Design of efficient classifier integration and performance evaluation in machine learning

K.S.Kavitha , Dr. K.V.Ramakrishnan , Manoj Kumar Singh

Abstract: Characteristics of any classifier heavily depend upon the nature of data set taken for training and verification. Area of applications like health care suffered from having the large and suitable dataset. Classifier designed for health care should show a better generalization and robustness characteristics so that end results presented by classifier can consider with high reliability and confidence. In this paper consistency problem associated with classifier has presented, which is a big issue from practical point of view. Defining committee of experts is one of natural way to increase the reliability in classifier design but at the same time, way of integration rules the end performance. To overcome problem of generalization and consistency of classifier, two methods for developing the mixture of classifier namely TMQD and MVFD are presented. Estimation of quality associated with a classifier is very challenging task for researcher, because there is no single parameter which could alone represents the absolute performance .To measure the quality of classifier rather than having the conventional parameters like sensitivity and specificity, receiver operating characteristics is always a better choice. But in practical environment of health care use of ROC hardly has seen. In this paper detail understanding of ROC and estimation of area under curve has also presented. Selection of threshold value is one of the most important factor to determine the performance of classifier. Dependency of threshold value with population and geographical area making difficult to decide a optimal value. A graphical approach has presented to select the best threshold value as according to environment and need.

Index Terms – Data Mining, Classifier, Classifier integration, ROC, Area under ROC, Sensitivity, Specifity, Heart Diseases, Neural Networks, .

1. INTRODUCTION

There is always good variation in data available if is taken from different places, under different circumstances and by different methods. Quality of classifier completely depends upon the quality of training and test data taken for learning and verification. From health care point of view this is very important that design solution methodology should have more generic characteristics rather than localization. This generic characteristic possible if there is very huge and diverse training and test data set available, but having such data set is a critical limitation of health care environment.
Another issue which affects the performance of classifier is design method adopted in solution. This is universally accepted by research community that no single approach generates satisfactory result in all different environment and situation. One possible design approach to increase the performance and robustness of classifier is to go for

• *K.S.Kavitha is Assistant Professor in the Dept of CSE & ISE, Auden Technology & Management Academy, Bangalore, India.* E-mail: *k_s_kavitha@yahoo.com.*
• *Dr. K.V.Ramakrishnan is with C.M.R .Institute of Technology., Bangalore, India.*
• *Manoj Kumar Singh is Director in Manuro Tech Research, Bangalore, India, E-mail: mksingh@manuroresearch.com*

This generic characteristic possible if there is very huge and diverse training and test data set available, but having such data set is a critical limitation of health care environment.
Another issue which affects the performance of classifier is design method adopted in solution. This is universally accepted by research community that no single approach generates satisfactory result in all different environment and situation. One possible design approach to increase the performance and robustness of classifier is to go for integration of number of classifiers, which have been developed in different circumstances and with different method. The nature of integration will decide the final performance of integrated solution. This is very similar to developing a committee of human expert to find the solution of a problem. Evaluation of performance is one of very challenging task always appears in front of designer, because wrong evaluation can appear as a disaster in practical health care situation. There is still not a unique method available which could estimate the true quality available with solution. Presently applied approaches like sensitivity, specifity or related some variants are very sensitive to population and it's nearly impossible to define the optimal threshold of decision with the used population .Receiver operating characteristics is one of better hope in this requirement. Area under the curve of ROC is one the important characteristics which can use to define the quality of classifier or integrated classifier. Because this is less sensitive and more accurate compare to other used parameters in practice. ROC plot provides a graphic representation of all possible true positive (Se) and false positive (1-Sp) fractions for an ordinal or continuous test.

