VOL. 11, NO. 22, NOVEMBER 2016 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences

www.arpnjournals.com

GROUP-BASED NEAREST NEIGHBOUR IN CERVICAL CANCER SCREENING

Noor Azah Samsudin, Aida Mustapha, Nureize Arbaiy and Isredza Rahmi A Hamid
Soft Computing and Data Mining Centre, Faculty of Computer Science and Information Technology, Malaysia
University Tun Hussein Onn Malaysia, Parit Raja, Batu Pahat, Johor, Malaysia
E-Mail: rahmi@uthm.edu.my

ABSTRACT

In a cervical cancer screening procedure, a patient's Pap smear slide is presented to determine presence of abnormalities. Conventionally, features of individual cells are measured and analysed in the initial screening step. Based on the analysis results at the cellular level, the Pap smear slide is classified as positive (abnormal) or negative (normal). However, each slide presents data on thousands of cells. Consequently, classifying the slide based on cell-by-cell analysis is very time consuming and prone to 'false negative' problems. In this paper, we propose group-based classification (GBC) approach to classify a slide by measuring the slide data as a whole instead of scrutinizing the cells individually. This means measuring the slide's features once from a group of cells to obtain a diagnosis. We apply two group-based nearest neighbour techniques; voting and pooling schemes to label each slide. The performances of the group-based nearest neighbour techniques are evaluated against existing k-nearest neighbour classifier in terms of accuracy and area under the receiver operating characteristic curve (AUC). The group-based nearest neighbour classifiers show favourable accuracy compared to the existing k-nearest neighbour classifier.

Keywords: cervical cancer, classification, K-Nearest neighbour.

INTRODUCTION

Cervical cancer is known to be a major healthcare issue for women all over the world. Indeed, about 528,000 new cases were reported worldwide in 2012. In the United States (US) alone, 12,340 of new cases were estimated, and 4,030 victims were expected to die from the disease in 2013 (American Cancer Society - Cancer Facts and Figure 2013). To reduce the mortality rate caused by cervical cancer, women are encouraged to undertake a cervical cancer screening procedure known as a Pap smear or Pap test, which was originally introduced by Dr Georgios Papanicolaou [1]. In the Pap test, a physician collects sample cells from a woman's cervix and smears or deposits them onto a glass slide. Subsequently, the Pap smear slide is sent to a pathological laboratory to be examined to classify whether the patient's slide has premalignant and malignant cells.

To classify a Pap smear slide, the cells are quantitatively measured and analysed. As discussed in [2], there are various approaches to classifying a slide. The two main approaches are rare event (RE) and malignancy associated changes (MACs). Classifying the slide based on a cell-by-cell analysis, as in the RE approach, is very time-consuming and prone to 'false negative' problems [3]. The MACs approach is an example of the group-based approach that solves the problems of RE-based analysis [30-33]. Instead of depending on the availability of diagnostic cells and spending time scrutinising thousands of cells individually, the MACs approach uses the summary statistics of the features in intermediate cells. However, using the summary statistics of these cells leads to increasing the number of features, or dimensions, which makes classifying the slide more difficult. As one method of addressing the limitations of using the summary statistics as features, group-based classification (GBC) is considered a potential solution that takes the raw measurements of the cells as a group of features.

This paper applies group-based classification (GBC) approach [5] to automate a medical diagnostic procedure-specifically, cervical cancer screening. Automated cervical cancer screening involves classifying Pap smear slides as either positive (abnormal) or negative (normal). In [5], an existing classifier-namely the nearest neighbour technique-was exploited to enable GBC approach. In this paper, we will apply these GBC techniques to a real-world MACs data set to evaluate their ability to classify Pap smear slides. The group-based nearest neighbour techniques will also be evaluated against the existing conventional MACs approach that is summary statistics-based classifier, which were developed using *k*-Nearest Neighbour (*k*-NN) technique, using two performance measures: classification accuracy and the area under receiver operating characteristic curve (AUC). This work differs from [5], where the group-based nearest neighbour techniques were evaluated on both, synthetic and public domain data sets.

