Communications for Statistical Applications and Methods
2019, Vol. 26, No. 2, 79–89

https://doi.org/10.29220/CSAM.2019.26.2.079
Print ISSN 2287-7843 / Online ISSN 2383-4757

Selection of markers in the framework of multivariate receiver operating characteristic curve analysis in binary classification

Sameera G[a], Vishnu Vardhan R[1,a]

[a]Department of Statistics, Pondicherry University, India

Abstract

Classification models pertaining to receiver operating characteristic (ROC) curve analysis have been extended from univariate to multivariate setup by linearly combining available multiple markers. One such classification model is the multivariate ROC curve analysis. However, not all markers contribute in a real scenario and may mask the contribution of other markers in classifying the individuals/objects. This paper addresses this issue by developing an algorithm that helps in identifying the important markers that are significant and true contributors. The proposed variable selection framework is supported by real datasets and a simulation study, it is shown to provide insight about the individual marker's significance in providing a classifier rule/linear combination with good extent of classification.

Keywords: multivariate receiver operating characteristic curve, precision, stepwise algorithm, variable selection

1. Introduction

The practical problems pertaining to binary classification have paved a path to statistical methodologies with a strong mathematical base. The problem of classifying individuals/objects can be done using a marker or a set of markers. Many researchers have developed methodologies over the past seven decades that use univariate and multivariate setup. The present paper is confined to a graphical tool that provides an aid to allocate individuals into one of two known populations/groups is the namely receiver operating characteristic (ROC) curve analysis. Many ideas were proposed on this ROC curve methodology under univariate setup of which Bamber (1975), Metz (1978), Hanley and McNeil (1982, 1983), Faraggi and Reiser (2002), Zhang (2006), Vishnu Vardhan and Sarma (2010), Balaswamy *et al.* (2014), or Vishnu Vardhan and Kiruthika (2015) are a few to mention. However, in a practical sense classification may not be appropriate by using a single marker since the information obtained using single marker might be insufficient to draw proper conclusions. Hence, it became necessary to consider available multiple markers in order to have a complete profile of an individual/object that enables to come across a strong mathematical basis and provides an ease for better understanding about the classification scenario (Su and Liu 1993; Liu *et al.*, 2005; Gao *et al.*, 2008; Sameera *et al.*, 2016). However, an extensive inquiry is required when several markers are being considered to classify an individual in order to overcome some practical issues such as increase in error rate, improper

[1] Corresponding author: Department of Statistics, Pondicherry University, R.V.Nagar, Kalapet, Puducherry - 605014, India. E-mail: rvvcrr@gmail.com

Published 31 March 2019 / journal homepage: http://csam.or.kr

use of variables, cost effectiveness, misinterpretation and misleading conclusions. This leads to the use of an important procedure to choose a subset of markers that are validated and support the criteria of the investigator. In statistical literature, such a procedure is coined as the 'screening' of important markers which is a well-known principle. Usually the screening process refers to thorough monitoring on the role of markers by means of forward, backward and stepwise algorithms that contribute towards minimizing the error rate and address the practical issues mentioned above. Sameera *et al.* (2016) proposed a Multivariate extension of ROC methodology to provide a linear combination using Minimax principle and proved that their model performs better than the model given by Su and Liu (1993). The process of selecting a subset of important markers was not addressed till date. This paper develops a stepwise algorithm for the multivariate ROC (MROC) curve (Sameera *et al.*, 2016). The proposed methodology encompasses MROC framework, precision of the linear combination (PLC) (Sameera and Vishnu Vardhan, 2016) and a strategy designed for variable selection. The schematic view of the paper is divided into two parts. The first part details the MROC and PLC methodologies; the second provides a variable selection algorithm that uses the first part and its practical implications.

