Journal of Theoretical and Applied Information Technology
20th December 2015. Vol.82. No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

OPTIMIZING SENSITIVITY AND SPECIFICITY OF ENSEMBLE CLASSIFIERS FOR DIABETIC PATIENTS

MOKHAIRI MAKHTAR*, MOHD KHALID AWANG, M NORDIN A RAHMAN, SYED ABDULLAH FADZLI, MUMTAZIMAH MOHAMAD

Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin,

22000 Tembila, Terengganu, Malaysia

*Corresponding author: mokhairi@unisza.edu.my

ABSTRACT

Predicting the right class for a certain disease in the medical-related field is very critical. The effects of misclassification of the class could be very risky because it may lead to the mistreatment of the patient. The most important classification performance measurements in medical fields are sensitivity, specificity and accuracy. This research aims to focus on the relationship between these three measurements. Misjudgements in classifying a person to a particular disease will prevent him/her from getting the correct treatment. Thus, the accuracy in classifying such medical data should be at the highest. Nevertheless, the most significant measurement is to have the highest sensitivity, because this will show that the classifier correctly classifies the patient who had a positive symptom of a particular disease. By using a single classifier, it is impossible to get the highest sensitivity. Thus, this paper proposed an ensemble method that aimed to increase the sensitivity as well as to improve the accuracy of the classification. The proposed method optimises the three performance measures by giving weights that composed of the proposed objective function. The results showed that the ensemble method is significant to achieve the highest accuracy of 76% with 84% sensitivity and 63% specificity for diabetic dataset from UCI medical data repositories.

Keywords: *Ensemble Classifiers, Classification Model, Medical Dataset*

1. INTRODUCTION

The paper will discuss the solutions for getting a better classification model for diabetic dataset by combining classifiers. Such technique is known as the ensemble method. The objective of this paper is to have a higher sensitivity and specificity by optimizing the ensemble methods. In the medical domain, *Specificity* and *Sensitivity* play an important role in describing the performance of classification models. Studies have been done in the medical field to classify lung nodules in X-ray chest radiographs using a multi-classifier approach that aimed to reduce false positive rate [1]. There were also previous researches that focused on the ensemble method by optimizing other performance measures such as false negative rate and false positive rate for toxicology applications [2]. From the literature, this research is necessary to conduct with the other dataset; with a newly-proposed ensemble method.

This paper proposed the ensemble method with different techniques for the objective function and diversity measure that aimed to achieve the highest *Sensitivity* and *Accuracy*. The new objective function proposed is composite of *Sensitivity*, *Specificity* and *Accuracy*. For the diversity measure, double-fault was used to find diverse classifiers to be combined in the ensemble.

The idea of the ensemble model is to have more expertise (predictive models) involved in decision-making rather than a single model used in predicting the output [3], as shown in Figure 1. The figure showed that an ensemble classifier has to undergo a number of data mining processes [4]. At the data level (Level D), different subsets of the dataset are created in order to make independent classifiers. Each classifier will be used for the next step in Level B. Diversity of an ensemble model can be obtained by using different subsets of feature selection (Level C) and different base classifiers

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

(Level B). Finally, Level A represented the different ways of combining classifier decisions. The final predictive model of the ensemble learning has proven to be a better performance compared to a single predictive model [5]. Thus, this technique will often increase the performance of a predictive model [6],[15],[16]. The number of ensemble methods have been developed, such as bagging and boosting to handle certain applications[4],[5].

Basically, the performance measure used in the classifier is the accuracy itself. By focusing only on the accuracy, the classifier may be biased in a certain class because normally the medical dataset is small and the class of patient who has the disease are limited. Thus, by optimizing other performance measures such as sensitivity and specificity, the accuracy of the classifier will be at the highest and the confidence level of sensitivity is optimum. The performance measures can be calculated based on a confusion matrix generated from classification models. The following is a table of a confusion matrix for the binary classification model.

