Clustering and Classification of a Qualitative Colorimetric Test

Marzia Hoque Tania
Anglia Ruskin IT Research Institute (ARITI)
Anglia Ruskin University
Chelmsford, United Kingdom
marzia.hoque@pgr.anglia.ac.uk

Khin T. Lwin
Anglia Ruskin IT Research Institute (ARITI)
Anglia Ruskin University
Chelmsford, United Kingdom

Antesar M. Shabut
Anglia Ruskin IT Research Institute (ARITI)
Anglia Ruskin University
Chelmsford, United Kingdom

M.A. Hossain
Anglia Ruskin IT Research Institute (ARITI)
Anglia Ruskin University
Chelmsford, United Kingdom

Abstract—**In this paper, we present machine learning based detection methods for a qualitative colorimetric test. Such an automatic system on mobile platform can emancipate the test result from the color perception of individuals and its subjectivity of interpretation, which can help millions of populations to access colorimetric test results for healthcare, allergen detection, forensic analysis, environmental monitoring and agricultural decision on point-of-care platforms. The case of plasmonic enzyme-linked immunosorbent assay (ELISA) based tuberculosis disease is utilized as a model experiment. Both supervised and unsupervised machine learning techniques are employed for the binary classification based on color moments. Using 10-fold cross validation, the ensemble bagged tree and k-nearest neighbors algorithm achieved 96.1% and 97.6% accuracy, respectively. The use of multi-layer perceptron with Bayesian regularization backpropagation provided 99.2% accuracy. Such high accuracy system can be trained off-line and deployed to mobile devices to produce an automatic colourimetric diagnostic decision anytime anywhere.**

Keywords—machine learning; artificial neural network; multi-layer perceptron; Bayesian regularization; diagnostic decision; colorimetric detection; tuberculosis

I. INTRODUCTION

The automatic color based analysis on mobile platform is capable to provide valuable indication of health status, environmental parameters, food safety, crop harvesting and forensic analysis. In this work, we investigated classification techniques for a plasmonic enzyme-linked immunosorbent assay (ELISA) based qualitative test to determine the presence or absence of tuberculosis (TB) antigen-specific antibodies. There are limited number of ELISA plate readers in Google play and App store such as [1–3]. However, these mobile applications are mainly for quantitative tests and need more technical improvements.

The detail bio-chemical based experiment of plasmonic ELISA is described in [4]. In our preliminary study [5], we have provided the colorimetric detection utilizing unsupervised learning for image processing and supervised machine learning techniques for classification. Using 71 samples, we achieved 97.2% accuracy with 5-fold cross validation (CV) via Random Forest (RF).

TABLE I.   DATASET

Class	Positive	Negative
Number of samples	79	175
Total	**254**	

The aim of this work is to improve the detection accuracy than [5] for a larger dataset (Table I). The plasmonic ELISA based tuberculosis wet-chemical experiment was conducted in University Putra Malaysia. This work utilized the same 18 color moments as [5], i.e. mean, mode, standard deviation, skewness, energy and entropy in L, a and b color channels. By exploring number of machine learning techniques in this study, the performance were evaluated for unsupervised (Section II) vs. supervised learning (Section III), parametric (Section III. A) vs. non-parametric method (Section III. B), feedforward vs. backpropagation (Section IV). We found non-parametric supervised learning to show better performance. The performance of multilayer perceptron with backpropagation was enhanced through regularization. The analysis was performed in MATLAB 2018a in Windows 10 system with 16GB RAM and Intel® Core™ i7-7700HQ CPU at 2.80Hz.

II. UNSUPERVISED MACHINE LEARNING

For a qualitative colorimetric test which can provide naked-eye detection, the close points in data space will be in the same class, which implies, the positive and negative samples should be distinguishable without training the system with the corresponding class label. For such a case, the decision boundaries should lie in low density regions. Thus, the

subjectivity of color interpretation for a binary classification can be easily removed by an unsupervised machine learning technique.