2. Methodology

2.1. Multivariate receiver operating characteristic methodology

Let $X_1, X_2, \ldots, X_k$ be the 'k' markers involved in the study. Let Π_0 and Π_1 be the two populations assumed to follow multivariate normal distribution with mean vectors μ_0, μ_1; covariance matrices Σ_0, Σ_1 and sample sizes n_0, n_1 respectively and $n = n_0 + n_1$. The basic definition of MROC curve is that it is a tradeoff between 1 − Specificity $(1 - S_p)$ and sensitivity (S_n). The expression for MROC curve is given as

$$y(x) = \Phi\left(\frac{b'(\mu_1 - \mu_0) - \Phi^{-1}(1-x)\sqrt{(b'\Sigma_0 b)}}{\sqrt{(b'\Sigma_1 b)}}\right), \tag{2.1}$$

where x is $(1 - S_p)$ and b is the vector of coefficients of linear combination of markers given as $b = [t\Sigma_1 + (1-t)\Sigma_0]^{-1}(\mu_1 - \mu_0)$; $0 < t < 1$ where t-value is obtained using trial and error method and optimal 't' is identified with the help of Youden's index, $J = \max[S_n + S_p - 1]$. S_n and S_p are the probabilities of correct identification of the two groups '0' and '1' respectively.

$$S_n = \Phi\left(\frac{b'\mu_1 - c}{\sqrt{(b'\Sigma_1 b)}}\right), \tag{2.2}$$

$$S_p = \Phi\left(\frac{c - b'\mu_0}{\sqrt{(b'\Sigma_0 b)}}\right), \tag{2.3}$$

where

$$c = \frac{b'\mu_1\sqrt{(b'\Sigma_0 b)} + b'\mu_0\sqrt{(b'\Sigma_1 b)}}{\sqrt{(b'\Sigma_1 b)} + \sqrt{(b'\Sigma_0 b)}}$$

is the cut point obtained at optimal 't'. Therefore, 'c' can be termed as optimal cut point. A test's performance can be explained by using an accuracy measure, area under the curve (AUC) which is defined as average sensitivity over the range of specificities and given as

$$\text{AUC} = \Phi\left(\frac{b'(\mu_1 - \mu_0)}{\sqrt{b'(\Sigma_1 + \Sigma_0)^{-1}b}}\right). \tag{2.4}$$

AUC lies between 0 and 1 and a test is said to be *perfect test* if its AUC is equal to 1; in addition, a test's *worst scenario* can be observed when AUC = 0.5.

2.2. Precision of linear combination

Once the linear combination of markers is obtained, its validity can be checked using the concept of PLC proposed by Sameera and Vishnu Vardhan (2016). PLC is based on an F ratio that comprises correlation measures to determine if the obtained linear combination is adequate for classifying individuals/objects into one of the two classes. The F ratio is obtained for testing the significance of the contribution of samples after a linear combination is considered through analysis of variance method,

$$F = \frac{n_1 + n_0 - k - 1}{k - 1} \frac{R^2(1 - r^2)}{(1 - R^2)(1 - r^2)} \sim F_{(k-1, n_1+n_0-k-1)}. \tag{2.5}$$

Here, 'R^2' indicates the multiple correlation coefficient and 'r^2' indicates the correlation within samples. The present paper is dedicated in developing a stepwise methodology to select 'p' important markers among a given set of 'k' markers using the concept of PLC. The reason for considering PLC as selection criterion is that it is specific to the MROC model and determines the precision with which the linear combination can classify an individual/object. The main objective is to combine markers to identify the classes better the use of this F ratio given in equation (2.5) is justified. In addition to testing the significance of the linear combination, it is mandatory to test the significance of individual markers in a variable selection. To meet this purpose, the partial F-statistic is formulated to validate the importance and role of individual markers in the linear combination. The partial F for marker 'l' when 'k' markers are considered in the model is defined in (2.6), where $R^2_{l.12,\ldots,l-1,l+1,\ldots,k}$ is the multiple correlation coefficient between marker 'l' on markers '$1, 2, 3, \ldots, l-1, l+1, \ldots, k$'.

$$\text{Partial } F_{l.12,\ldots,l-1,l+1,\ldots,k} = \frac{n_1 + n_0 - k - 1}{k - 1} \frac{R^2_{l.12,\ldots,l-1,l+1,\ldots,k}(1 - r^2)}{\left(1 - R^2_{l.12,\ldots,l-1,l+1,\ldots,k}\right)(1 - r^2)} \sim F_{(k-1, n_1+n_0-k-1)}. \tag{2.6}$$

2.3. Marker (variable) selection - stepwise algorithm

In usual variable selection algorithms, initially a single variable will be included in the model for significance. The same logic cannot be applied in the proposed stepwise algorithm since it is based upon the concept of Precision which depends on correlation measures. Hence, it is imperative to begin with a pair of markers (variables) instead of a single marker. The following steps detail the procedural flow of how the algorithm is executed.