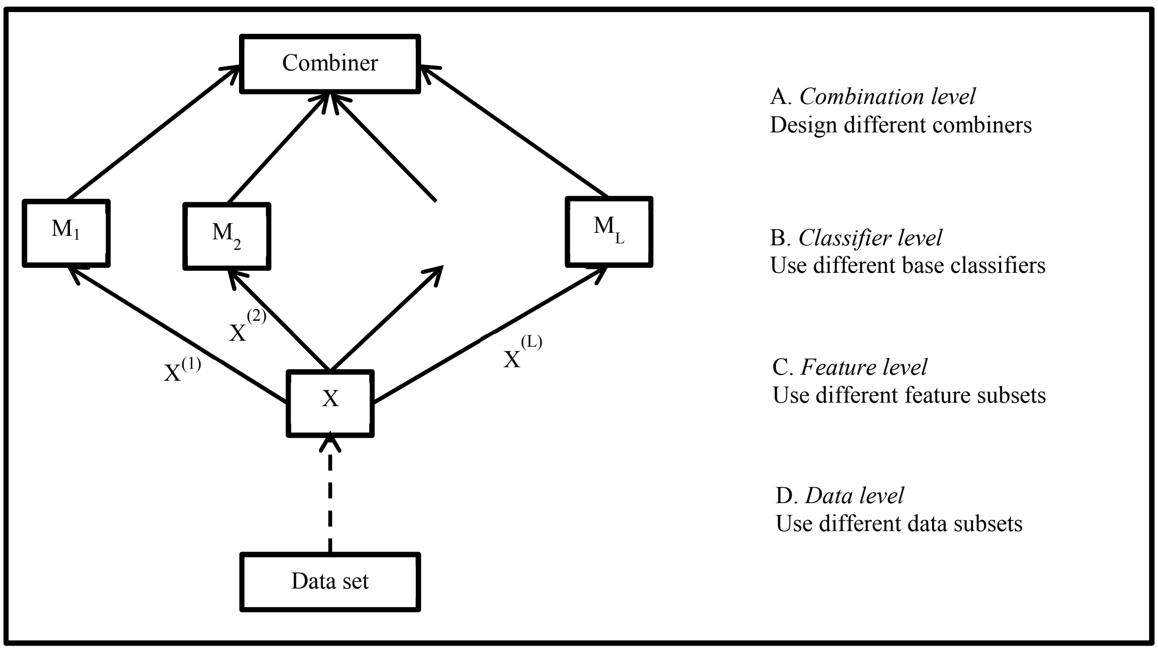

Figure 1: Approaches to Building Classifier Ensembles (Kuncheva, 2005)

Table 1: Confusion Matrix of Binary Classification: True Positive (TP), True Negative (TN), False Negative (FN) and False Positive (FP)

		Actual	
		Positive	Negative
Predicted	Positive	*TP*	*FP*
	Negative	*FN*	*TN*

TP is the number of correct predictions for the positive output (e.g. Yes),

FP is the number of incorrect predictions for the negative output (e.g. No),

FN is the number of incorrect predictions for the positive output, and

TN is the number of correct predictions for the negative output.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

From the confusion matrix, the *Acc, Sensitivity* and *Specificity* can be defined as follows;

$$Acc = \frac{TP+TN}{TP+FP+FN+TN}$$

$$Specificity = \frac{TN}{TN+FP}$$

$$Sensitivity = \frac{TP}{TP+FN}$$

Our methodology addressed the model selection, where comparing the performance of the classifier for each class will lead to useful diverse predictive models for the class of interest from model ensembles. This paper will demonstrate the technique of ensemble and combine classifier decisions using a majority voting with the focus to improve sensitivity and specificity.

The rest of the paper is structured as follows: Section 2 presents related works on applied ensemble methods. Experiments and results are discussed in Section 3. The paper ends with conclusions on the current work and further directions.

2. THE PROPOSED ENSEMBLE METHOD FOR MEDICAL DATASET

In this paper, the predictive model performance measures are focused on Accuracy (*Acc*), Sensitivity and Specificity. *Acc* is the proportion of correct predictions for the positive output (correctly identifying those with the disease), *Sensitivity* is the proportion of incorrect predictions for the positive class and *Specificity* is the proportion of incorrect predictions for the negative class (correctly identifying those without the disease).

In order to build a high quality and robust ensemble method, all three performance measures will be combined as a ranking value that helps in selecting classifiers from a collection of models using an objective function which is a composite of the three performance measures; *Acc*, *Sensitivity* and *Specificity*. The ensemble proposed in this paper followed 3 principles;

a) Objective function to optimise the construction of ensemble models,
b) Diversity measure to search candidate classifiers to be combined that are diverse to each other,
c) Decision fusion strategy to manipulate the class prediction of ensemble models.