The term rate is often used by most of researchers in the context of these measures, but is a misnomer because there is no time component. To generate a simple ROC plot, the cutoff value is systematically shifted over the range of observed test values, and Se together with (1 – Sp) are established for each of these, say k, different operating points. The resulting k pairs (x, y coordinates: 1 -Sp, Se) are then graphically displayed as an ROC plot in a unit square that is defined by the x axis (1-Sp) and y axis (Se), each having a length of 1 unit. The use of the false-positive fraction (i.e., 1 -Sp rather than Sp) is logically justified because ROC plots jointly consider the fraction of positive test results for both the diseased and no diseased group, which helps to resolve some mathematical issues associated with their presentation. The connection of the different operating points leads to a staircase trace within the unit square that originates from the upper right corner (where Se = 0 and 1-Sp=0) and ends at the lower left corner (where Se = 1 and 1 – Sp=1). The interesting feature of this plot is that it characterizes the given test by its trace in the unit square, irrespective of the original unit and range of the measurement. ROC plots can be used, therefore, as universal tools for test comparison even when the tests are quite different in their cutoff values and in their units and ranges of measurement. ROC plots for diagnostic tests with perfect discrimination between negative and positive reference samples (i.e., no overlap of test values) pass through the coordinates (0, 1), which is equivalent to a Se and Sp of 1. Consequently, the area under such ROC plots (area under curve [AUC]) would be 1.We assume in our presentation that higher values of the diagnostic marker are indicative of disease, which is by no means a prerequisite for the use of a ROC analysis. If the opposite were true, SE would decrease and SP would increase if the cutoff value were shifted from low to high values. The inverse relationship between the two diagnostic parameters for different cutoff values is a general finding and occurs if the test values for the diseased and no diseased subpopulations have different mean values.

2. RELATED WORK

The integration of multiple classifiers promises higher classification accuracy and robustness than can be obtained with a single classifier. Chibelushi, C.C etc.[1] Proposed an adaptive technique for classifier integration based on a linear combination model. The proposed technique is shown to exhibit robustness to a mismatch between test and training conditions. It often outperforms the most accurate of the fused information sources. E. Kim etc.[2] proposed a combining method, which harness the local confidence of each classifier in the combining process. This method is at the confluence of two main streams of combining multiple classifiers: classifier fusion and classifier selection. Presented method learns the local confidence of each classifier using training data and if an unknown data is given, the learned knowledge is used to evaluate the outputs of individual classifiers. Because of the lack of a clear guideline or technique for selecting classifiers which maximize diversity and accuracy, the development of techniques for analyzing classifier relationships and methods for generating good constituent classifiers remains an important research direction. A framework based on the Bayesian Belief Networks (BBN) approach to classification has presented by Samuel [3]. In the proposed approach the multiple-classifier system is conceived at a meta-level and the relationships between individual classifiers are abstracted using Bayesian structural learning methods. For improving identification rate and real time of ensembles learning algorithm, the diversity of ensemble classifiers is analyzed and a novel combination algorithm with pruning function of multiple classifiers is presented in [4] Min Fang . A coincident errors measure of classifiers is presented for the compound error probability by which classifiers are partitioned, and some classifiers in a partition are pruned. The voting weights of pruned classifiers are assigned according to diversity between classifiers, so that optimize classifier set and voting weights for integration are obtained. Weighting individual classifiers in a multiple classifier system based on their local within-class accuracies is proposed by Shiliang Sun[5]. For an example to be classified distance metric learning is applied to determine the within-class nearest neighbors. Then the local within-class accuracy of an individual classifier for classifying this example is judged by its performance on these neighbors, which is further used to weight the individual classifier. [When a multiple classifier system is employed, one of the most popular methods to accomplish the classifier fusion is the simple majority voting. However, when the performance of the ensemble members is not uniform, the efficiency of this type of voting generally results affected negatively. R. M. Valdovinos [6] presented a functions for dynamic weighting in classifier fusion .Peter Revesz[7] propose *classification integration* as a new method for data integration from different sources. We also propose *reclassification* as a new method of combining existing medical classifications for different classes. We introduce general *classification integration* and *reclassification* methods that create new classes by combining in a flexible way the existing classes without requiring access to the raw data. The flexibility is achieved by representing any linear classification in a constraint database The area under the

receiver operating characteristic curve is the most commonly used measure of the ability of a biomarker to distinguish between two populations. Some markers are subject to substantial measurement error. Under normality assumptions, the authors Enrique F. Schisterman etc.[8]; develop a confidence interval procedure for the area under the receiver operating characteristic curve that adjusts for measurement error. This procedure assumes the availability of data from a reliability study of the biomarker. A simulation study was used to check the validity of the proposed confidence interval. Furthermore, it was shown that not adjusting for measurement error could result in a serious understatement of the effectiveness of the biomarker. Alan Herschtal[9]introduces RankOpt, a linear binary classifier which optimizes the area under the ROC curve (the AUC). Unlike standard binary classifiers, RankOpt adopts the AUC statistic as its objective function, and optimizes it directly using gradient descent. A common approach to training neural network classifiers in a supervised learning setting is to minimize the mean-square error (mse) between the network output for each labeled training sample and some desired output presented by Lee, W.H etc[10] . In the context of landmine detection and discrimination, although the performance of an algorithm is correlated with the mse, it is ultimately evaluated by using receiver operating characteristic (ROC) curves. Atapattu etc.[11] applied the AUC measure for wireless application. they comprehensively analyze the AUC of an energy detector with no-diversity reception and with several popular diversity schemes.