• Algorithm

1. List out the $\binom{k}{2}$ combinations of markers.

2. Compute the F ratio using equation (2.5) for the listed $\binom{k}{2}$ combinations and select the combination that has the highest F-value which exceeds $F_{(1, n_0+n_1-2)}$ at a fixed level of $\alpha = 0.05$ (say). If none of the combinations have a significant F-value, select the combination that has highest F-value among the combinations. This is to allow a pair of markers to initially enter the model and be eliminated in the later stages if observed as insignificant. Let the markers included in the model be 'k_1' and 'k_2'.

3. In order to select the next marker that can be included into the model compute partial F using equation (2.6) for each of the remaining markers. Select the marker with maximum partial F-value

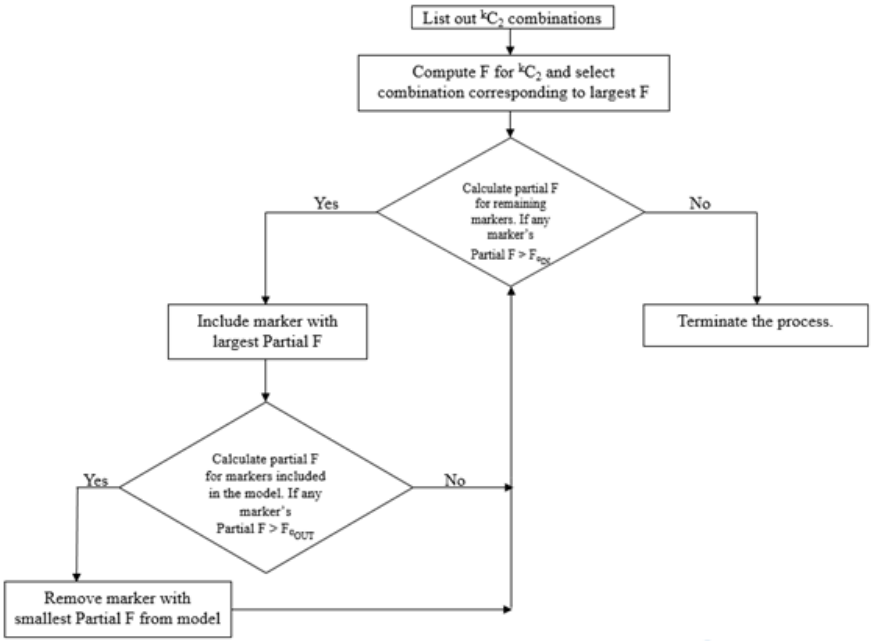


Figure 1: *Logical process of the stepwise algorithm.*

which is significant at fixed level α_{IN}. If no markers are found to be significant, then the process is terminated and the linear combination cannot be considered if its F ratio is insignificant.

4. Now, conducting a backward elimination step will help to remove any insignificant marker that is added to the model. To achieve this, compute the partial F of markers which are included in the model and remove that marker which has the least insignificant partial F ratio at a fixed level α_{OUT}.

5. Repeat steps 3 and 4 till a stage comes where no marker can be added into/removed from the model.

Figure 1 details out the logical process of the proposed stepwise algorithm.

Using the algorithm given in Figure 1, a subset of markers can be identified and linearly combined to arrive at a classification rule for classification of new individuals/objects. For illustration purposes in real datasets, the α_{IN} and α_{OUT} values are considered to be 0.05 and 0.10 respectively. However, the choice of α_{IN} and α_{OUT} can be varied upon experimentation. The next section depicts the use of the proposed methodology with the help of real and simulation datasets.

3. Results and discussion

3.1. Real datasets

The functionality of the proposed stepwise algorithm is demonstrated using three datasets by Statlog Heart data (Michie *et al.*, 1994), Norton - Neonatal audiology data (Norton *et al.*, 2000) and Vertebral Column data (Guilherme and Ajalmar 2011). Further, the Vertebral column dataset contains three categories of individual status. For the present context of binary classification, it is divided into two

segments, one with samples that belong to Spondylolisthesis (SL) and Normal individuals and the other comprised of Normal and Disk Hernia individuals.

