Int. J. Advance Soft Compu. Appl, Vol. 11, No. 3, November 2019
ISSN 2074-8523

Diabetic Retinopathy Classification using Support Vector Machine with Hyperparameter Optimization

Nur Izzati Ab Kader[a,1], Umi Kalsom Yusof[a,2,] and Syibrah Naim[a,3]

[a]School of Computer Science, Universiti Sains Malaysia, 11800 USM, Pulau Pinang

[1]nurizzati.ucom13@student.usm.my, [2]umiyusof@usm.my, [3]syibrah@usm.my

Abstract

Diabetic Retinopathy (DR) is one of the frequent comorbidities of diabetes worldwide. Diabetic eye screening has become a challenging task for ophthalmologist as they need to deal with large number of patients to be diagnosed, creating a need to develop tool that may help ophthalmologist to classify the severity of DR in order to establish an adequate therapy. Previous researchers have studied machine learning to propose an automatic DR classification using the clinical variables. However, it needs to be improvised especially in terms of accuracy. Hence, this paper aims to propose an optimal or near-optimal DR classifier using the Support Vector Machine with hyperparameter optimization. This study considered three classes of diabetic patients which were patients who do not have DR (NODR), patients with non-proliferative DR (NPDR) and patients with proliferative DR (PDR), instead of focusing only on two classes (NO DR, DR). The radial basis function, polynomial, sigmoid kernel and their respective hyperparameters were tested in this study in order to find the best kernel and combination of hyperparameters that can improve the performance of SVM. The results obtained show that SVM-radial kernel with cost value,64, 0.03 gives the best accuracy at 85.45%.

Keywords*: Diabetic Retinopathy, Classification, Hyperparameter, Optimization, Support Vector Machine*

1 Introduction

The global prevalence of Diabetes Mellitus (DM) has been increasing year on year. According to the statistic from WHO Global Report, the number of diabetes cases has almost quadrupled since 1980 with 108 million patients to 2016 with 422 million [1]. This dramatic rise is largely due to the rise in type 2 diabetes and

among the factors contributing to this steep rise include the increased consumption of high-calorie food, lack of exercise, and increased prevalence of obesity. Diabetic Retinopathy (DR) is the most common microvascular complication of Diabetes Mellitus and it could affect 1 in 3 persons with DM. DR is characterized by damage to the blood vessels of the retina. In the beginning, it may be asymptomatic at an early stage, but without any treatment or delay in treatment can cause permanent blindness to the diabetic patients [2].

With frequent screenings and early treatments to the diabetic patients, it could reduce the chance of vision loss and it could decrease the overall load on health care centers. However, with the increasing number of cases nowadays, abnormal retinal classification becomes a challenging task for ophthalmologist as they need to deal with a large number of patients to be diagnosed on a daily basis. It is quite resource consuming to conduct a yearly screening to all of the patients. The screening tasks are done manually in most countries [3]. The process is carried out through an inspection with naked eyes. The inspections are carried out using an ophthalmoscope to identify relative characteristics such as to differentiate between normal healthy vessels and abnormal vessels. Usually, experts identify relative characteristics such as to differentiate between normal and abnormal retina manually based on their experience which can lead to inconsistency during the grading process [4].

Issue of variability in grading arises from this manual process because the boundaries between the grades may differ between observers [5] and it could also be prone to error [6]. Recently, the information on the screening diabetic patients have been properly recorded. This database is significantly valuable for assessment of the risks in the development of DR. Amongst the solutions which has been proposed by previous researchers is to come out with a DR classification that can help ophthalmologist to ease the grading process. Several DR classifiers have been developed by using clinical variables proposed by previous researchers. However, there is still some space for improvement especially in the accuracy of the classification.

Therefore, this study is proposed to classify DR with the objective to find a DR classifier with optimal or near-optimal performance matrices using Support Vector Machine (SVM) with hyperparameter optimization. SVM is chosen to be adapted in this study to improve the sensitivity and/or specificity of the DR diagnosis. The hyperparameter optimization technique is incorporated into this study in order to increase the SVM performance. There are some significant advantages of this study and they will be elaborated throughout this study. First and foremost, this dataset encompasses three classes of diabetic patients which are patients that do not have DR (NODR), patients with non-proliferative DR (NPDR) and patients with proliferative DR (PDR). Usually, a DR classification is conducted using only two classes which are to classify whether a person is being diagnosed with DR or not. Besides, the clinical variables used in this study are selected by doctors. Thus, there will be no issue on the validity of the features

used. This classification can assist doctors in easing the process of decision-making regarding the type and medication to be prescribed. Thus, unnecessary testing and checkups can be prevented.

2 Related Work

In the area of DR classification, the researchers have studied DR with different intelligent approaches and aims. A few studies have been conducted to develop DR classification using clinical variables by adopting machine learning. Previously, a clinical decision support system (CDSS) for DR has been developed by [2] using random forest, logistic regression, decision trees and ensemble models. It was built from a demographic and lab data with the aim to detect patient's susceptibility to retinopathy.