The paper is organised as follows: Section 2 describes the classifiers used in the experiment. Section 3 presents the data set and discusses the experiment conducted using the classifiers on the MACs data. Section 4 presents the results and finally, Section 5 offers some conclusions.

THE CLASSIFIERS

We have categorised the classifiers to be used in our experiments into three major types: summary statistics-based classifier (conventional MACs approach) [10], one-step and two-step group-based classifiers. We have developed the summary statistics-based classifiers using the *k*-NN [11, 12] techniques. The *k*-NN classifier, which is used to estimate the benefit of variants to classify

training samples, is a non-parametric classifier for which assumptions on data distribution are not required. However, it requires us to select a value for parameter *k*. The procedure to select *k* value for our experiments will be presented in Section 3.

One-step GBC refers to the pooling scheme for group-based k-NN -namely GBkNN(P) [5]. The two-step group-based classifier is the voting scheme of group-based k-NN -namely GBkNN(V). The voting and pooling schemes of GBkNN were chosen because of their promising performance shown in the Iris data set, the Pap smear data set and three types of synthetic data sets proposed by [13]. The group-based NN classifiers have been applied to simulated data thus far, whereas the experiments in this paper use a real-world data set.

Table-1 and Table-2 are to be used to describe the GBkNN(V) and GBkNN(P). As shown in Table 1, let X_{TE} denotes a test set of N_{TE} cells, such that $X_{TE} = \{ \mathbf{x}^1, \ldots, \mathbf{x}^{N_{TE}} \}$. Note that, each i^{th} cell, $\mathbf{x}^i$ is represented by multiple features. Let v_l^i denotes votes for each class based on the number of nearest neighbours from class *l* for i^{th} sample. Note that v_l^v denotes total votes for class *l* based on v_l^i and, v_l^p denotes total votes for class *l*. Let c_l^i denotes class label for each i^{th} cell and c_l denotes a class label for X_{TE}. A discussion on how the GBkNN(V) and GBkNN(P) will be used to classify the group of cells in X_{TE} is briefly described here.

We have demonstrated the possibility of extending non-parametric nearest neighbour methods to GBC using accumulated information gained from voting processes namely, GBkNN(V). The formulae applied in the voting schemes and the base of class label decisions are presented in Table- 1 and Table- 2 respectively. In our experiment, GBkNN(V) is presented with the raw measurements for a group of cells that represent a slide. Using the *k*-NN classifier, these cells will be assigned class labels. After all cells in the group have been labelled, the number of cells designated normal and abnormal will be counted, and the majority vote will be used to determine the final class label.

We have demonstrated the possibility of accumulating information on a group of samples based on pooling scheme namely, GBkNN(P). In GBkNN(P), the total votes of the neighbours from every class are accumulated for the group. The group is then labelled in accordance with the largest total votes. The formulae applied in the pooling scheme and the base of class label decisions are presented in Table-1 and Table-2 respectively.

In this study, we have applied the *k*-NN to automate the existing MACs data classification approach that is the *k*-NN classifier is presented with the slide summary statistics of the cells' features, such as mean and/or variance of X_{TE}. A slide is assigned a class label that has majority voting of *k* neighbours. The use of *k*-NN allows us to compare the existing MACs classification approach with the proposed GBkNN(V) and GBkNN(P).

Table-1. Group based nearest neighbour classifiers: group-based K-NN, voting and pooling scheme [5].

Test set, X_{TE}	**class *l* Metric**	**GB *k*-NN**
$\mathbf{x}^1$: : : $\mathbf{x}^{N_{TE}}$	v_l^1 : : : $v_l^{N_{TE}}$	$c_l^1 = \underset{l}{\text{argmax}}\left(v_l^1\right)$: : : $c_l^{N_{TE}} = \underset{l}{\text{argmax}}\left(v_l^{N_{TE}}\right)$
	$v_l^p = \sum_{i=1}^{N_{TE}} v_l^i$	$v_l^v = \sum_{i=1}^{N_{TE}} c_l^i$
	pooling schemes	voting schemes

Table-2. Pooling and Voting schemes for the group-based K-NN[5].