As the image segmentation in [5] was achieved by k-means [6] clustering using colors in AB space, in this work we used k-means to determine the presence or absence of TB antigen-specific antibodies. The two-phase iterative algorithm was used to minimize the sum of point-to-centroid distances for 18 features, summed over 2 clusters present in here. At first the squared Euclidean distance was used, later replaced by other proximity measures such as correlation, cosine and Manhattan distance, which played no difference in the result.
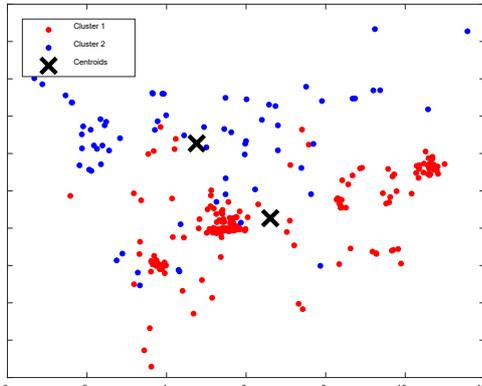
Fig. 1. Cluster assignments and centroids using mean color values in a and b channel

A dataset of n x p matrix can be presented as $X_{ij} \in \mathbb{R}^{nxp}$, cluster centroid location $C_{ij} \in \mathbb{R}^{kxp}$, where, k=2 and row j implies the centroid of cluster j. Using k-means++ algorithm [7], 10 different sets of seeds were selected. In this way the clustering process were repeated 10 times using new initial cluster centroid positions. In return, k-means provided the lowest within-cluster sums of point-to-centroid distances, which is a 2x1 vector, where element j is the sum of point-to-centroid distances within cluster j.

■ Accuracy ■ Sensitivity ■ Specificity

	kmeans	QD	RF	KNN	BRANN
Accuracy (%)	68.1	93.7	96.1	97.6	99.2
Sensitivity (%)	42.0	88.9	91.4	96.2	100
Specificity (%)	80.4	94.8	98.8	96.6	98.86
Rank of training speed	4	1	3	2	5
Rank of prediction speed	4	3	1	2	5
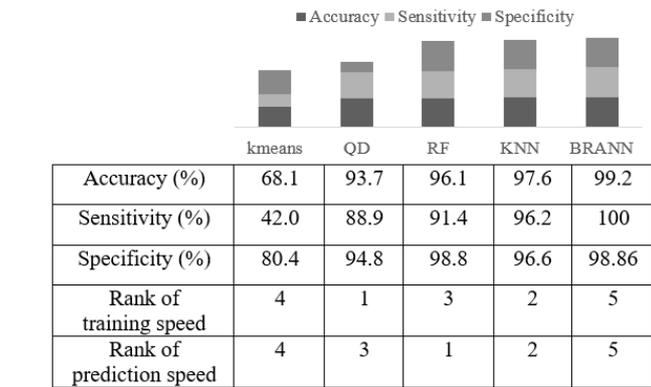
Fig. 2. The computation performace of selected classifiers. QD= Quadratic discriminant, RF= Random forest, KNN= k-nearest neighbors, BRANN= Bayesian regularization aritifical neural network

It is difficult to produce 18-dimensional illustration on paper; thus, the distribution of only mean a and b color value of

the samples are showed in Fig. 1. From the ratio between correctly classified samples and classified samples, the correct rate for k-means clustering was found to be 85.43%. The accuracy was 68.1% (Fig. 2), which is considerably lower than the performance achieved by Random Forest in [5]. The conceivable reason for such poor performance is the use of k-means using dependent variables and the translation invariance.

The color channels l, a and b are independent variables. Therefore, the use of k-means clustering in [5] for image segmentation is justified. We used Euclidean distance in our earlier study satisfying the underlying assumptions. In this work, the presented dataset does not provide easily distinguishable naked-eye detection of TB antigen-specific antibodies. Therefore, the decision boundary does not lie within the mean color value of the low density regions in l, a and b.

As mentioned earlier, the meaningful attributes were extracted from the pixel lists in LAB color space. Thus, variables derived from the histogram in l, a and b are not independent. Moreover, some of the features have more direct dependency on the other features as well. Such dependencies cannot be comprehended by k-means.

III. SUPERVISED MACHINE LEARNING

A. Parametric methods

The parametric methods are independent of the number of samples, easier to perceive and interpret and fast to learn. Among the parametric learners such as linear discriminant [8], quadratic discriminant and logistic regression [8, 9], the quadratic discriminant analysis showed better performance (93.7% accuracy). Although feedforward neural network can be considered as a parametric method, it is discussed later in the neural network section.

B. Non-parametric methods

This work focuses on colorimetric tests of wet-chemicals, where the color of samples can vary due to ambient condition, position of the sample in the assay plate, bio-chemicals and shape of the well. Therefore, for a robust operation, the system should be flexible while making assumptions regarding the form of the mapping function. More generalization can be achieved for the unseen data by such algorithms. Hence, the non-parametric models were explored.

The RF [11] and KNN [12] provided better accuracy than the others e.g. support vector machines, decision trees and ensemble methods such as boosted tree.