Besides, another study was performed on ensemble classifiers to determine whether a patient is at risk of developing DR. They explored the use of two kinds of ensembles: dominance-based rough set balanced rule ensemble and fuzzy random forest ensemble. The study employed the clinical variables which represents main risk factors to perform the pre-diction. In another study, [7] performed a study on prediction of DR using Naive Bayes [8]. A predictive system has been developed to predict prevalence of DR in Malaysia using 140 diabetic patients by adopting a voting mechanism to select the final results of Decision Tree and Case Base Reasoning. [9] built DR classifier to predict the risk of DR using data from 55 type 1 diabetes patients. They applied Neural network, Classification and Regression Tree (CART), Hybrid Wavelet Neural Network (HWNN), classification-based Rule Induction with C5.0 and merged their result using a voting mechanism.

While these machine learnings have been adopted in some form, they are limited in several ways. First and foremost, most of the studies done in the past focused only on two classes (NO DR, DR) which is still general for DR grading. Very little work had been done with regards to DR classification focusing on three classes of DR (NO DR, NPDR, PDR). Secondly, the current accuracy yield for DR classification is still low and needs to be improvised. To the best of our knowledge, the highest accuracy recently yielded by [2] with an overall accuracy of 92.76%. However, this result only concerns the two classes of DR. Therefore, this study fills the gap in the literature by proposing an algorithm (SVM with hyperparameter optimization) with the motivation to improve upon the results.

2.1. Support Vector Machine

SVM is a classifier introduced by Corinna Cortes and Vladimir Vapik [10]. It is a mathematical model with a learning routine used for classification of input data received by a computing system and also for regression task. SVM works by generating a hyperplane to discriminate between two classes after the input data has been transformed into high-dimensional space with objective to maximizing

the margin between classes [10, 11]. The specialty of SVM is that it possess a good generalization, robust to noisy data and can efficiently perform non-linear classification using the kernel trick. The kernel-trick function is to allows the construction of a classifier without the feature space [12]. In real world application problem, there are different types of data. It might be present in the case of linear or non-linear.

Linear case: In linear case, SVM deals with binary classification problem, with the goal to find the optimum hyperplane that is represented through Equation (1).

$$H = w \,.\, x + b = 0 \tag{1}$$

Based on the equation, w is a vector of the hyperplane and x represents the data and b is a bias added to the hyperplane. In order to obtain an optimum hyperplane for the class yi 2 -1,+1, the margin has to be maximized and the error has to be minimized. -1 denoted training example, x for the first class and +1 denoted for the other class. It can be achieved by searching through Lagrange multiplier, and the problem can be formulated as Equation (2).

$$f(x) = sign\ (\sum_{i=1}^{N} yi.Li < x.xi > +b) \tag{2}$$

Li denoted the Lagrange multiplier. Lagrange multiplier is a strategy to find the local minima or local maxima of a function. It is solved through Equation (3).

$$\lambda\ (x, u) = \phi(x)\ < u, f(x) > \tag{3}$$

Nonlinear case: In the case of nonlinear separable data, the finding solution is to produce a soft margin that is particularly adopted to noised data. It is different from the separable case as the nonlinear separable case is quite complicated to solve. Fig. 1 shows the figure of linear and nonlinear case. Among the options to help SVM solve nonlinear cases is through the kernel function.

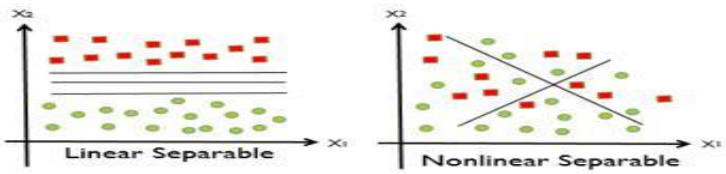

Fig. 1. Linear separable and nonlinear separable data

SVM introduces the kernel function K_{xn},X_i which can help to separate the data easily. It is used to transform input data, X from an input space into a high dimensional space (feature space) through nonlinear mapping in Equation (4).

$$(\phi : X \rightarrow F) \tag{4}$$

In the high dimensional space, the dot product operation can be substituted by kernel function. Three kernels function are employed in this study which are radial basis function (RBF), polynomial and sigmoid. These three kernels are chosen as they are able to deal with nonlinear data, as the data used in this study is a nonlinear separable data.