A neonatal dataset is used to provide better understanding about the iterative steps of the proposed stepwise algorithm. An R code is developed to perform the stepwise algorithm where the first column in the data frame holds the status of the individual and subsequent columns contain markers in the following order: 2 = ear, 3 = sitenum, 4 = currage, 5 = gender, 6 = DPOAE 65 at 2kHz ($y1$), 7 = TEOAE 80 at 2kHz ($y2$), and 8 = ABR ($y3$).

- Iteration 1: There are 7 markers in the study, 7C2 = 21 combinations are listed. On computing the F ratio (p-values) for the listed combinations, the combination 6, 8 is observed to have the highest significant F ratio (F ratio = 10.2102, p-value = 0.001). Hence, the markers in the model are DPOAE 65 at 2kHz ($y1$) and ABR ($y3$).

- Iteration 2: The process of forward selection is executed for the variables 2, 3, 4, 5, and 7. At this stage, the corresponding F and p-values are computed and marker 7, which has highest partial F (partial F ratio = 561.7274, p-value = 0.000) is appended to the previously identified combination 6, 8. Now, the subset of markers in the model is 7, 6, 8 = TEOAE 80 at 2kHz ($y2$), DPOAE 65 at 2kHz ($y1$), ABR ($y3$).

- Iteration 3: Backward elimination is executed in this iteration to examine if the included markers 7, 6, and 8 are significant enough to remain in the model. The partial F and p-values for these markers are found to be significant and are retained.

- Iteration 4: The process of forward selection is repeated again on markers 2, 3, 4, and 5 of which marker 2 is included into the model because of its significant partial F-value (partial F ratio= 4.3488, p-value = 0.004).The markers in the model at this stage are 2, 7, 6, 8 = ear, TEOAE 80 at 2kHz ($y2$), DPOAE 65 at 2kHz ($y1$), ABR ($y3$).

- Iteration 5: This step verifies the significance of the included markers; consequently, it is shown that all are significant and allowed to remain in the model.

- Iteration 6: In this step, the significance of the markers 3, 4, and 5 is tested for possible inclusion into the model. As none of the markers are significant, the iterative procedure is terminated and the final list of markers involved in the model contain 2, 7, 6, and 8 i.e., ear, DPOAE 65 at 2kHz (y1), TEOAE 80 at 2kHz (y2), and ABR ($y3$).

In addition to the above described iterative procedure for the Neonatal dataset, the model's significance is tested using PLC. The full model is observed to provide an accuracy of 68.39% but the linear combination obtained in this case is found to have an insignificant F-value (1.1953 sig. = 0.305). The application of a stepwise algorithm helped identify a subset of 4 markers and the linear combination is observed to be significant with an accuracy of 66.07%. However, the AUC of a linear combination can be considered and interpreted accurately only if its corresponding linear combination is significant. Further, the use of such an insignificant linear combination leads to the misinterpretation of markers and misleads the conclusions. It is better to consider the AUC whose linear combination is proven to be significant along with individual markers' significance. The results obtained for the Heart dataset show that the accuracy is 93.66% when all 13 markers are included in the full model. However, the linear combination in this case does not have a significant F-value. On the use of proposed algorithm, 7 markers are included in the model and observed to have a significant/validated linear combination.

Table 1: Coefficients and partial F-values for full model and stepwise model - real datasets