3. CLASSIFIER INTEGRATION

In [12] we have shown that a better classifier design is possible by utilizing the neural network and genetic algorithm. We are using the same classifier in this paper for our further research in designing the mixture of expert and evaluation of its performance. Performance of GA over different parts of data as a training data set and test data has shown in fig(1) ,fig(2),fig(3) and in fig(4).From the graphs this is very clear that variation in perfromance noticible and nature of vriations are different for different data set.This issue generate the pratical utilization of such a classifier having knowldege of leraning from a data set and verfification from another available data set. This is well known that quality of learning affected by the information available in the training data set.If size of data set is small or there is less diversity available within the data,its not possible to perform well in test data set having some different information compare to training data set.At the same time having very large tarining data set with lots of diversity will make learning difficult.If there is a small test data set for varification, uncertainty in result is more and very important issue, reliabilty of its performance in practice will started to appear .To overcome this problem, integration of classifiers which are having knowledge from different environment for same objective can be considered.From health care prospective having large data set is another chellenge in practice.In the presented work ,available data set divided into number of subgroup data sets and each classifier obtained the learning with a subgrop data set and validated with other available subgroup data set.For each traning data set classifier got the independent learning for number of times.All independent learning knowledge obtained from same training data set form a group.In result number of groups are availavle having the knowledge from different data set and verfied with different data set.This create the diversity in knowledge data and in verfication data.At the end the group decision integrated with two different method of mixture.Because all classifier have got the knowledge on same plateform hance rather than having the weighted approach ,equlity impression or voting technique are more preferable.Two different techniques have presented for integration of classifiers namely (1)T MQD:Thresholding the mean of group quantive decision (2) MVFD :Majority voting of group final decision .Both methods are having two stage process to get the final decision.In the TMQD first stage is to get the quantitive result of each group then mean of quantive decision given by each member taken as shown in eq(1), and in second stage the mean of all grops compared with the defined threshold for final decision as shown in eq(2). In the MVFD ,the first stage is to get the mean of quantive decision by each memebr and thresholding applied to get final decision from a group as shown by eq(3) ,eq(4)and eq(5).In second stage a voting scheme based on majority created for final decision as shown by eq(6).

$$f = \frac{1}{mn}\sum_{j=1}^{n}\sum_{i=1}^{m} Cf_{ij} \qquad (1)$$

$$D_{op(TMQD)} = \begin{Bmatrix} 1\ if\ f \geq thr \\ 0\ if\ f < thr \end{Bmatrix} \qquad (2)$$

$$G = \frac{1}{m}\sum_{i=1}^{m} Cf_i \qquad (3)$$

$$H = \sum_{j=1}^{n} x_j \qquad (4)$$

$$where\ \ x_j = \begin{Bmatrix} 1\ if\ \ G_j \geq thr \\ 0\ if\ \ G_j < thr \end{Bmatrix} \qquad (5)$$

$$D_{op(MVFD)} = \begin{Bmatrix} 1\ if\ H \geq E_n \\ 0\ if\ H < E_n \end{Bmatrix} \qquad (6)$$

Where Cf is quantitive output of each classifier in a group. m represents the total number of classifier in a group (here $m = 10$) and n represents the total number of classifiers($here\ n = 4$). Thr is the value of threshold in decision taken here 0.5 and E_n represents the majority voting value,in $E_n = n/2$;

4. DATA SUBGROUPING AND CLASSIFIER INTEGRATION

Heart disease Data set taken from[5].Total number of data available in data set are 270 . Among these data four subgroups have created containg training data set namely D1,D2,D3 and D4,each subgroup having the size of 150 data for training and 120 data for verification.Each subgroup having 60 new data compare to its preceder as it shown in table(1).This method create the balance of diversity as well similarity in environment of learning and verification.