Radial Basis Function Kernel: RBF kernel or also known as radial kernel is among the most commonly used in classification. In this kernel, two parameters that play a significant role are cost and gamma. The cost parameter function is to exclude the misclassification in training data, while gamma parameter controls the distance training data can reach. This kernel can be defined through Equation (5).

$$\text{radial} : K(x,x_i) = C^{-\Upsilon \| x_i - x_j \|2} \tag{5}$$

Polynomial Kernel: Polynomial kernel is also popularly used for classification. In this kernel, there are two extra parameters other than cost and which are degree and coefficient. It is represented by Equation (6).

$$\text{polynomial} : K(x,x_i) = (<x.x_i> + 1)^d \tag{6}$$

Sigmoid Kernel: Sigmoid kernel is a kernel that must be carefully tuned with a certain value of parameters, it can be non-positive definite. Wrongly defining the range of parameters for the sigmoid kernel will definitely bring wrong results. In this kernel, an extra parameter is coefficient r that controls the shifting parameter which then controls the threshold for mapping. Sigmoid kernel is represented through Equation (7).

$$\text{sigmoid} : K(x,xi) = \tanh (<x:xi> + r) \tag{7}$$

Table 1 shows the summary of kernels with their parameters respectively.

Table 1: Kernels in Support Vector Machine

Kernels	Parameters
Radial	C and ϒ
Polynomial	C, ϒ , r and d
Sigmoid	C, ϒ and r

C:cost,ϒ: gamma, r : coefficient, d:degree

3 Data and Methods

This section explains about the data used and the procedure of the method used. The dataset used in this study is from an Eye Clinic of the Sakarya University Educational and Research Hospital, located in the city of Adapazari, the capital of the Turkish province of Sakarya.

3.1. Data from the Electronic Health Record

The dataset consists of 385 diabetic patients, which can be divided into three categories: 79 patients were not suffering from DR (NODR), 161 patients presented NPDR and 145 patients presented PDR. There are two types of attributes: numerical (Hemoglobin (HGB), Glycated Hemoglobin (HbA1C), High-Density Lipoprotein (HDL), Triglyceride, Low Density Lipoprotein (LDL), Diabetes Duration, Glucose, Creatine and URE) and categorical (Class NODR, Class NPDR, Class PDR). These attributes represent the reading of health records for diabetic patients. Details of the dataset are shown in Table 2.

Table 2: Description of features in dataset

Features	Description
Glycated Hemoglobin	Shows average level of sugar over the past 2 to 3 months
Hemoglobin	Carries oxygen in blood to the cells of the body
High-Density Lipoprotein	Carries LDL cholesterol away from the arteries to liver
Low-Density Lipoprotein	A bad cholesterol that contributes to fatty build ups in arteries
Diabetes Duration	Length of time they have suffered with diabetes
Triglyceride	A type of fat found in blood that the body uses for energy
Glucose	Indicates concentration of blood sugar at a single point in time
URE	Indicates blood urea concentration
Creatine	Facilitates recycling of energy

3.2. Support vector Machine with Hyperparameter Optimization

3.2.1. Hyperparameter Optimization

In every machine learning model, there are a number of hyperparameters that needs to be focused on prior to execution. Hyperparameters can be defined as the external parameters of a model, that are estimated without considering actual observed data. It is different from model parameter which is internal of a model [13].

The word tuning is also frequently used to refer to the parameters that have to be carefully paid attention to, i.e. optimized the parameters with respect to performance. The users of any algorithms can set the parameters value to any specific values or conduct a tuning strategy in order to choose the best parameters for the specific data used.

Therefore, hyperparameter optimization can be defined as the amount of performance gained by setting the considered hyperparameter to the best possible value instead of the default value [14]. It can be understood as a black box search of an x, for example for a given function f:A Bd - > R, the value f(x) is small, where the function can be stochastic. The parameters that affect the learning algorithm can be adjusted based on the model presented. The goal is to optimize the performance metric of the algorithm and to ensure that the model does not bloat its data by tuning.

In this study, the aim is to optimize the hyperparameters in SVM, to ensure that the SVM runs with the optimal parameter values. This study has three classes of data. The problem is formulated based on one-against-one method in order to ensure that SVM is able to handle this multiclass case. Based on this method, SVM is built for each one pair of classes to differentiate the samples of one class from other classes. This method works by construct k (k -1)/2 classifiers where each one is trained using the data from two classes. The value of k the number denoted the number of classes.

3.2.2. Grid Search Strategy

The hyperparameter optimization in this study is conducted using the grid search strategy. In the first step, the range of parameters for each parameter are defined. Grid search strategy is one of the parameters searching techniques that is used to tune the parameter. The reason grid search is chosen as the search strategy is because the grid search is easy to implement, and it can work in parallelization. The grid search basic principle applied in this study work by running the SVM algorithm iteratively with different combinations of parameters within the specific range.

In order to get the optimal solution of parameters, the search range is quite large, and it would take time to run the algorithm. The SVM model is trained and

assessed using all the combinations of the values of parameters, that is called grid. Even though it is simple to carry out, the grid search method might be costly as expense grows exponentially with the number of hyperparameters and the discretize values for each parameter. Too large of a search region can also waste the computational resource.