Group-based classifier	**Pooling scheme**	**Voting scheme**
GB *k*-NN	$\mathbf{c}_l = \underset{l}{\text{argmax}}\left(v_l^p\right)$	$\mathbf{c}_l = \underset{l}{\text{argmax}}\left(v_l^v\right)$

These classifiers have the same ultimate aim: to determine a slide's class membership. However, they differ in their approach. The *k*-NN classifier uses summary statistics to classify each slide. The group-based nearest neighbour classifiers use the raw measurements of the cells as a group to classify each slide. The one-step group-based nearest neighbour, GBkNN(P) classifier accumulates information from the group of cells and classify the slide as a whole. The two-step group-based nearest neighbour classifier, GBkNN(V) requires sample cells to be labelled individually prior to labelling the slide. The next section will explain how the experiments are conducted in the study.

EXPERIMENTAL SETUP

Pap smear data

The Pap smear data set used in our research was obtained from the Cytology Department, Queensland Medical Laboratory (QML). The data set consists of MACs cell measurements for a set of Papanicolaou-stained cervical smear slides. We used data of 139 slides for our experiments. According to the QML's diagnosis, 99 slides were classified as normal (negative) and the other 40 slides were classified as abnormal (positive) [6, 7]. For our experiments, we randomly selected data for 1,000 cells from each slide. Each measured cell presented a total of 29 features, referred to as $\{F_1, \ldots, F_{29}\}$. The descriptions of these features can be found in [8].

We applied a normalisation transform to ensure that all of the measurements of the features for the 139,000 cells (1,000 cells × 139 slides) had a zero mean and unit variance [9] such that all measurements were scaled to the range ~N(0,1) [9]. The transformation was class label independent; therefore, the mean and variance of each class may be different.

Feature selection

The ultimate aim of a feature-selection phase is to determine the features that define class memberships in order to eventually improve classification accuracy [21, 22, 24]. There are two main tasks in the feature-selection phase of our experiments: first, determine the optimal number of features needed to represent each slide; second, determine the optimal subset of features to represent the slides.

As noted earlier, the MACs data consist of 139 slides, with each comprising 1,000 cells. Each cell contains a set of 29 features, $\{F_1,\ldots,F_{29}\}$. The brief descriptions of these features are documented in Table 4; they are also the standard features presented in [23]. We determined mean, μ_i and standard deviation, σ_i for every feature, F_i of each slide. As a result, each slide is represented by a set of summary statistics, $\{ \mu_1 \ldots \mu_{29}, \sigma_{1},\ldots, \sigma_{29} \}$ as features. We used μ_i and σ_i for the feature-selection phase because the conventional MACs approach uses summary statistics as features to classify each slide. Note that to estimate the effectiveness of our feature-selection method, we also included an additional random feature, F_{30}. We generated 139 observations using a uniform distribution function in MATLAB for feature F_{30}. The μ_{30} and σ_{30} of F_{30} were also determined.

To determine the optimal number of features for our experiments, we used the feature set of summary statistics, $\{\mu_1 \ldots \mu_{30}, \sigma_{1},\ldots, \sigma_{30} \}$. The approach used to determine the optimal number of features was to apply an inner hold-out within 10-fold cross-validation. Our data consist of 99 slides from the normal class and 40 slides from the abnormal class. The data are divided into a validation set and a training set, and in every fold, approximately 90 per cent of the slides are left for training. In effect, the training set contains about 90 normal slides and 36 abnormal slides for each fold. Therefore, applying the inner hold-out, means that in every fold, 30 per cent of the training set slides were held as an inner test set partition and 70 per cent of the training set slides were assigned to an inner training set.

