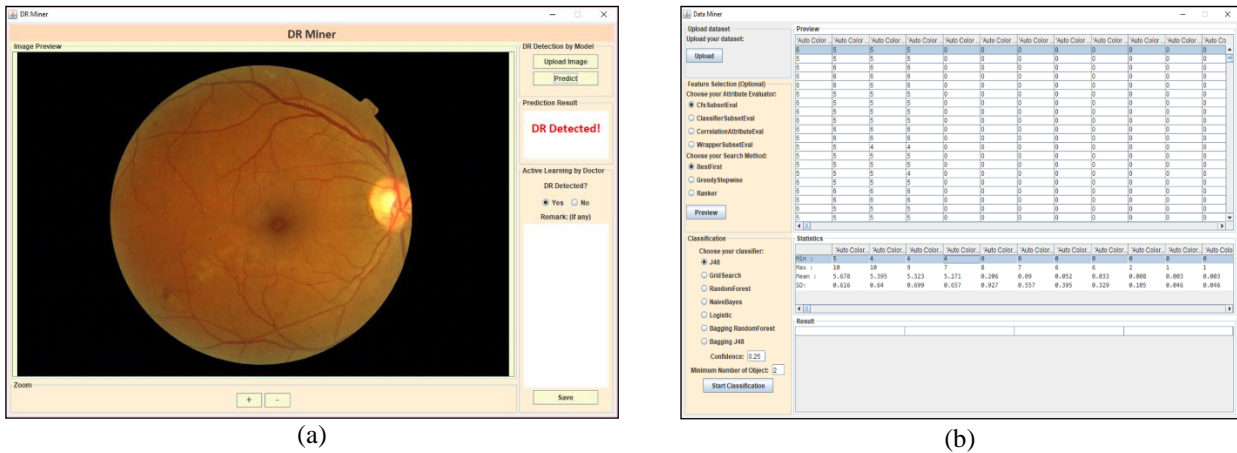


Figure 3. DR Miner contains the graphic user interface for both (a) ophthalmologists, and (b) data scientists.

4. Results and Discussion

Table 2. The Results of Utilising Various Learning Algorithms and Bagging Approaches. Each of the Bolded Numbers is the Best Result of the Evaluation Metrics Used in This Study.

	Sensitivity	Specificity	Accuracy	ROC
Single Learning Algorithms				
C4.5(c=0.1, m=6)	83.6%	80.3%	82.0%	80.4%
NB	77.6%	59.1%	68.4%	69.3%
BN	82.1%	77.3%	79.7%	90.1%
KNN (k=3, L ₁)	80.6%	89.4%	85.0%	86.4%
Bagging				
Random Forrest	74.6%	81.8%	78.2%	89.7%
Bagged LibSVM (cost=1, gamma=0.001)	70.1%	83.3%	76.7%	82.4%
Bagged C4.5 (c=0.1, m=3)	85.1%	81.8%	83.5%	88.0%
Bagged KNN (k=3, L ₁)	85.1%	89.4%	87.2%	91.3%

* c and m are parameters for C4.5 representing confidence factor and number of samples per leaf
 * k is the number of the nearest neighbours, and L₁ is the Manhattan distance used in KNN.

We had evaluated several single learning algorithms and bagging approaches, as shown in Table 2. The parameters of the algorithms were fine-tuned to give optimal performance in DR detection. The results showed that the single learning algorithms gave only average results. On the contrary, the bagging approaches gave relatively better results than the single learning algorithms. The top performer is the bagged K-Nearest Neighbours (bagged KNN), and it gave the best results in these four evaluation metrics, namely, sensitivity, specificity, accuracy and the Receiver Operating Characteristic (ROC). Thus, bagged KNN shall be the primary bagging approach used by DR Miner for detecting DR. The UK National Institute for Clinical Excellence

(NICE) recommends at least 80% of sensitivity of DR screening modalities. Such requirement is also stated by “Clinical Practice Guidelines: Screening of Diabetic Retinopathy” released by the Ministry of Health Malaysia in the year 2011. The bagged KNN used in this study achieved a sensitivity of 85.1%. Other metrics used to measure the bagged KNN also showed satisfactory results, where 89.4%, 87.2%, and 91.3% were achieved for specificity, accuracy and ROC, respectively.

5. Conclusion

In this study, we aim to assist the ophthalmologists in diagnosing patients by building an application capable of auto detecting DR. The application also helps organisations, particularly clinics, in accumulating and analysing fundus photos of DR. This application can solve the problems of Malaysia as stated in the Literature Review Section.

In searching for the best data mining approach for detecting the DR, we conducted the empirical study, as explained in the previous sections. In general, the single learning algorithms gave satisfactory results for detecting DR. To improve the detection results further, we used bagging approaches that can provide good stability and variances lower than the single learning algorithms. Out of all bagging approaches we evaluated, the bagged KNN gave the best results as compared with the other bagging approaches as well as the single learning algorithms.

However, there are rooms for improvement. We are considering deep learning in our future work to elevate further the performance of the application to a higher state. The application will also be developed further to be capable of detecting different grades of DR.