TABLE 1

CREATION OF SUBGROUP DATA SET

Data No.			Tr.Data
1	-	30	D1
31	-	60	
61	-	90	D2
91	-	120	
121	-	150	D3
151	-	180	
181	-	210	D4
211	-	240	
241	-	270	
1	-	30	
31	-	60	

Training data result

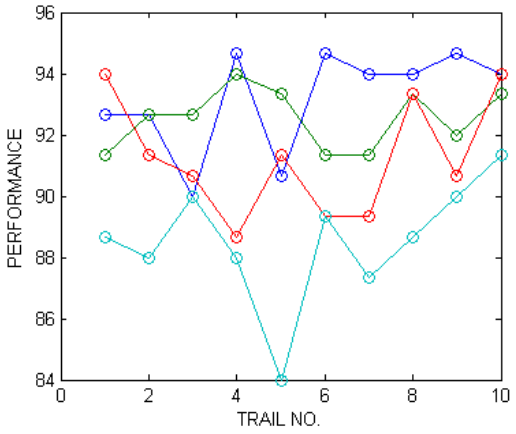


Fig .1 performance of True result for training data

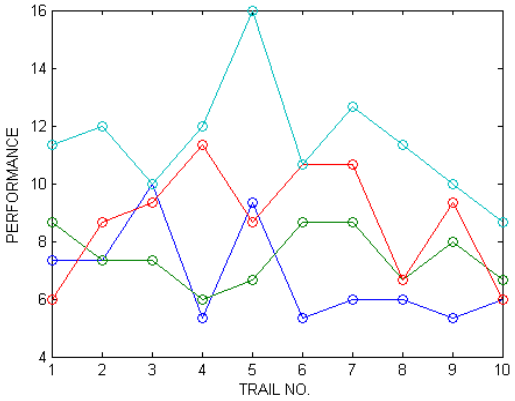


Fig.2 performance of False result for training data

Test data result

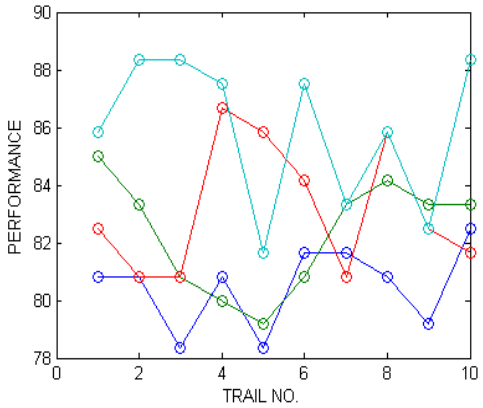


Fig.3 performance of True result for test data

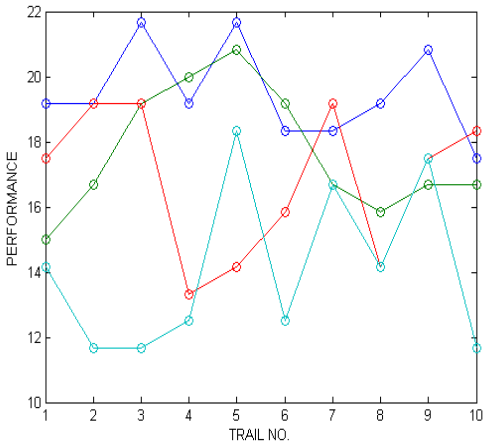


Fig .4 performance of False result for test data

TABLE 2

Performance of different methods in integration of experts.