The slides were then used in the inner training set to select 10 feature subsets from the feature set $\{\mu_1 \ldots \mu_{30}, \sigma_{1},\ldots, \sigma_{30} \}$ using the 'plus-l-take-away-r' algorithm [25]. Each selected feature subset was evaluated using the Mahalanobis criterion function [9]. The 10 training set partitions of the cross-validation resulted in 10 iterations of the inner hold-out approach. Ten feature subsets were therefore obtained, with each containing up to 10 features. Determining the optimal number of features using the Mahalanobis criterion function alone is insufficient because the criterion is known to increase whenever a new feature is added to the subset [26]. Therefore, we used a logistic regression classifier [27, 28] to evaluate combinations of the features from size 1 to 10 for each subset. This means that for every training set partition, the inner training set is used to train the logistic regression classifier, while the inner test set is used to estimate the ROC curve (AUC) of the logistic regression classifier. At the end of the 10-fold cross-validation, 10 sets of the AUC for every feature subset size are obtained. The results of the mean AUC were then plotted against the feature subset size. We then considered the number of features with the maximum mean AUC and identified eight features. The approach used to determine the number of features found eight features to be of optimal size. It should be noted that determining the feature subsets of size eight within the 10-fold cross-validation is likely to result in different feature subsets to be used by different validation sets in the classifier performance-evaluation phase.

The *k*-NN classifier was presented with the selected features originally from the feature set $\{\mu_1 \ldots \mu_{30}, \sigma_{1},\ldots, \sigma_{30}\}$. However, for group-based classifiers, the aim is to classify the slide using the cells' raw measurements instead of the slide summary statistics. Therefore, to present the slides to the group-based classifiers, the selected features are traced to the original feature set $\{F_1,\ldots, F_{30}\}$. For example, if σ_{12}, μ_{15} and σ_{20} are selected for summary statistics-based classifiers, then F_{12}, F_{15} and F_{20} are the features presented to the group-based classifiers. Using the selected features from the summary statistics to trace the features for group-based classifiers may allow us to directly compare the classifiers independently of the features used. Therefore, we believe that evaluating the performance of the group-based classifiers with a potentially (pessimistically) biased feature subset is acceptable. The results of our feature-selection methods are presented in Table-3.

Estimation of k value

When designing a *k*-NN classifier, it is important to select an optimal number of neighbours, which are referred to as *k**. The *k** selection may influence a *k*-NN classifier's performance when evaluated on a validation set partition. In our experiments, for the *k*-NN classifier, the same inner partition of the training set for feature selection is used to determine *k** in the 10-fold cross-validation. In this way, each fold may utilise a different *k** to evaluate the *k*-NN classifier on its corresponding validation set.

In *k** selection, the inner test set is used to evaluate the *k*-NN classifier's performance on a set of odd values in the range of 3 to 25-for example, $k = \{3, 5, 7,\ldots,25\}$. The odd values are considered because the voting neighbours are expected from two classes; therefore, it is important to prevent ties. The maximum value of 25 is chosen because there will be a maximum number of 26 abnormal slides in every training set partition. The AUC for the *k*-NN classifier is evaluated for every *k* value. In each fold, the *k** is selected based on the maximum AUC on the inner test set.

Table-3. Feature subsets selected in every fold for the eight-feature experiments. M refers to the mean of the feature and S refers to the standard deviation of the feature. F refers to the raw measurement of the feature.