	1	2	3	4	5	6	7	8	9	10
Accuracy	96.1	97.6	94.5	95.3	95.3	94.9	94.5	95.7	73.6	75.2

1-Euclidean 2-Manhattan 3-Chebychev 4-Minkowski
5-Mahalanobis 6-Cosine 7-Correlation 8-Spearman
9-Hamming 10-Jaccard
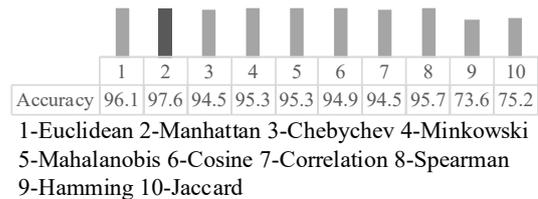
Fig. 3. Accuracy (%) of Fine KNN using different distant metrics

The most common distant metric for KNN is the Euclidean distance. For the fine KNN, which uses medium prediction speed in MATLAB, i.e. ~1 second, the number of neighbor (k) was chosen to be 1. After standardizing the data, such prevalent parameters provided 96.1% accuracy (Fig. 3). The KNN tends to perform better with Euclidean distance when higher number of neighbors are chosen. However, the increment in k degraded the performance. The medium KNN with k=10 and the coarse KNN with k=100 provided 92.1% and 69.7% accuracy, respectively. Thus, other distance metrics were explored.

Among the distance metrics, in addition to Euclidean distance, number of other Minkowski distances were explored in this work e.g. Hamming and Manhattan distance. These two are equivalent when all features are in binary form, which is not the case in here. In this work, both Hamming and Jaccard distance provided poor performance. In case of Jaccard, the number of common attributes is divided by the number of attributes that exists in at least one of the two objects. The performance from Jaccard distance was predictable from the clustering performance.

The Chebyshev, correlation and cosine distance showed similar performance. The cosine (or angular) distance is proportional to the squared Euclidean distance, which dissatisfies the triangular inequality. The KNN cosine works well when the subset of original data is consistent with the original data. From this context, a large dataset, maintaining the same camera to sample distance and ambient condition may work well. The magnitude of the vector is another point of concern for the presented dataset, for which cosine KNN is not a good choice for our case.

The best performance from KNN was achieved using Manhattan (also known as the city block) distance. The 10-fold cross validated accuracy was 97.6%, which is higher than the accuracy achieved in our earlier work [5] with 5-fold cross validation for a smaller dataset using RF. This is another case of Minkowski distance, where p=1. It has L^1 norm, whereas the Euclidean distance is of L^2-norm. The Manhattan distance usually performs better in case of high dimensional data.
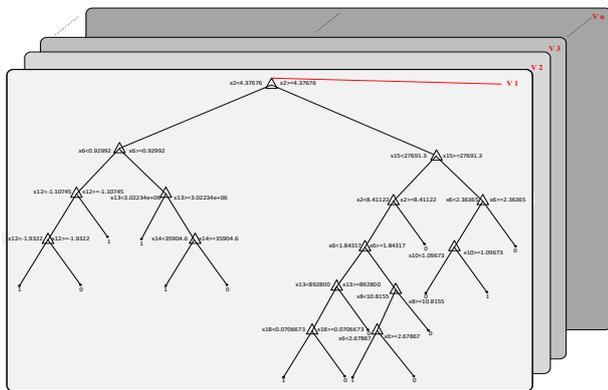
Fig. 4. Random Forest representation. V= Voting

In the presented dataset (Table I) dimension reduction occurred in two stages. At first, the images of the plasmonic ELISA samples were scaled prior to segmentation. In the second stage, the meaningful attributes were extracted from the pixel list of segmented images. Thus, Table I deals with only 18x254 data. However, the number of attributes, i.e. the dimensions may not be comparable in some cases. This can be perceived from the impact of different features. In our preliminary work [5], we were able to discard one of the features (entropy in L) as it had no influence on the classification, which is not the case for the current dataset. This could be the reason behind better performance from Manhattan distance than the standardized Euclidean distance, as the features were not squared.

In this work, the Breiman's RF [11] consisting bagging and random selection of features to create a controlled variance-decision tree collection was implemented using 30 base learners. Few of the decision thresholds can be visualized from Fig. 4, where the score from the classifier can be expressed as the likelihood ratio, i.e. p $(x|class_1)$ / p $(x|class_2)$. The randomization in RF is supposed to reduce the generalization error by reducing the correlation among the decision trees.