The objective function method (*ObjF*), which applies to the ensemble selection, was implemented to optimise the selection of models and the combination method is as follows;

$$ObjF = ((v_1 * Acc) + (v_2 * Sensitivity) + (v_3 * Specificity))$$

where;

ObjF is an objective function to be used as a ranking value,

v_1, v_2 and v_3 are the weights for *Acc*; *Sensitivity* and *Specificity*, respectively,

The values of v_1, v_2 and v_3 are between 0 and 1. The sum of (v_1+v_2+v_3) equals to 1.

The most important principle in constructing an ensemble is the ensemble models must consist of diverse classifiers that had been combined and the variety of classifiers are measured using diversity measures. This paper applied a double-fault measure as a diversity measure. The double-fault measure calculated the diversity between classifiers to find which classifiers are least related to a base classifier [3]. It is the ratio of incorrect predictions by both classifiers as follows;

$$DF_{i,k} = \frac{N^{00}}{N^{11}+N^{10}+N^{01}+N^{00}}$$

The agreement for a relationship between both binary classifiers *i* and *k* is presented in Table 2.

Table 2: Relationship Between a Pair of Classifiers

	D_k correct(1)	D_k wrong(0)
D_i correct(1)	N^{11}	N^{10}
D_i wrong(0)	N^{01}	N^{00}

where;

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

N^{11} is the number of correct predictions made by both classifiers *i* and *k*,

N^{10} is the number of correct predictions made by classifier *i* and incorrect predictions made by classifier *k*,

N^{01} is the number of correct predictions made by classifier *k* and incorrect predictions made by classifier *i*,

N^{00} is the number of incorrect predictions for both classifiers *i* and *k*.

A simple majority voting is applied to the combination of the models in the ensemble as a decision-fusion strategy to build upon the proposed combined performance measure and can be formalised as follows [1];

$$SMV_{oting_i} = \frac{C_i \sum_1^n}{n}$$

where;

i is the index of instances in a classifier.

n is a number or classifiers in an ensemble, and

C is a classifier.

If the value of $SMVoting_i \geq 0.5$, then the predicted class will be *yes* and if $SMVoting_i < 0.5$ the predicted class will be *no*.

The ensemble method proposed can be summarised as Figure 2.

```
Start
  1. Generate Pool of Classification models
  2. Construction of Ensemble Model
       a. Calculate the Diversity measure among the models
          in the pool
       b. Select the most diverse model to be combined
          (ensemble model)
       c. Calculate the objective function (ObjF) of the
          ensemble model
  3. Select final ensemble models that fit the objective
     function.
End
```

Figure 2: The Algorithms of Ensemble Method

The limitation of the method depended on the diversity of the models in a repository. The number of diverse models in the repository may improve the accuracy, sensitivity and specificity. Otherwise, the performance of the ensemble might be the same as the single classifier or maybe just slightly improved. The experiment and results based on the proposed ensemble method will be discussed in the next section.

3. EXPERIMENTS AND RESULTS

A collection of models are generated using Weka, a Java class [13],[14] to make a pool of models. The classifier algorithms used were k-nearest neighbour classifier (weka. classifiers.lazy.IBk), decision trees (weka. classifiers.trees.J48), numerical prediction algorithms (weka. classifiers.rules.JRip) and other ensemble methods [16]. The collections of models are generated based on the Pima Indian Diabetes dataset form the UCI repository. The repository of datasets is maintained by University of California Irvine (UCI) to facilitate the research in data mining and knowledge

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

discovery [12]. The researchers have selected the dataset reported by other researchers and are similar in terms of the number of classes and coming from a medical domain. Table 3 showed the properties and the distribution of the dataset.

Table 3: Summary of the Pima Indian Diabetes Datasets Used in This Paper

No. of Instances	No. of Features	No. of Classes	Class Distribution
768	9	2	500:268

Over 50 predictive models are generated by different combinations of datasets, algorithms, and model parameters. The feature selection algorithm applied to the original full datasets was the Correlation-based Feature Selection (CFS). The researchers used the feature selection to find sets of attributes that are highly correlated with the target classes. Each dataset was processed using Weka with 10-fold cross validation and the numbers of classifiers as mentioned earlier.