Cross-validation	Eight-feature subsets
1	[S_6 S_7 S_8 S_{22} M_{27} S_{28} S_{29}] [F_6 F_7 F_8 F_{22} F_{27} F_{28} F_{29}]
2	[M_1 M_4 S_4 M_5 M_6 M_8 S_{19} S_{21}] [F_1 F_4 F_5 F_6 F_8 F_{19} F_{21}]
3	[S_3 S_8 S_{14} M_{21} M_{22} S_{26} M_{29} S_{29}] [F_3 F_8 F_{14} F_{21} F_{22} F_{26} F_{29}]
4	[S_6 S_{16} S_{19} M_{23} M_{24} S_{25} S_{26} S_{27}] [F_6 F_{16} F_{19} F_{23} F_{24} F_{25} F_{26} F_{27}]
5	[S_8 S_9 S_{14} M_{24} M_{26} S_{28} M_{29} S_{29}] [F_8 F_9 F_{14} F_{24} F_{26} F_{28} F_{29}]
6	[S_3 M_5 M_6 S_{13} M_{18} S_{22} S_{25} M_{28}] [F_3 F_5 F_6 F_{13} F_{18} F_{22} F_{25} F_{28}]
7	[S_2 S_3 S_4 S_6 S_7 S_{16} S_{23} S_{25}] [F_2 F_3 F_4 F_6 F_7 F_{16} F_{23} F_{25}]
8	[S_2 S_9 S_{11} S_{17} S_{25} M_{26} M_{28} S_{28}] [F_2 F_9 F_{11} F_{17} F_{25} F_{26} F_{28}]
9	[S_{12} M_{17} M_{20} M_{22} S_{26} M_{28} M_{29} S_{29}] [F_{12} F_{17} F_{20} F_{22} F_{26} F_{28} F_{29}]
10	[S_3 M_{12} S_{13} S_{14} M_{15} M_{24} M_{25} S_{29}] [F_3 F_{12} F_{13} F_{14} F_{15} F_{24} F_{25} F_{29}]

It should also be noted that the same k^* is used to implement the GBkNN(V) and GBkNN(P) to enable the direct comparison of the group-based classifiers performance with the k-NN classifier independent of the k^* value. Therefore, evaluating the performance of the GBkNN(V) and GBkNN(P) with a potentially (pessimistically) biased k^* value is acceptable.

Performance evaluation

To evaluate the performance of the classifiers, stratified 10-fold cross-validation was conducted. A study by Breiman *et al.* [14] demonstrated that 10-fold cross-validation is an almost unbiased approach for estimating classification performance. Stratified 10-fold cross-validation requires the data set to be divided into 10 subsets of about equal size. Each subset is assigned a representative proportion of slides from every class in the data set. Thus, each subset contains approximately 10 per cent of the slides from each class, in this case consisting of approximately 10 normal slides and four abnormal slides. In every fold, one subset is assigned as a validation set, and another nine subsets are assigned as a training set.

The training set is used to design the classifiers, and the validation set is used to evaluate the performance of the classifiers. The k-NN classifier was developed to demonstrate MAC-based classifier. Therefore, the summary statistics of the cells are applied to represent each slide in the validation set. For the group-based classifiers, the original measurements of the cells are used instead of slide summaries because group-based classifiers are developed to analyse a group of cells to label a slide. However, in our experiments, only 100 cells (selected at random) were used from every slide of the validation set rather than all 1,000 cells in order to reduce the computational complexity.

One of the performance measures used in evaluating a classifier's performance is their accuracy, which is determined using equal misclassification costs. The accuracy is the probability that the slides in the validation set are classified correctly by the classifiers. We will obtain 10 sets of accuracy measures for each classifier from the 10-fold cross-validation and then use the mean accuracy and standard deviation of the accuracy for each classifier for performance evaluation.

In addition to the mean accuracy and its standard deviation, we will also use the area under the ROC curve (AUC), which was initially presented in [15]. The ROC curve gained prominence in medical data analysis because it enabled observations to be made on the consequences of decisions [16-19]. In our experiments, the ROC analysis was used to evaluate the capability of each classifier in labelling abnormal slides from the normal slides. Using the ROC analysis, four possible outcomes are obtained as in Table-4. When a slide is correctly classified as normal, it is labelled True Negative (TN). If a slide is incorrectly classified as abnormal, it is labelled False Positive (FP). Similarly, when a slide is incorrectly classified as normal, it is labelled False Negative (FN), and when a slide is correctly classified as abnormal, it is labelled True Positive (TP).