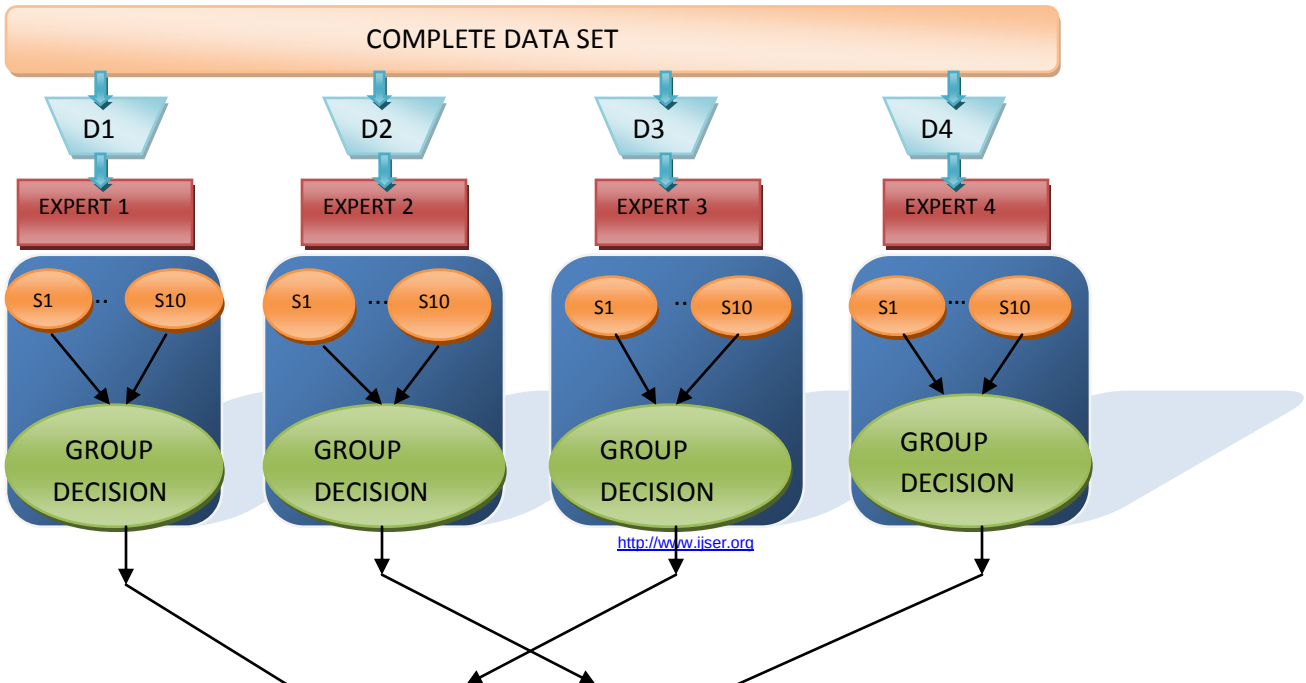
	TCR	TFR	TPV	FPV	TNV	FNV
TMQD	90.7	9.2	86.6	13.3	94.0	6.0
MVFD	91.1	8.8	87.5	12.5	94.	6.0

Fig.5 Structure and integration of classifier

Each classifier is having the computing architecture of neural network and learning completed by Genetic algorithm. For a given subgroup of data set a classifier is having independent 10 times learning. Because there are four subgroup of dataset hence in results there are four different classifiers. The construction of integration shown in fig (5).Decision of integrated classifier generated using TMQD and MVFD methods. Performances of both systems are shown in table (2).From the result obtained by both method it is appeared that MVFD performed better compare to TMQD.

4.1 Selection of cut-off values

Cut-off values for diagnostic tests can be derived using different methods amongst which the Gaussian distribution method is most commonly used. Based on this method, a cut-off value is defined as the mean plus two standard deviation (2SD) of the negative reference sample.



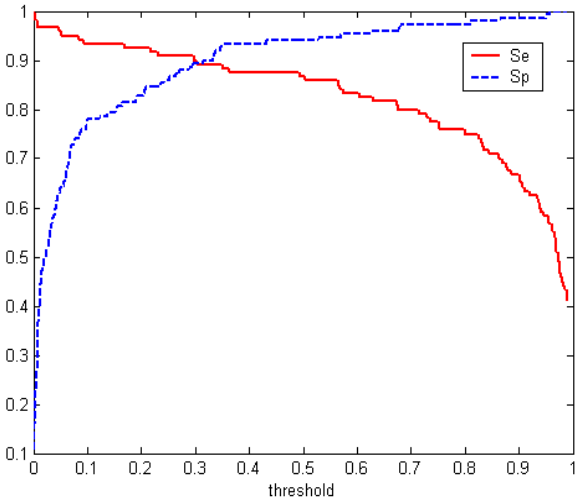


Fig.6 Threshold selection characteristics

The rationale of the 2SD procedure is to establish a cut-off value providing an SP of 97.5%.The procedure is clearly not adequate if the test values follow a skewed or multimodal distribution, as is often the case. Moreover, the procedure does not consider the Se; this is the most important disadvantage. Two parameters (SE and SP) are necessary to fully describe the probabilities of the four possible test outcome (TP, TN, FP, FN).Cut-off value and the resulting SE/SP can be obtained for a pre-selected SP(or SE).A plot of Se and Sp as function of the cut-off values as shown in fig(6) provides an useful visualization and can also be used to derive two cut-off values for the definition of intermediate test results. Optimally the cut-off selection procedure is an informed decision that takes into account the epidemiologic situation(e.g. prevalence in the target population) and the relative consequences of FP and FP test result(which may differ for every different decision making situation).As an example, given a disease of low prevalence and high cost of false-positive diagnosis, it may be advisable to choose a cut-off at the lower part of the curve to maximize SP.If on the other hand, the disease occurs at high prevalence and missing any diseased subject has serious consequences, a cut-off value towards the upper part of curve would be selected to maximize SE.