From this experiment, it showed that the *Acc* for Pima Indian Diabetes dataset increased and outperformed all the other ensemble methods with the given following values of *CRV* (v_1=0.6, v_2=0.2 and v_3=0.2) using double-fault as the diversity measure. The value *of* the proposed ensemble was 0.76 higher than the other ensemble methods (Bagging = 0.76, AdaBoost (0.75 and Bayes 0.74). Table 4 showed that the ensemble implemented could predict the diabetic class (yes) with the highest *Acc* and maximised *Sensitivity* and *Specificity.*

Figure 3 showed that the ensembles constructed were able to get the highest *Acc* compared to the other ensemble methods such as Bagging, Boosting, and Bayes.

Table 4: Performance measures (Acc, Sensitivity and Specificity) of the Pima Indian Diabetes Data

Ensemble Method	*Acc*	*Sensitivity*	*Specificity*
(Proposed Ensemble)	**0.76**	**0.83**	**0.63**
Bagging	0.74	0.83	0.61
AdaBoost	0.75	0.84	0.60
Bayes	0.74	0.83	0.71

.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

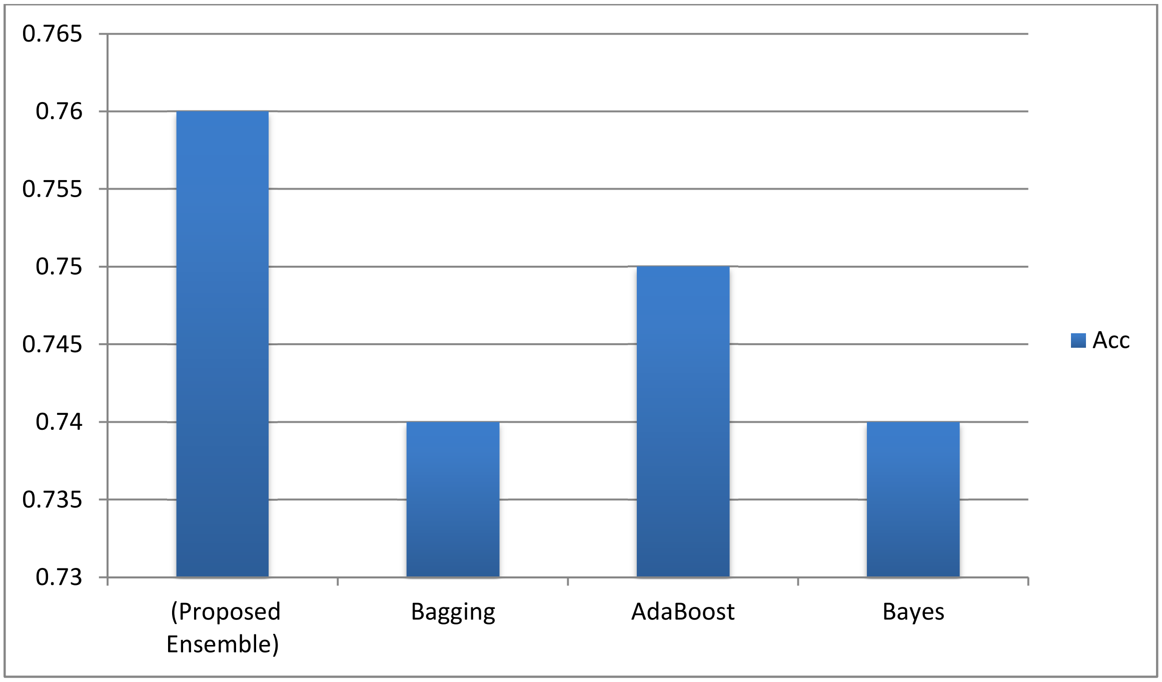

Figure 3: Accuracy of different ensemble methods

4. CONCLUSIONS

This study showed that the combining of the classification models with suitable techniques in calculating the diversity of models and the optimisation value of the objective function is able to achieve the highest *Acc* and maintain a higher *Sensitivity and Specificity*. As a conclusion, the results showed that by combining performance measures (*Acc*, *Sensitivity* and *Specificity*), as proposed within this study, the *Acc* has increased, as well as *Sensitivity* and *Specificity*. The results also showed that the objective function *(ObjF)* proposed by the given values of v_1=0.6, v_2=0.2 and v_3=0.2 and double-fault as the diversity measure may as well be applied to other domains and datasets. The results may be better improved by generating the number of diverse classifiers to be included in the repository. The diverse model to be combined in ensemble could increase a higher *Acc*. Apart from that, other diversity measures can be applied in order to find the most diverse classifier to be combined in the ensemble. To prove this, further study will have to be done to another domains and datasets.