The searching in grid search will cover the whole search space. Method for selecting the best hyperparameter used in this study is Mean Squared Error (MSE), a statistical method proposed by [15] which consider the variance and bias in parameter estimates and model predictions. It can be calculated based on the formula in Equation (8).

$$MSE: \sqrt[1/n]{\sum_{i=1}^{n} e_i^2} \tag{8}$$

The process of searching is terminated when it reach the maximum predefined range. The grid point with the lowest test error is then chosen. Fig. 2 shows an illustration of grid search strategy. In the illustration, the parameters are set to the range within (0,0.5,0.1) which means the searching interval is between 0 and 0.5 with increment of 0.1 for each grid step in searching for optimal parameters. Fig. 3 shows the proposed SVM with hyperparameter optimization.

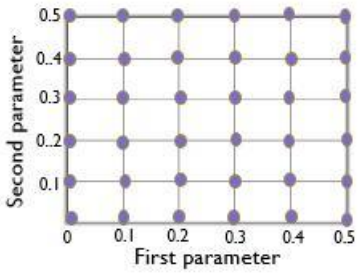

Fig.2. Grid search strategy

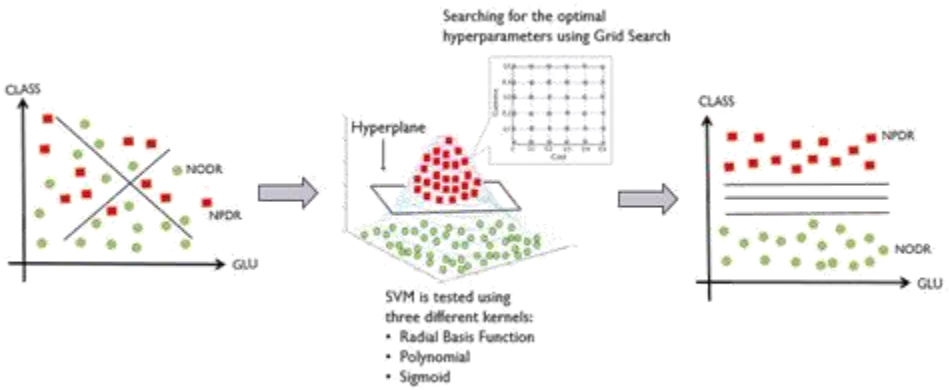

Fig.3. The proposed SVM with hyperparameter optimization

3.3. Performance Evaluation

Performance evaluation is beneficial for comparing the quality of classification across systems [16]. The accuracy of the classifier is indicated by the percentage of the test dataset that are correctly classified by the classifier. It is calculated using the value of true positive (TP), true negative (TN), false positive (FP) and false negative (FN). Besides, confusion matrix is also used to measure the general performance of classifier using confusion matrix. It determines the ability of the classifier to produce accurate diagnosis for DR [17]. Confusion matrix is a table that consists of performance of classification model which are known true values. It contains information regarding actual and predicted classification done by a classification system [18]. Besides, sensitivity and specificity also can be measured from the confusion matrix in order to get more specific information on performance of classifier. Sensitivity measures relevant instances selected while specificity measures the exactness of said classifier.

The words sensitivity and specificity have their origins in screening tests for diseases. Sensitivity is defined as the probability of test that says a person has the disease when in fact they really do have the disease. In other words, it measures how likely it is for a classifier to pick the presence of a disease in a person who has it. [19] suggests that the index, sensitivity is the first priority to be considered. This is because in the reality of medical-care case, if a patient is found to be true positive but he/she is not cured further, he/she will suffer an irreparable damage or in DR permanent blindness.

While specificity defines the probability that the classifiers says a person does not have the disease when in fact they are disease free. It is also an important measure to be considered. An ideal classification should have high sensitivity and high specificity value. Thus, in this paper we are looking for high sensitivity and specificity classifier, with emphasizing on sensitivity value as one of the motivations for this research is to minimize cases of visual loss.

Besides, in order to determine which classifier has the best performance, it is good to evaluate the classifiers with additional evaluation metrics. F-measure is a harmonic mean of precision (positive predictive value) and recall (exactness of classifier) [16]. According to Van Rijsbergen (1979), F-measure is defined as a combination of recall (R) and precision (P) with an equal weight in the following formula in Equation (10). According to [20], precision can be defined as the probability that a randomly chosen predicted instance (positive) will be relevant while Recall is how close we are to a specific target on average.

4 Results, Analysis and Discussions

The computational experiment in this study were done on an Intel Core i3-3110M CPU 2.4 GHz, on a 64-bit windows 8 operating system. The software used was the R software version 2016.

To build the SVM with hyperparameter optimization model, the formulated problem was tested on a benchmark dataset. This dataset corresponds to the DR classification problem of three classes. The overall ow of the process was shown in Figure 3. In the beginning, the dataset was partitioned into training and testing data; the proportion of elements in each part was 70% and 30% respectively. A 10 fold cross validation was then applied to the training set.