Table-4. Confusion matrix.

True class label	Predicted class label		
	Negative	Positive	
Negative	TN	FP	Total_Negative
Positive	FN	TP	Total_Positive
	Total Negative Prediction	Total Positive Prediction	N

On each validation set partition, a set of the four outcomes will be obtained at various decision threshold values. The decision thresholds are varied based on the scores assigned to each slide. Table-5 indicates how these scores are assigned to each slide by each classifier. The values of the probabilities of true positives (1-β) and false positives (α) were then plotted as a ROC curve.

Table-5. Scoring system of summary statistics-based classifiers and group-based classifiers.

Classifiers	Scoring system
k-NN	Number of NN from Class 1 against total number of *k* neighbours
GBkNN(V)	Number of cells assigned to Class 1 against total number of cells in the test slide
GBkNN(P)	Total votes of neighbouring cells that belong to Class 1 against total voting *k* neighbours of the test set samples (k × total cells in the group)

We will consider two ways of combining the ROC curves from the 10 different validation partitions in presenting the experiment results: pooling and averaging [20]:

Averaging: We will calculate the ROC curve (AUC) at each successive point of (P(TP), P(FP)) pair for every validation set. The AUC is to be estimated using the formula of trapezoidal numerical integration described in [18]:

$$\mathrm{AUC} = \sum_{\mathrm{i}} ((1-\beta_{\mathrm{i}}.\Delta\alpha) + \frac{1}{2}(\Delta(1-\beta).\Delta\alpha))$$

where

$$\Delta(1-\beta) = \left(\frac{TP_i}{TP_i + FN_i}\right) - \left(\frac{TP_{i-1}}{TP_{i-1} + FN_{i-1}}\right)$$

$$\Delta\alpha = \left(\frac{FP_i}{TN_i + FP_i}\right) - \left(\frac{FP_{i-1}}{TN_{i-1} + FP_{i-1}}\right)$$

There will be 10 sets of AUC to be estimated for the 10 validation set partitions. We will use the mean and standard deviation of the 10 AUC values to evaluate the performance of the classifiers, as recommended by [20]. The mean AUC will give an estimate of the true area and an estimate of its standard error, which is determined from the standard deviation of the 10 AUC values.

Pooling: We will also present an average or 'group' ROC curve for each classifier, as described in [20]. Thus, we will pool the frequencies of TPs and FPs at varying decision thresholds, as defined in Table 6, for every classifier. At the end of the experiments, we will obtain 10 sets of these frequencies. These frequencies will then be averaged to plot a single ROC curve for every classifier.

RESULTS AND DISCUSSIONS

This section presents the results of the experiments. Table-6 shows the mean accuracy (Mean Acc) and mean AUC for the experiments with eight features. It also presents the plot of the average ROC curves for the classifiers in Figure-1. Overall, group-based classifiers show higher mean accuracy and mean AUC than the summary statistics-based classifier. Figure-1 shows that both, GBkNN(V) and GBkNN(P) classifiers perform better than the *k*-NN classifier. GBkNN(P) has the best performance among the *k*-NN classifiers. In fact, GBkNN(P) has the maximum AUC for the experiment using eight features. Thus, the GBkNN(V) and GBkNN(P) classifiers are examples of extended non-parametric classifier applicable in Pap smear slide classification problem.

Table-6. Results of means accuracy (Mean Acc) and mean area under the receiver operating characteristics (ROC) curve (AUC).