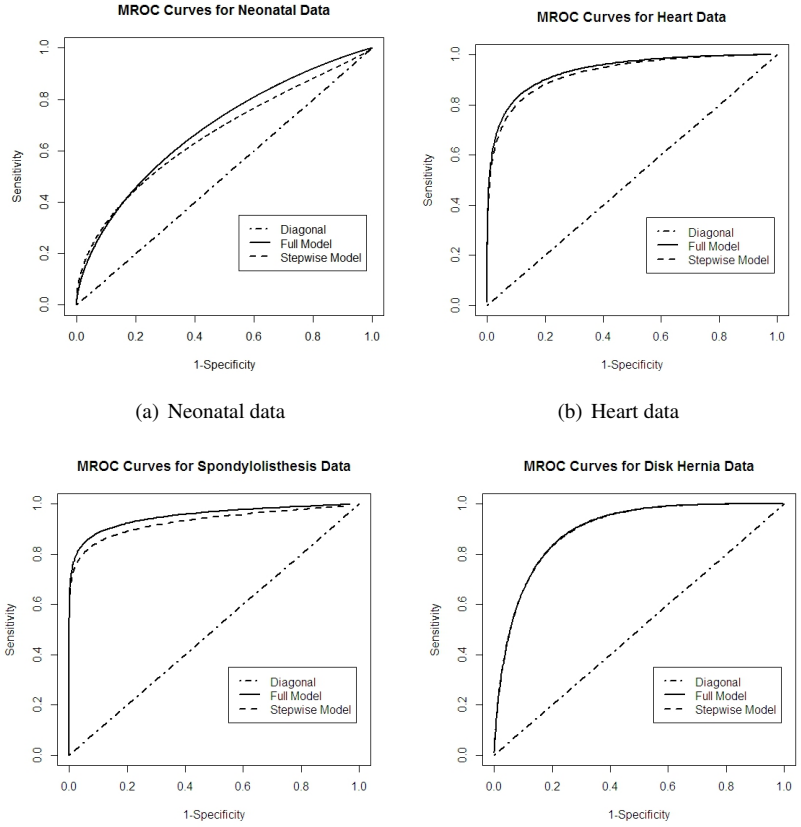
Dataset	Full model				Stepwise model			
	Variables	Coefficients	Partial F	sig.	Variables	Coefficients	Partial F	sig.
Neonatal (n = 5056)	Ear	−0.0963	0.4878	0.818	Ear	−0.1100	4.3488	0.005
	sitenum	0.2264	5.6181	0.000	$y2$	0.0258	267.4902	0.000
	currage	0.0295	5.6410	0.000	$y1$	0.0405	268.1827	0.000
	gender	−0.0350	0.2889	0.942	$y3$	0.1574	7.0262	0.000
	$y1$	0.0410	29.7824	0.000				
	$y2$	0.0210	29.8079	0.000				
	$y3$	0.1817	1.3841	0.217				
	Model F (sig.) = 1.1953 (0.305NS)				Model F (sig.) = 5.3341 (0.001*)			
Heart (n = 270)	Age	−0.0220	0.2128	0.998	nMBV	1.2523	2.2158	0.042
	rBP	0.0187	0.0906	1.000	CPT	0.8904	3.2811	0.004
	SC	0.0051	0.0712	1.000	EIA	1.3849	4.2655	0.000
	MHR	−0.0252	0.2792	0.992	SPSTs	0.5759	9.8091	0.000
	Oldpeak	0.4243	0.3525	0.978	Oldpeak	0.5194	11.2808	0.000
	nMBV	1.2687	0.1359	1.000	Sex	1.0877	3.1066	0.006
	thal	0.5344	0.2190	0.997	Thal	0.5123	7.0360	0.000
	Sex	1.3227	0.1355	1.000				
	CPT	0.8294	0.1260	1.000				
	FBS	−0.7239	0.0349	1.000				
	rECGr	0.3584	0.0397	1.000				
	EIA	1.0912	0.1590	0.999				
	SPSTs	0.3983	0.3353	0.982				
	Model F (sig.) = 0.4897 (0.920NS)				Model F (sig.) = 7.2033 (0.000*)			
Spondylolisthesis (n = 250)	PI	5.3239	21972280	0.000	PT	−12.3646	26133190	0.000
	PT	−5.3321	8470798	0.000	SS	−12.3594	37534480	0.000
	LLA	0.0903	1.9174	0.092	PI	12.3776	67786610	0.000
	SS	−5.3588	12166540	0.000	LLA	0.0740	5.8392	0.000
	PR	−0.1123	0.5763	0.718	GS	0.1282	4.5012	0.002
	GS	0.1006	1.5598	0.172				
	Model F (sig.) = 2.6084 (0.026*)				Model F (sig.) = 3.4729 (0.009*)			
Disk Hernia (n = 160)	PI	14.8544	52420430	0.000	LLA	−0.0321	18.3480	0.000
	PT	−14.7483	19189290	0.000	PT	−14.7275	24337630	0.000
	LLA	−0.0368	15.2383	0.000	PI	14.8343	66484970	0.000
	SS	−15.0232	35695070	0.000	SS	−15.0035	45272420	0.000
	PR	−0.1483	2.1034	0.068	PR	−0.1484	2.6678	0.034
	GS	0.0356	0.8572	0.511				
	Model F (sig.) = 3.7778 (0.003*)				Model F (sig.) = 4.7639 (0.001*)			