5. ROC CURVE OVERVIEW

The tradeoff at different thresholds between obtaining more true positives at the expense of additional false positives is visualized in an ROC curve by plotting the tradeoff for every possible threshold. This yields a curve like that in Figure 7. As when estimating accuracy, this plot is obtained by building a model from a set of training data and then evaluating the model against a set of test data, often within a cross-validation process. The output of the model for each case in the test data is then compared against each possible threshold, producing a point for each threshold in the plot. These points are plotted in a unit square, with the vertical location of the point for each threshold corresponding to the percentage of positive cases in the test data that are correctly labeled as positive when using the model at that threshold. The horizontal location of the point for each threshold is the percentage of negative cases in the test data that are incorrectly labeled as positive when using the model at that threshold. Note that this means neither axis represents possible thresholds, but rather the possible thresholds are distributed along the length of the curve. Given this initial description, there are several characteristics of ROC curves worth nothing. All curves start in the bottom left corner, representing a threshold at which all cases are classified as negative, and end in the upper right corner, representing a threshold at which all cases are classified as positive. Better curves are closer to the upper-left corner (if one curve is above another at a

given point on the horizontal axis, the higher curve is better at detecting true positives, while generating the same percentage of false positives as the lower curve). Curves should also always be above the diagonal (indicated in Figure 7), as a curve below the diagonal indicates that a model is generating more false positives than true positives (in which case, inverting the output of the model would provide a better model).While the information presented in an ROC curve can help a researcher choose an appropriate threshold, ROC curves are especially appealing because they allow models to be compared independent of what threshold will be used in an application. When the curve of one model is completely above the curve of another model, it is clear that the model will perform better regardless of what threshold is used. But if two curves cross, the determination of which model is better again depends on what threshold will be used. While there is no single solution to this problem in the general case, many researchers have obtained good results using the area under the ROC curve as a single measure of the quality of a model. The area under an ROC curve also has very useful statistical properties, which we will discuss later in this section.

5.1 Computing an ROC Curve

While plotting a curve over every possible threshold may sound computationally expensive, the computation is actually very simple and inexpensive. A model is first evaluated against each case in the test data, outputting larger scores to indicate greater confidence that a case is positive. The cases are then sorted by their score. All of the points in the plot can then be computed in a single pass through the sorted cases. Each distinct score encountered in this pass represents a possible threshold. A point is plotted for that threshold based on what percentage of positive cases in the test data have scores greater than or equal to the threshold and what percentage of negative cases in the test data have scores greater than or equal to the threshold. Note that these counts of positive and negative cases can be maintained during the pass through the sorted cases, so they do not need to be computed from scratch at each threshold.

5.2 Area under an ROC Curve

The area under an ROC curve is equal to the probability that a randomly selected positive case will receive a higher score than a randomly selected negative case. In this section, we present the computation of this probability, and therefore the area under the ROC curve, using pair-wise comparisons. Equations needed to work with and analyze ROC curves given in eq (7)- eq (10). When using a set of test data to estimate the probability that a randomly selected positive case will receive a higher score than a randomly selected negative case, we compare the scores assigned by a model to each case in the test set. We define a function for comparingSp, the score of a positive case, withSn, the score of a negative case, as:

$$C(S_p, S_n) = \begin{cases} 1 & if \quad S_p > S_n \\ 0.5 & if \quad S_p = S_n \\ 0 & if \quad S_p < S_n \end{cases} \tag{7}$$

We then compute the average value of this comparison function over every pair of positive and negative cases:

$$A = \frac{1}{n_p * n_n} \sum_{i=1}^{n_p} \sum_{j=1}^{n_n} C\left(s(P_i), s(N_j)\right) \tag{8}$$

Where n_p and n_n are the number of positive and negative test cases. The resulting estimate of the area under an ROC curve is known as A. As when plotting the ROC curve, A can be computed in a single pass after sorting the cases in the test data by their scores,