References

1. Antal, B., & Hajdu, A. 2014. An ensemble-based system for automatic screening of diabetic retinopathy. *Knowledge-Based Systems*, 60, 20-27.
2. Breiman, L. 1996. Bagging predictors. *Machine learning*, 24(2), 123-140.
3. Chaum, E., Karnowski, T., Govindasamy, V., Abdelrahman, M., & Tobin, K. 2008. Automated diagnosis of retinopathy by content-based image retrieval. *Retina*, 28(10), 1463-1477.
4. Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. 2002. SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16, 321-357.
5. Cheung, N., Mitchell, P., & Wong, T. 2010. Diabetic retinopathy. *The Lancet*, 376(9735), 124-136.
6. Zheng, Y., He, M. and Congdon, N. 2012. The worldwide epidemic of diabetic retinopathy. *Indian Journal Of Ophthalmology*, 60(5), 428.
7. D. Fleming, A., Philip, S., A. Goatman, K., J. Prescott, G., F. Sharp, P., & A. Olson, J. 2011. The Evidence for Automated Grading in Diabetic Retinopathy Screening. *Current Diabetes Reviews*, 7(4), 246-252.
8. Decenci re, E., Zhang, X., Cazuguel, G., Lay, B., Cochener, B., & Trone, C. et al. 2014. Feedback on a publicly distributed image database: the messidor database. *Image Analysis & Stereology*, 33(3), 231.
9. Fern andez, A., Garc a, S., Galar, M., Prati, R., Krawczyk, B., & Herrera, F. Learning from Imbalanced Data Sets.
10. Gardner, G., Keating, D., Williamson, T., & Elliott, A. 1996. Automatic detection of diabetic retinopathy using an artificial neural network: a screening tool. *British Journal Of Ophthalmology*, 80(11), 940-944.
11. Goodfellow, I., Bengio, Y., & Courville, A. 2016. Deep learning. MIT press.
12. Hartnett, M., Baehr, W., & Le, Y. 2017. Diabetic retinopathy, an overview. *Vision Research*, 139, 1-6.
13. Huang, J., Ravi Kumar, S., Mitra, M., Zhu, W., & Zabi, R. 1997. Image indexing using color correlograms. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. USA.
14. Huang, J., Kumar, S. R., Mitra, M., Zhu, W. J., & Zabi, R. 1999. Spatial color indexing and applications. *International Journal of Computer Vision*, 35(3), 245-268.
15. Hussain, A., Mkpojiogu, E.O.C., Fadzil, N.M., Hassan, N.M. (2017). The UX of a pregnancy app on mobile device. *AIP Conference Proceedings*, 1891, art. no. 020061.
16. Hussain, A., Mkpojiogu, E.O.C., Yusof, M.M. (2016). Perceived usefulness, perceived ease of use, and perceived enjoyment as drivers for the user acceptance of interactive mobile maps. *AIP Conference Proceedings*, 1761, art. no. 020051.
17. Hussein, Z., Wahyu Taher, S., Gilcharan Singh, H., & Siew Swee, W. 2016. Diabetes Care in Malaysia: Problems, New Models, and Solutions. *Annals Of Global Health*, 81(6), 851.

17. LeCun, Y., Bengio, Y., & Hinton, G. 2015. Deep learning. *nature*, 521(7553), 436-444.
18. Ministry of Health Malaysia (MoH). 2017. Diabetic retinopathy screening.
19. Moutray, T., Evans, J., Armstrong, D., & Azuara-Blanco, A. 2016. Different lasers and techniques for proliferative diabetic retinopathy. *Cochrane Database Of Systematic Reviews*.
20. Fleming, A.D., Olson, J.A., Philip, S., Goatman, K.A., Sharp, P.F., Fonseca, S., Prescott, G.J., Sotland, M.G.S. and McNamee, P. 2008. Manual vs. automated: the diabetic retinopathy screening debate. *Ophthalmology Times*, 4(2).
21. Somasundaram, S.K. & Alli, P. 2017. A Machine Learning Ensemble Classifier for Early Prediction of Diabetic Retinopathy. *Journal Of Medical Systems*, 41(12).
22. Sopharak, A., Uyyanonvara, B., Barman, S., & Williamson, T. 2008. Automatic detection of diabetic retinopathy exudates from non-dilated retinal images using mathematical morphology methods. *Computerized Medical Imaging And Graphics*, 32(8), 720-727.
23. Tufail, A., Rudisill, C., Egan, C., Kapetanakis, V., Salas-Vega, S., & Owen, C. et al. 2017. Automated Diabetic Retinopathy Image Assessment Software. *Ophthalmology*, 124(3), 343-351.
24. UzZaman, A., & Kawnain Bashir, S. 2016. Diabetic retinopathy detection using image processing. BRAC University.
25. Who.Int. 2019. Diabetes. Retrieved 3 November 2019, from <https://www.who.int/health-topics/diabetes>
26. Zaki, W., Zulkifley, M., Hussain, A., Halim, W., Mustafa, N., & Ting, L. 2016. Diabetic retinopathy assessment: Towards an automated system. *Biomedical Signal Processing And Control*, 24, 72-82.
27. Zhang, P., Zhang, X., Brown, J., Vistisen, D., Sicree, R., Shaw, J., & Nichols, G. 2010. Global healthcare expenditure on diabetes for 2010 and 2030. *Diabetes Research And Clinical Practice*, 87(3), 293-301.