To handle multiclassification in this study, the one against one method (pairwise coupling) was applied. In the first iteration of training, data from NODR and NPDR were considered. The target is to distinguish the data of NODR from the data in NPDR. During the training phase, the kernel will be applied to the SVM to help the SVM separate the nonlinear data of NODR and NPDR. For the first training, radial kernel was used. The radial kernel has its own parameters. In order to obtain the best result, the parameters must be optimized. The grid search strategy was used during the parameter tuning. Table 3 shows a summary of parameters setting. The range of the parameters are defined with considering the range defined in [21]. The performance is obtained from the k th validation data, and it is evaluated based on the dispersion error.

Table 3: Parameter setting for SVM

Parameters	Range for tuning
Cost (C)	2(2:10)
Gamma (ϒ)	(0,0.1,0.01)
Coefficient (r)	2(10;2)
Degree (d)	(0,3,1)

The output from the training phase is the best SVM-radial model found in the search process. The best model is the model with the lowest value of dispersion error. From the best model, the information on the best hyperparameters were obtained. The best model is then evaluated on the testing data. In the end, the algorithm will produce the result based on the performance measure stated in Section 3. The same process repeated using the data from the other two classes (NPDR and PDR). This same experiment also will be repeated using polynomial and sigmoid kernel. Their performance will be compared to determine the best SVM model.

Table 4 presents the summary of the classification error for each kernel and its optimal value. From the table, it is found that the polynomial kernel has the lowest error classification compared to the other two kernels. Meanwhile, the kernel with the highest classification error is the sigmoid kernel. Besides, as the cost and are the most influential parameters in SVM, the performance of gamma for each kernel against the error with respective to the constant cost value of 4 is plotted in order to see a pattern inside the hyperparameter optimization. From Fig. 4, it can be seen that the performance for each of the kernels are not consistent. For example, for radial kernel, the error reading is fluctuated. It is keep increasing and decreasing alternately. From the graph, it can be concluded that the pattern of performance is not final thus all values within the specific range must be tested. Thus, the grid search strategy is the right choice for hyperparameter optimization as it tested all the value within the range and not random tested.

Table 4: Summary of classification error for each kernel

Kernel	Optimal Hyperparameter				Error
	C		r	d	
Radial	64	0.03	n/a	n/a	0.3765
Polynomial	64	0.06	0.0325	3	0.3739
Sigmoid	128	0.01	0.0313	n/a	0.3978

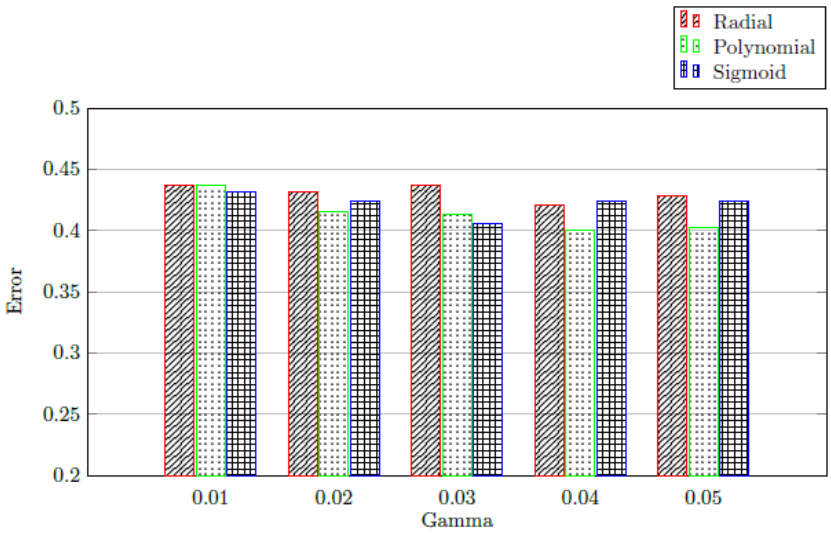

Fig.4. Comparison of hyperparameters performances between radial, polynomial and sigmoid kernel

From the result, it can be seen that the polynomial and radial kernel produce very similar performances, where they are only differed by 0.0026 of classification error. Based on their accuracy, they differ only by 0.52% which is less than 1%. As the performance of polynomial kernel and radial are very close to each other, the other performance measure that has to be considered is the sensitivity of the models. Previously, [19] suggested that the index, sensitivity has to be a priority in making a decision because in the reality of medical care cases, if a patient is true positive but he/she has not cured further, he/she will suffer an irreparable damage or in DR, it is permanent blindness. Fig. 5 shows the sensitivity analysis for the three kernels.

Currently, the focus is on the sensitivity of the polynomial and radial kernel. Based on the graph, the sensitivity between the two kernel models is also close to each other. However, it is obvious that the radial kernel gained highest performance in two classes out of three. Only in the class of NPDR, the polynomial outperformed the sensitivity of radial with higher reading of 6.77%. Thus, in this case, it can be concluded that, both polynomial and radial basis function are suitable to be used for this kind of data. However, radial basis function is decided as the best model based on the higher sensitivity performance in two classes of DR compared to polynomial kernel.