Classifiers	Eight features	
	Mean Acc± std dev	**Mean AUC ± std dev**
Summary statistics based classifier: k-NN	0.757±0.077	0.807±0.080
Two-step GBC : GBkNN(V)	0.943±0.100	0.966±0.068
One-step GBC : GBkNN(P)	0.950±0.089	0.978±0.048

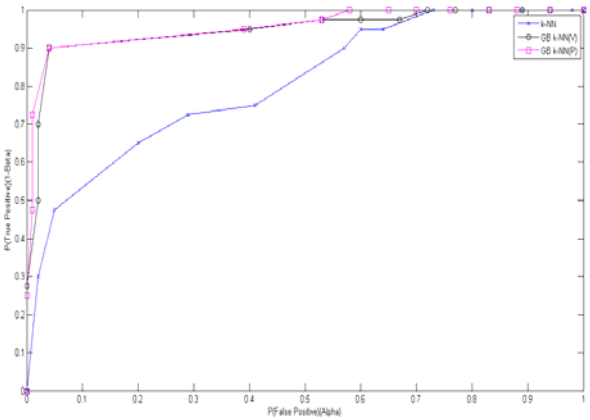

Figure-1. ROC curve of *k*-NN, GB *k*-NN(V) and GB *k*-NN(P) classifiers for eight features.

CONCLUSIONS

The MACs data classification problem clearly demonstrates that the additional prior knowledge that a group of cells belongs to the same slide (patient) can be effectively utilised to reduce misclassifications. The apparent finding from the results is that the one-step GBC is more effective than classifications that utilise a two-step approach. In conclusion, the GBC approach is applicable to detecting the MACs phenomenon using the raw measurements of the cells as a group.

ACKNOWLEDGEMENT

This project is part of the study funded by Ministry of Higher Education under Research Acculturation Grant Scheme (RAGS), grant number R045 and Gates IT Solution Sdn. Bhd.

REFERENCES

[1] G. N. Papanicolaou and H. F. Traut, "The diagnostic value of vaginal smears in carcinoma of the uterus," American Journal of Obstetrics and Gynecology, vol. 42, pp. 193-206, 1941.

[2] B. Nordin and E. Bengtsson, "Specimen analysis by rare event, cell population, and/or contextual evaluation," in Automated Cervical Cancer Screening, H. K. Grohs and O. A. N. Husain, Eds. New York: IGAKU-SHOIN Medical Publishers, 1994, pp. 44-51.

[3] R. M. DeMay, "Common problems in papanicolaou smear interpretaion," Archives of Pathology and Laboratory Medicine, vol. 121, pp. 229-238, 1997.

[4] N. A. Samsudin and A. P. Bradley, "Group-based meta-classification," in 19th International Conference on Pattern Recognition Tampa, Florida: IEEE, 2008.

[5] N. A. Samsudin and A. P. Bradley, "Nearest neighbour group-based classification," Pattern Recognition, vol. 43, pp. 3458-3467, 2010.

[6] A. Mehnert, A. P. Bradley, and P. Bamford, "An independent review and evaluation of the Cytometrics Project's classifiers for Papanicolaou stained cervical slides," Queensland: Centre for Sensor Signal and Information Processing, 2002, pp. 1-61.

[7] A. J. H. Mehnert, "Image analysis for the study of chromatic distribution in cell nuclei with application to cervical cancer screening," in School of Information Technology and Electrical Engineering. vol. Phd: The University of Queensland, Australia, 2003.

[8] A. Doudkine, C. MacAulay, N. Poulin, and B. Palcic, "Nuclear texture measurements in image cytometry," Pathologica, vol. 87, pp. 286-299, 1995.

[9] A. Webb, Statistical Pattern Recognition. London: Arnold, 1999.

[10] W. L. Martinez and A. R. Martinez, Computational Statistics Handbook with MATLAB: Chapman & Hall/CRC, 2002.

[11] B. V. Dasarathy, Nearest neighbor norms: NN pattern classification techniques, ed ed., 1991.

[12] T. M. Cover and P. E. Hart, "Nearest neighbor pattern classification," IEEE Trans., vol. IT-13, pp. 21-27, 1967.

[13] K. Fukunaga, Introduction to statistical pattern recognition, Second ed.: Academic Press, 1990.