The RF achieved 96.1% accuracy using 10-fold cross validation, where the sensitivity and specificity is 91.36% and 98.30% respectively. However, it (Table I) is not a balanced dataset; the ratio is higher than 2:1. The use of balance dataset in [5] provided 100% sensitivity and 88.9% specificity while testing the trained model. The cross-validated result showed 97.1% sensitivity and 97.2 % specificity. As [5] utilized 5-fold cross validation, in order to perform a direct comparison, in this work, we performed 10-fold CV using the dataset of [5]. The overall accuracy was reduced to 94.37% (97.2% accuracy by 5-fold CV), where sensitivity was 94.29% and specificity was 94.44%. RF being a non-parametric method, the classifier was able to learn from the new data introduced in this work (Table I). However, as [5] involved a balanced data, the specificity and sensitivity is balanced for both 5-fold and 10-fold cross validated result. Nevertheless, overfitting is apparent in both cases (in this work as well as in [5]).

The quest of reducing false positive and false negative rate is difficult for an unbalanced data. Before implementing any Bayes optimal decision rule for RF in future, a balanced data should be considered. For a robust system, more focus should be given on generalization so that the trained system can function adequately for the unseen data.

IV. ARTIFICIAL NEURAL NETWORK

In this work, we employed a multilayer perceptron (MLP) with feedforward neural network [13]. The hidden layers used hyperbolic tangent sigmoid transfer function (tanh(N), N= input), whereas the output layer was implemented by linear transfer function. The feedforward network provided 94.9% accuracy.

Using error calculation from the training output and the target, the backpropagation can adjust the weights and biases of the input and hidden layers. To improve the performance furthermore, Bayesian regularization was deployed. A combination of squared errors and weights were minimized to produce a corrected combination, thus a better generalization is produced. Using Bayesian regularization artificial neural network (BRANN), the weights and bias values were updated by Levenberg-Marquardt optimization [14].
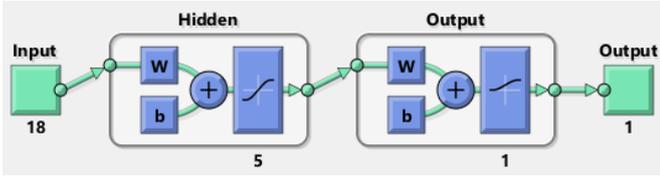
Fig. 5. Network architechture of MLP with Bayesian regularization backpropagation

The network architecture is illustrated in Fig. 5. The hidden layers were varied from 2 to 10. The best result was achieved when the number of hidden layers is either 5 or 6. For a more stable network, 6 hidden layers were chosen.
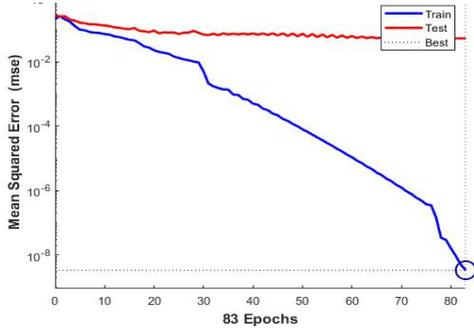
Fig. 6. Training and testing performance of MLP with Bayesian regularization backpropagation

No validation stopping was used, as the regularization was utilized. In this way, the training was carried out until an optimal combination of errors and weights was obtained. The aim of regularization is to build a more generalized model.

Fig. 7. Confusion matrix of MLP with Bayesian regularization backpropagation

The performance of training and testing stage can visualized from the mean squared normalized error with respect to epochs (Fig. 6). The best training performance was $3.3411e^{-09}$ at epoch 83. The error tends to reduce after more epochs of training, however it effects the training time. In an ideal case, the testing graph in Fig. 6 was supposed to follow the training one. Although the error was low for both cases, with more epochs the minimization of testing error stopped earlier, which may have occurred due to overfitting. This issue needs to be further investigated in future.

Reference	CV	Accuracy	Specificity	Sensitivity
	k-fold	(%)	(%)	(%)
[5]	5	97.2	97.2	97.1
This work	10	99.2	98.86	100

Table II shows comparison between the best performance achieved in this work and our earlier work on a smaller dataset [5]. The stable network with cross validation provided 99.2% accuracy (Fig. 7), outperforming [5] and rest of the classifiers of this work (Fig. 2). Based on Clopper-Pearson's 95% confidence interval [15], the accuracy can vary in between 97.18% to 99.90%. The specificity and sensitivity were 98.86% and 100% respectively. The positive predicted value can be calculated from the confusion matrix to be 97.53%, whereas the negative predictive value was 100%. A balance between specificity and sensitivity can be obtained by using equal ratio of positive and negative samples. In the absence of a larger dataset, a resampling technique may also be utilized. However, the MLP implemented with Bayesian regularization backpropagation is slower than all other classifiers mentioned in this work (Fig. 2). Due to 10-fold CV, the computation time was reduced more. To deploy the system on point-of-care (POC) platforms with limited storage and processing capacity, a trade-off would be required to optimize the computation complexity while maintaining the accuracy of BRANN. In such cases, the system can be trained off-line or a server-based approach can be integrated to provide the detection in real time.