Acknowledgements. This work is supported by UniSZA.

REFERENCES

[1] C. Pereira, L. Alexandre, A. M. Mendonc ̧a, and A. Campilho, A multiclassifier approach for lung nodule classification, *LNCS., 4142* (2006), 612–623.
[2] M. Makhtar, L. Yang, D. Neagu and M. J. Ridley, Optimisation of classifier ensemble for predictive toxicology application, *in* ‘*Proceedings of the 14th International Conference on Modelling and Simulation (UKSim2012)*’, IEEE, (2012), 1–6.
[3] L. Rokach, Taxonomy for characterizing ensemble methods in classification tasks: A review and annotated bibliography, *Comput. Stat. Data Anal., 53* (2009), 4046–4072.
[4] L. I. Kuncheva, Combining Pattern Classifiers: Methods and Algorithms, Wiley-Interscience, (2005).
[5] T. Woloszynski, and M. Kurzynski, , A probabilistic model of classifier competence for dynamic ensemble

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

selection, *Pattern Recogn., 44* (2011), 2656–2668.

[6] T. G. Dietterich, Ensemble methods in machine learning, *in* J. Kittler & F. Roli, eds, 'First International Workshop on Multiple Classifier Systems', *Lecture Notes in Computer Science, Springer - Verlag, New York*, (2000), 1–15.

[7] L. Breiman, Bagging Prediction. *Machine Learning*, 24 (1996), 123-140.

[8] R. E. Schapire, Y. Freund, P. Bartlett, and W. S. Lee, "Boosting the Margin: A New Explanation for the Effectiveness of Voting Methods". *Annals of Statistics*, 6 (1998), 1651-1686.

[9] D. Neagu, M. V. Craciun, S. A. Stroia, and S. Bumbaru, "Hybrid Intelligent Systems for Predictive Toxicology - a Distributed Approach", I*nternational Conference on Intelligent Systems Design and Applications*, (2005), 26-31.

[10] W. Wang, D. Partridge, and J. Etherington, "Hybrid ensembles and coincident-failure diversity*", Proc. IJCNN '01 International Joint Conference on In Neural Networks*, 4 (2001), 2376-2381.

[11] A. Kumaravel, "Comparison of two Multi-Classification Approaches for Detecting Network Attacks", *WASJ*, 27 Issue 11 (2013), 1461-1465.

[12] S. D. Bay, D. Kibler, M. J. Pazzani and P. Smyth, The uci kdd archive of large data sets for data mining research and experimentation, *in 'SIGKDD Explorations*', (2000), pp. 14–18.

[13] I. Witten, E. Frank, and M. Hall, Data Mining: Practical Machine Learning Tools and Techniques, Morgan Kaufmann series in data management systems, Elsevier Science & Technology, (2011).

[14] I. H. Witten, E. Frank, L. Trigg, M. Hall, G. Holmes and S. J. Cunningham, Weka: Practical machine learning tools and techniques with Java implementations, *in* '*In Proceedings of ICONIP/ANZIIS/ANNES'99 Workshop on Emerging Knowledge Engineering and Connectionist-Based Information Systems*', (1999), pp. 192–196.

[15] L. Li, R. Stolkin, L. Jiao, F. Liu and S. Wang. A compressed sensing approach for efficient ensemble learning. *Pattern Recognition*, (2014).

[16] A. Dal Pozzolo, O. Caelen, Y. A. Le Borgne, S. Waterschoot and G. Bontempi, Learned lessons in credit card fraud detection from a practitioner perspective. *Expert Systems with Applications*, *41* Issues 10 (2014), 4915-4928.

[17] N. S. Kamarudin, M. Makhtar, S. A. Fadzli, M. Mohamad, F. S. Mohamad, M. F. A. Kadir, Comparison Of Image Classification Techniques Using Caltech 101 Dataset, *Journal of Theoretical and Applied Information Technology*, 71 Issues 1 (2015), pp. 79-86.