Wilcoxon statistic, commonly used to compare the level of a quantitative variable in two populations, will recognize that the area under the ROC curve can be analyzed in terms of A because A is equivalent to the Wilcoxon statistic .The Wilcoxon statistic is well-studied, and this equivalence means that a simple computation can be used to obtain the standard error for a given A', which we can then use to test the significance of a difference in the area under two ROC curves. Defining the terms D_p andD_n:

$$D_p = (n_p - 1)\left(\frac{A}{2-A} - A^2\right) \quad ;$$
$$D_n = (n_m - 1)\left(\frac{2*A^2}{1+A} - A^2\right) \tag{9}$$

The standard error for A is:

$$SE(A) = \sqrt{\frac{A(1-A) + D_p + D_n}{n_p * n_n}} \tag{10}$$

Using (7) and (8) we can test the significance of difference between the area under two curves using a Z test ,where Z is:

$$Z = \frac{A_1 - A_2}{\sqrt{SE(A_1)^2 + SE(A_2)^2}} \quad (11)$$

In the case where we want to test whether a model is significantly more predictive than chance, we use $A_2 = 0.5$ and $SE(A_2) = 0$ .The significance of the Z value is then checked in a table.

5.3 ROC Curve Discussion

This section has presented ROC curves and A, the area under an ROC curve, together with statistical tests for examining the significance of A and comparing values of A'. Presented visually, ROC curves allow inspection of a model's fundamental tradeoff between true information than is conveyed by a straightforward notion of accuracy. When comparing two models, ROC curves make it clear that a curve entirely above another represents a model that will perform better regardless of what threshold is used. In the case where two curves cross, A' can be used as a measure of which model is better overall. Broadly depends upon Area under ROC curve accuracy of classifier can be defined like:

(a) Non-informative if AUC <=0.5;
(b) Less accurate if $0.5 < AUC \leq 0.7$)
(c) Moderately accurate if $0.7 < AUC \leq 0.9$)
(d) Highly accurate if ($0.9 < AUC \leq 1$)
(e) Perfect if $AUC = 1$

TABLE 3
DIFFERENCE IN AREA UNDER ROC COMPARE TO CLASSIFIER TMQD

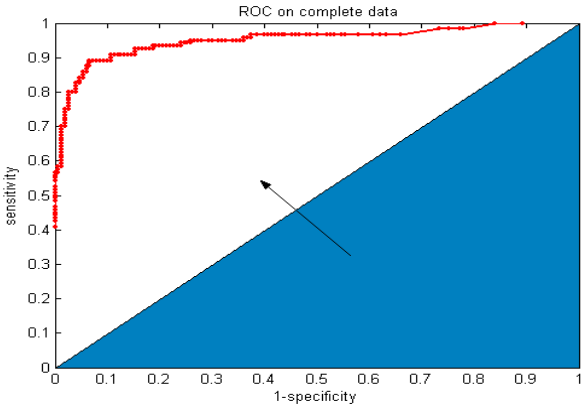
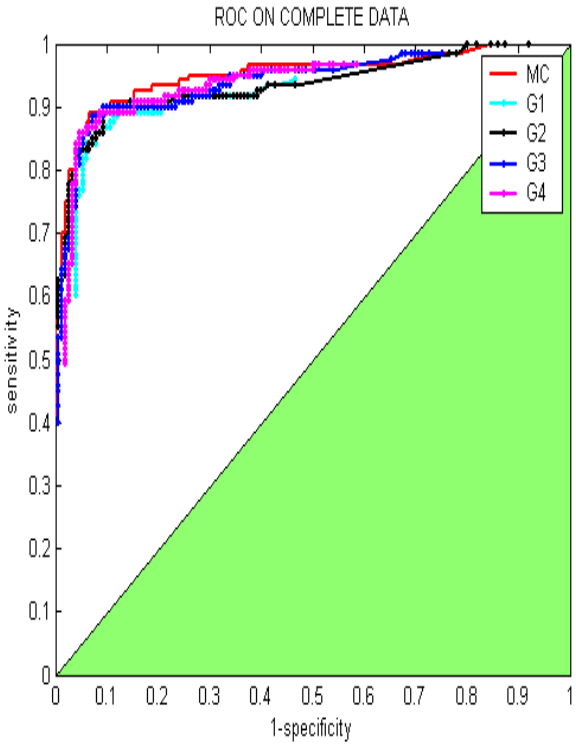


Figure (7) ROC for developed TMQD integration of classifier



Figure(8) ROC for (i) Mixture classifier (ii)G1:group1 (iii)G2:group2 (iv)G3:group3 (v) G4:group4

ROC curve for integrated classifier based on TMQD is shown in the fig (7).comparison for this classifier with other individual classifier has shown in fig (8).ROC curve of integrated classifier is always above to all ROC curve defined by individual classifier. This show the performance enhance by integrated classifier over their ingredients. Using eq. (7), to eq.(10) area under each curve and their standard error calculated. Differences in the curve estimated by Z-test as given in eq. (11) and results are shown in table (3).