V. CONCLUSION

The colorimetric tests on the point-of-care platform using machine learning techniques can bring benefits to millions of population. In this paper, we implemented clustering and classification techniques to automate a qualitative colorimetric decision, i.e. TB antigen-specific antibody detection. Using color moments as the feature-set, the supervised techniques showed better performance than the unsupervised machine learning. The dependency among the variables could not be well-presented by k-means clustering, resulting only 68.1% accuracy. On the other hand, among the supervised techniques, the non-parametric methods performed better than the parametric methods. The higher order discriminant analysis showed 93.7% accuracy, which was the best performance by any paramedic methods.

In our preliminary study, the best performing classifier was RF. In this study, KNN showed better performance than RF. 97.6% accuracy was achieved by fine KNN using Manhattan distance. The highest accuracy of 99.2% was achieved by the neural network (MLP) using Bayesian regularization backpropagation. To the best of our knowledge, this is the best accuracy achieved for the plasmonic ELISA based TB image analysis on any platform including desktop, server or mobile enabled application. Our future work will aim for more generalization of the system maintaining the accuracy achieved in this paper.

REFERENCES

[1]   Alidans srl, "AssayColor," *Android App on Google Play*, 2015. [Online]. Available: https://play.google.com/store/apps/details?id=com.alidans.assaycolor. [Accessed: 10-Jan-2017].

[2]   Enzo Life Sciences inc., "Enzo ELISA Plate Reader," *Android App on Google Play*, 2015. [Online]. Available: https://play.google.com/store/apps/details?id=com.enzo.elisaplatereader. [Accessed: 21-Sep-2017].

[3]   Sicasys Software GmbH, "Spotxel® Reader," *Google Play*, 2017. [Online]. Available: https://play.google.com/store/apps/details?id=com.sicasys.spotxel&hl=en. [Accessed: 12-Jan-2018].

[4]   N. M. Bakhori, N. A. Yusof, J. Abdullah, and H. Wasoh, "Immuno Nanosensor for Ultrasensitive and Affordable Naked Eye Detection of Tuberculosis," *Sensors*, vol. 18, no. 6, pp. 1–10, 2018.

[5]   K. J. Abuhassan *et al.*, "Automatic Diagnosis of Tuberculosis Disease Based on Plasmonic ELISA and Color-based Image Classification," in *2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2017, pp. 4512–4515.

[6]   S. Lloyd, "Least squares quantization in PCM," *IEEE Trans. Inf. Theory*, vol. 28, no. 2, pp. 129–137, Mar. 1982.

[7]   D. Arthur and S. Vassilvitskii, "k-means++: The Advantages of Careful Seeding," in *Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, 2007, pp. 1027–1035.

[8]   R. A. Fisher, "The Use of Multiple Measurements in Taxonomic Problems," *Ann. Eugen.*, vol. 7, pp. 179–188, 1936.

[9]   D. R. Cox, "The Regression Analysis of Binary Sequences," *J. R. Stat. Soc. Ser. B*, vol. 20, no. 2, pp. 215–242, 1958.

[10]   S. H. Walker and D. B. Duncan, "Estimation of the Probability of an Event as a Function of Several Independent Variables," *Biometrika*, vol. 54, no. 1 and 2, pp. 167–179, 1967.

[11]   L. Breiman, "Random Forests," *Mach. Learn.*, vol. 45, pp. 5–32, 2001.

[12]   MATLAB & Simulink, "Classification Using Nearest Neighbors," *MathWorks, United Kingdom*, 2018. [Online]. Available: https://uk.mathworks.com/help/stats/classification-using-nearest-neighbors.html. [Accessed: 06-Apr-2018].

[13]   F. Rosenblatt, *Principles of neurodynamics; perceptrons and the theory of brain mechanisms*. Washington DC: Spartan Books, 1961.

[14]   M. T. Hagan and M. B. Menhaj, "Training Feedforward Networks with the Marquardt Algorithm," *IEEE Trans. Neural Networks*, vol. 5, no. 6, pp. 989–993, 1994.

[15]   C. J. Clopper and E. S. Pearson, "The use of confidence or fiducial limits illustrated in the case of the binomial," *Biometrika*, vol. 26, no. 4, pp. 404–413, Dec. 1934.