[14] L. Breiman, J. Friedman, R. Olshen, and C. Stone, Classification and Regression Trees. Belmont: Wadsworth, 1984.

[15] A. P. Bradley, "The Use of the Area under the ROC Curve in the Evaluation of Machine Learning Algorithms," Pattern Recognition, vol. 30, pp. 1145-1159, 1997.

[16] R. T. Constable, P. Skudlarski, and J. C. Gore, "An ROC approach for evaluating functional brain magnetic resonance imaging and postprocessing protocols," Magnetic Resonance in Medicine, vol. 34, pp. 57-64, 1995.

[17] R. F. Raubertas, L. E. Rodewald, S. G. Humiston, and P. G. Szilagyi, "ROC curves for classification trees," Medical Decision Making, vol. 14, pp. 169-174, 1994.

[18] A. P. Bradley, "Machine learning for medical diagnostics: Techniques for feature extraction, classification, and evaluation," in School of Information Technology and Electrical Engineering vol. PhD: The University of Queensland, Australia, 1996.

[19] M. S. Chesters, "Human visual perception and ROC methodology in medical imaging," Physics in Medicine and Biology, vol. 37, pp. 1433-1476, 1992.

[20] J. A. Swets and R. M. Pickets, Evaluation of Diagnostic Systems: Methods from Signal Detection Theory. New York: Academic Press, 1982.

[21] M. Dash and H. Liu, "Feature selection for classification," Intelligent Data Analysis, vol. 1, pp. 131-156, 1997.

[22] B. D. Ripley, Pattern Recognition and Neural Networks. Cambridge, UK: Cambridge University Press, 1996.

[23] A. Doukine, C. MacAulay, N. Poulin, and B. Palcic, "Nuclear texture measurements in image cytometry," Pathologica, vol. 87, pp. 286-299, 1995.

[24] A. K. Jain, R. P. W. Duin, and J. Mao, "Statistical pattern recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, pp. 4-37, 2000.

[25] J. Kittler, "Feature set search algorithms," in Pattern Recognition and Signal Processing, C. H. Chen, Ed. The Netherlands: Sijthoff and Noordhoff, 1978, pp. 41-60.

[26] A. Jain and D. Zongker, "Feature selection: Evaluation, application, and small sample performance," IEEE Transactions on PAMI, vol. 19, pp. 153-158, 1997.

[27] D. G. Kleinbaum, Logistic regression: A self-learning text. New York: Springer-Verlag, 1994.

[28] J. A. Anderson, "Logistic discrimination," in Classification, Pattern Recognition, and Reduction of Dimensionality. vol. 2, P. R. Krishnaiah and L. N. Kanal, Eds. Amsterdam: North Holland Publishing Company, 1982, pp. 169-191.

[29] C. R. Hicks, Fundamental concepts in the design of experiments: Saunders College Publishing, 1993.

[30] P. C. Bamford, "Segmentation of cell images with application to Cervical Cancer Screening," in School of Information Technology and Electrical Engineering. vol. Phd: The University of Queensland, Australia, 1999.

[31] P. Jackway and P. Bamford, "Modifications to the AcCELL-SAVANT cytometer to enable the classification of Pap smears by the method of malignancy associated changes," Cooperative Research Centre for Sensor Signal and Information Processing (CSSIP), SPRI Building, Technology Park, Mawson Lakes, South Australia2000.

[32] R. A. Kemp, C. MacAulay, D. Garner, and B. Palcic, "Detection of malignancy associated changes in cervical cell nuclei using feed-forward neural networks," Analytical Cellular Pathology, vol. 14, pp. 31-40, 1997.

[33] D. M. Garner, A. Harrison, C. MacAulay, and B. Palcic, "CytoSavantTM and its use in automated screening of cervical smears," in Compendium on the computerized cytology and histology laboratory, Tutorials of cytology G. L. Wied, P. H. Bartels, D. L. Rosenthal, and U. Schenck, Eds. Chicago, Illinois, 1994, pp. 346-352.