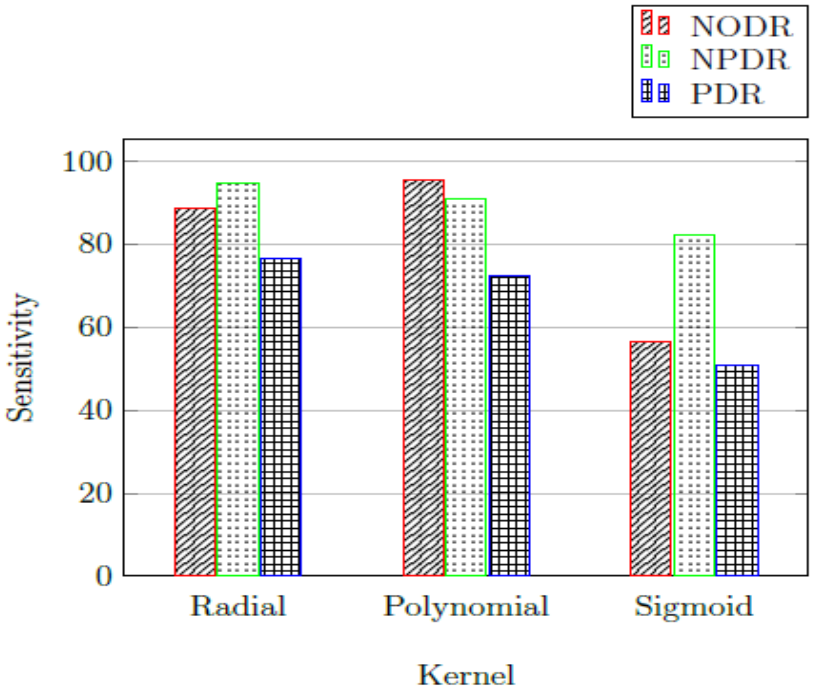

Fig.5. Comparison of sensitivity between radial, polynomial and sigmoid kernel

The details performance of the best SVM model is shown in Table 5. The rest performance measure observed is accuracy of the algorithm. The accuracy of

the SVM optimized is indicated by the percentage of the test dataset that are correctly classified. The accuracy obtained for classification is high, which is 85.45%.

Besides, the performance of optimized SVM also have been precisely analyzed using the other additional metrics. The value of sensitivity and specificity for each class in optimized SVM is high. High sensitivity values bring a meaning that the classifier has a low incorrect negative prediction. High specificity means that the testing data for each class have classified with high precision.

Table 5: Result of the best SVM model based on each classes of DR

Algorithm	Accuracy	Class	Sensitivity	Specificity	Precision	Recall	F-Measure
SVM	85.45	NODR	0.9494	0.9804	0.9260	0.9494	0.9375
		NPDR	0.8882	0.8438	0.8033	0.8882	0.8437
		PDR	0.7655	0.9376	0.8810	0.7655	0.8192

F-measure is then calculated. In the F-measure, the objective of the experiment is to find the algorithm with the highest value, or 1. It shows high F-measure, 0.8668 (in average) with a good precision and recall performance.

4.1. Comparisons of algorithms

Comparison results were established between SVM with radial kernel and SVM with polynomial and sigmoid kernel. They were compared based on the performance metrics other than sensitivity, as it was discussed in previous section. Table 6 shows the performance result for each kernel based on each classes of DR.

For the class of NODR, the performance of SVM with radial kernel was high which was more than 0.9260% of each performance metrics. It gained the highest performance for precision, recall and F-measure. The performance of its sensitivity was lower than SVM with polynomial kernel but higher than SVM with sigmoid kernel. Lower sensitivity means that SVM with radial kernel has lower capability than SVM with polynomial kernel to correctly identify patients for NODR compared to polynomial kernel, as the sensitivity measures refer to the ability of an algorithm to correctly identify a person without the disease. Compared to sigmoid kernel, SVM showed a better performance especially in terms of precision. The precision of radial shows the largest difference where radial has 16.43% higher performance that sigmoid.

In the class of NPDR, SVM also showed a better performance compared to polynomial and sigmoid in terms of specificity and precision, while for the other two metrics recall and F-measure, it has lower performance compared to polynomial. Having a higher specificity value and higher precision value compared to the other model means that, SVM with radial kernel has high ability to recognize people without NPDR. At the same time, it also shows high

relevance in the result as the precision measured on the fraction of the relevant result from all the result retrieved.

In the classes of PDR, the performance of SVM with radial kernel also gained the highest performance in specificity and precision when compared to the other kernel. However, it has lower recall and F-measure when compared to polynomial but higher than sigmoid. With a higher performance of precision but lower recall, it means that SVM-radial has a higher ability to return only relevant results of the DR classification compared to other kernels. However, it has a difficulty in identifying the true relevant samples from the results.