Classifier	AUC	SE	Z-test value
Mixture classifier	**0.9346**	**0.0160**	-
Group1	0.9103	0.0189	0.9813
Group1	0.9153	0.0184	0.7915
Group1	0.9217	0.0177	0.5407
Group1	0.9237	0.0175	0.4597

6. CONCLUSION

Problem of designing mixture of classifier for critical applications like in heath care has presented. Two important qualities of classifier, consistency and reliability have achieved by use of combining number of classifier having knowledge over some different learning environment. This provides the facility to use the diversity in knowledge for decision. Two different methods proposed to integrate the classifiers outcome. Performance shows rather than integrating the quantitive decision of different classifiers it is better to use their final decision for integration. Advantage achieved by integrated classifier has compared with individual in terms of area under ROC curve as a performance parameter. It has seen integrated classifier outperform all individual classifier. A graphical approach of determining threshold value also presented. Decision of threshold taken by this method definitely will give lots of help in real application to decide the optimal value.

REFERENCE

[1] Chibelushi, C.C.; Deravi, F.; Mason, J.S.D.; Adaptive classifier integration for robust pattern recognition; Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions , Dec 1999 ,Volume: 29 Issue:6 : 902 - 907

[2] E. Kim and J. Ko, Dynamic Classifier Integration Method; LNCS, 2005, Volume 3541/2005, 916-917, springer.

[3] Samuel Chindaro, Konstantinos Sirlantzis and Michael Fairhurst; Modelling Multiple-Classifier Relationships Using Bayesian Belief Networks ; Lecture Notes in Computer Science, 2007, Volume 4472/2007, 312-321,

[4] Min Fang ; A Novel Multiple Classifiers Integration Algorithm with Pruning Function; Fuzzy Systems and Knowledge Discovery, 2008. FSKD '08. Fifth International Conference on 18-20 Oct. 2008 Volume:2 On page(s): 86 – 90.

[5] Shiliang Sun; Weighting Individual Classifiers by Local Within-Class Accuracies ; Lecture Notes in Computer Science, 2009, Volume 5552/2009, 802-809,

[6] R. M. Valdovinos and J. S. Sánchez; Combining Multiple Classifiers with Dynamic Weighted Voting ; Lecture Notes in Computer Science, 2009, Volume 5572/2009, 510-516

[7] Peter Revesz,Thomas Triplet; Classification integration and reclassification using constraint databases. Artificial Intelligence in Medicine Volume 49, Issue 2 , Pages 79-91, June 2010, Elsevier

[8] Enrique F. Schisterman; Statistical Inference for the Area under the Receiver Operating Characteristic Curve in the Presence of Random Measurement Error. Am. J. Epidemiol. (2001) 154 (2): 174-179

[9] Alan Herschtal Telstra Corporation, Clayton, Victoria, Australia Bhavani Raskutti Optimising area under the ROC curve using gradient descent; ICML '04 Proceedings of the twenty-first international conference on Machine learning ACM New York, NY, USA ©2004.

[10] Lee, W.-H. Gader, P. D. Wilson, J. N. Optimizing the Area Under a Receiver Operating Characteristic Curve With Application to Landmine Detection ; Geoscience and Remote Sensing, IEEE Transactionson
Issue Date: Feb. 2007 Volume: 45 Issue: 2 On page(s): 389 – 397

[11] Atapattu, S.; Tellambura, C.; Hai Jiang; Analysis of area under the ROC curve of energy detection; Wireless Communications, IEEE Transactions on Issue Date: March 2010 Volume: 9 Issue:3 On page(s): 1216 – 1225.

[12] .Kavitha ,Ramakrishnan ,Singh," Modeling and design of evolutionary neural network for heart disease detection". IJCSI International Journal of Computer Science Issues, Vol. 7, Issue 5, September 2010 ,ISSN (Online): 1694-0814

[13] Heart attack dataset from
http://archive.ics.uci.edu/ml/datasets/Heart+Disease