Table 6: Comparison of DR classification for radial, polynomial and sigmoid kernel

Class	Techniques	Accuracy	Specificity	Precision	Recall	F-Measure
NODR	Radial	**85.45**	0.9494	**0.9804**	**0.9260**	**0.9494**
	Polynomial	**85.97**	**0.9869**	0.9474	0.9114	0.9492
	Sigmoid	59.74	0.9412	0.7831	0.8228	0.8245
NPDR	Radial	**85.45**	**0.8882**	**0.8438**	0.8033	0.8437
	Polynomial	**85.97**	0.7946	0.7700	**0.9565**	**0.8532**
	Sigmoid	59.74	0.6339	0.5260	0.5652	0.5449
PDR	Radial	**85.45**	0.9376	**0.8810**	0.7655	0.8192
	Polynomial	**85.97**	**0.9833**	0.7241	**0.9054**	**0.8268**
	Sigmoid	59.74	0.7708	0.5736	0.5103	0.5401

The optimized SVM is also compared to SVM that is non-optimized. The parameters of non-optimized SVM was run using a default parameter value at C=1, G=0.11 and was tested using radial kernel. They were compared based on the performance metrics. From the Table 7, optimized SVM obtained a better result in all the metrics for the three classes of DR. Based on the accuracy, optimized SVM shows an improvement with 8.83% higher than non-optimize SVM. It is a good improvement for DR classification as the increase in accuracy means increasing the ability of the algorithm to stage the severity of the disease.

Table 7: Result for DR classification for SVM with optimization and SVM without optimization

Class	Techniques	Accuracy	Sensitivity	Specificity	Precision	Recall	F-Measure
NODR	Optimize	**85.45**	**0.9494**	**0.9804**	**0.9260**	**0.9494**	**0.9375**
	Non-optimize	76.62	0.8481	0.9771	0.9054	0.8481	0.8758
NPDR	Optimize	**85.45**	**0.8882**	**0.8438**	**0.8033**	**0.8882**	**0.8437**
	Non-optimize	76.62	0.7950	0.7545	0.6995	0.7950	0.7442
PDR	Optimize	**85.45**	**0.7655**	**0.9376**	**0.8810**	**0.7655**	**0.8192**
	Non-optimize	76.62	0.6897	0.7812	0.6897	0.7813	0.7327

5 Conclusion

In previous years, the method of finding the best values for hyperparameters has been a manual effort. The researchers commonly set the values based on past experiences using the machine learning algorithms to train models. However, it is worth to mention that the best settings of hyperparameter might be changed with different nature of data. Thus, it is hard to prescribe the hyperparameter value based on previous experience. Therefore, an alternative configuration such as an automated and guided searching is needed, as in this case hyperparameter optimization was employed.

The predetermined of the range value (refer Table 2) with the grid search strategy method were an important factor to find the optimal pair value hyperparameters. From the graph in Figure 4, it can be seen that the pattern of SVM performance is not consistent along the increasing of the hyperparameter value. For example, with the increasing value in gamma and cost, it does not necessarily increase the value of the SVM's accuracy. Thus, it is important to try all the combinations of parameters in order to get an optimal hyperparameter for SVM.

Grid search is an excellent strategy used to find the best hyperparameters as it tests for each hyperparameters within defined ranges instead of testing the combination of parameters randomly. However, it has to be noted that a very large search region only wasted the computational resource. In contrast, if the range is set for a very small search region, it might not return the best outcome. Thus, the provided value must be in a suitable range.

Besides, the number of hyperparameters in the kernel will definitely affect the performance of SVM, such as their computational time. For example, the computational time for the radial kernel is smaller compared to polynomial and sigmoid kernel as the radial kernel is only deal with two parameters, while polynomial and sigmoid deal with more than two parameters. Therefore, polynomial and kernel will have more combination of hyperparameters to test and need an extra time to execute all the experiments.

In this study, the performance of radial kernel and polynomial is very close to each other. However, SVM with radial kernel was chosen as the best model. The first factor was because of sensitivity in radial outperformed polynomial kernel in two out of three classes. As the sensitivity measure is more significant compared to the other additional metrics, SVM radial was chosen as the best kernel. Based on the literature, the radial kernel was used often for classification compared to polynomial. One of the reasons was because the computational time for polynomial was too long as it has to deal with more parameters compared to radial. Thus, it can be concluded that radial kernel is more efficient to be used for higher sensitivity and small computational time in experiment.

This proposed SVM with hyperparameter optimization contributes to a better result for DR classification. The implementation of the proposed DR classification with an excellent performance able to serve as an aid in assisting experts in the diagnosis of DR. Besides, it is highly important to classify and categorize the stage of severity of DR in order to provide an adequate therapy. With good management, the cases of visual loss can be prevented. With the healthcare industry continually looking to improve the efficiency and throughput, this study seems to be a satisfactory solution that can provide quick result and timely manage eye screening. Further studies will be conducted to improve the performance of these classification techniques by using larger datasets. A more advanced technique such as hybrid supervised machine learning can also be incorporated.

ACKNOWLEDGEMENTS

The authors wish to thank Universiti Sains Malaysia for the support it has extended in the completion of the present research through the support of USM Fellowship 2017/2018.

References

[1] Global reports on diabetes, Tech. rep., World Health Organization (2016).

[2] S. Piri, D. Delen, T. Liu, H. M. Zolbanin (2017). A data analytics approach to building a clinical decision support system for diabetic retinopathy: Developing and deploying a model ensemble, *Decision Support Systems,* 101, 12-27.

[3] W. M. D. W. Zaki, M. A. Zulkiey, A. Hussain, W. H. W. Halim, N. B. A. Mustafa, L. S. Ting (2016). Diabetic retinopathy assessment: Towards an automated system, *Biomedical Signal Processing and Control*, (24),72-82.

[4] M. Abdalla, A. Hunter, B. Al-Diri. (2015). Quantifying retinal blood vessels' tortuosity|review, in: *Science and Information Conference (SAI)* 2015, (pp. 687-693). IEEE.

[5] T. Mapayi, J.-R. Tapamo, S. Viriri, A. Adio. (2016). Automatic retinal vessel detection and tortuosity measurement, *Image Analysis & Stereology,* 35 (2),117-135.

[6] B. Wu, W. Zhu, F. Shi, S. Zhu, X. Chen. (2017). Automatic detection of microaneurysms in retinal fundus images, *Computerized Medical Imaging and Graphics*, 55, 106-112.

[7] H. Evirgen, M. CMerkezi,. (2014). Prediction and diagnosis of diabetic retinopathy using data mining technique, *Turkish Online Journal of Science & Technology*,4 (3), 32-37.

[8] V. Balakrishnan, M. R. Shakouri, H. Hoodeh, et al. (2012). Predictions using data mining and case-based reasoning: a case study for retinopathy, *World Academy of Science, Engineering and Technology, International Journal of Medical, Health, Biomedical, Bioengineering and Pharmaceutical Engineering*, 6 (3), 55-58.

[9] M. Skevo_lakas, K. Zarkogianni, B. G. Karamanos, K. S. Nikita. (2010). A hybrid decision support system for the risk assessment of retinopathy development as a long term complication of type 1 diabetes mellitus. *Engineering in medicine and biology society (EMBC), 2010 annual international conference of the IEEE*, pp 6713-6716.

[10] K. Ananthapadmanabhan, G. Parthiban, Prediction of chances. (2014). Diabetic retinopathy using data mining classification techniques, *Indian Journal of Science and Technology*, 7 (10), 1498-1503.

[11] W. Yu, T. Liu, R. Valdez, M. Gwinn, M. J. Khoury. (2010). Application of support vector machine modeling for prediction of common diseases: the case of diabetes and pre-diabetes, *BMC Medical Informatics and Decision Making*, 10 (1) 16.

[12] H.-Y. Huang, C.-J. Lin (2016). Linear and kernel classification: When to use which?, in: *Proceedings of the 2016 SIAM International Conference on Data Mining, SIAM*, pp. 216-224.

[13] C. Riggelsen. (2008). Approximation methods for efficient learning of Bayesian networks, *IOS Press*, 168.

[14] P. Probst, B. Bischl, A.-L. Boulesteix. (2018) Tunability: Importance of hyperparameters of machine learning algorithms, *arXiv preprint arXiv:1802.09596*.

[15] S. Wu, K. McAuley, T. Harris, Selection of simplified models: Ii. (2011). Development of a model selection criterion based on mean squared error. *The Canadian Journal of Chemical Engineering*, 89 (2), 325-336.

[16] Z. C. Lipton, C. Elkan, B. Naryanaswamy. (2014). Optimal thresholding of classifiers to maximize f1 measure, 225-239.

[17] K. V. Varma, T. M. L. Allam Appa Rao, P. N. Rao. (2014). A computational intelligence approach for a better diagnosis of diabetic patients, *Computers and Electrical Engineering*, 40, 1758-1765.

[18] K.Priyadarrshini, Dr.I.Laksahmi. (2018). Prediction analysis of diabetes using bayesian network and naïve bayes technique, *International conference on advancements in computing technologies*, 4(2),71-74.

[19] C.-H. Weng, T. C.-K. Huang, R.-P. Han. (2016). Disease prediction with different types of neural network classifiers, *Telematics and Informatics*, 33 (2) 277-292.

[20] D. M. Powers. (2014). What the f1 measure doesn't measure, Tech. rep., Technical report, *Beijing University of Technology, China & Flinders University, Australia*.

[21] B. Yekkehkhany, A. Safari, S. Homayouni, M. Hasanlou. (2014). A comparison study of different kernel functions for svm-based classification of multi-temporal polarimetry sar data, *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 40 (2) , 281.