The accuracy of this stepwise model is 92.59% which when compared to the accuracy of the full model leads to a conclusion that the stepwise model is sufficient for classification. In some situations, we may come across a case where the linear combination will be significant for the full model. But it requires further investigation of the individual markers' significance. The support of datasets from the Vertebral Column data is taken to show this kind of scenario. With respect to the Spondylolisthesis and Disk Hernia datasets, the linear combinations of full model are noticed to be significant at 94.77% and 89.61% respectively. However, 3 markers (LLA, PR, and GS) from Spondylolisthesis and 2 markers (PR and GS) from Disk Hernia are found to have insignificant partial F-values. In Spondylolisthesis dataset, the elimination of marker PR from the model resulted in the extraction of the actual contribution of other markers LLA and GS, which are observed to be significant in a stepwise model. Similarly in the Disk Hernia dataset, the removal of marker GS from the model helped extract the true contribution of marker PR. The accuracies of stepwise models for Spondylolisthesis and Disk Hernia datasets are 91.92% and 89.49%, respectively with minimum discrepancy from full

Table 2: Measures of MROC curve and optimal cutpoint - Real datasets

Dataset	Model	Opt c	$(1-S_p, S_n)$	AUC
Neonatal	Full	0.6012	(0.3673, 0.6327)	0.6839
	Stepwise	−1.2829	(0.3838, 0.6162)	0.6607
Heart	Full	7.2733	(0.1377, 0.8623)	0.9366
	Stepwise	8.6494	(0.1508, 0.8492)	0.9259
Spondylolisthesis	Full	−9.1092	(0.1057, 0.8943)	0.9477
	Stepwise	6.0845	(0.1259, 0.8741)	0.9192
Disk Hernia	Full	−23.2065	(0.1845, 0.8155)	0.8961
	Stepwise	−23.1334	(0.1859, 0.8141)	0.8949

MROC = multivariate receiver operating characteristic; AUC = area under the curve.



(a) Neonatal data (b) Heart data

(c) Spondylolisthesis data (d) Disk Hernia data

Figure 2: *MROC curves for full model and stepwise method - Real Datasets. MROC = multivariate receiver operating characteristic*

models. This example shows the need to validate every marker included in the model even though a significant linear combination is witnessed.

Table 1 provides the coefficients and partial F along with significance that are computed for all datasets; in addition, Table 2 lists the measures of the MROC curve and the optimal cut point for the considered datasets. Accordingly, Figure 2 visualizes the MROC curves. The MROC curves which

are displayed in Figure 2 support the above given description.

3.2. Simulation study

Simulation studies detail the algorithmic proposed for variable selection as well as observe the sensitivity of α_{IN} and α_{OUT}. Here, two datasets were generated and two stepwise models were produced by considering two combinations for the level of significance: $\alpha_{\text{IN}} = 0.05$ & $\alpha_{\text{OUT}} = 0.10$ (Model 1) and $\alpha_{\text{IN}} = 0.01$ & $\alpha_{\text{OUT}} = 0.05$ (Model 2). The mean vectors and covariance matrices for groups '0' and '1' are

$$\mu_0 = \begin{pmatrix} 1.4986 \\ 3.3835 \\ 38.4541 \\ 1.5581 \\ -8.9171 \\ -11.7704 \\ -3.8821 \end{pmatrix}, \qquad \mu_1 = \begin{pmatrix} 1.4631 \\ 3.7986 \\ 38.5984 \\ 1.5704 \\ -4.8718 \\ -7.9295 \\ -3.2761 \end{pmatrix},$$

$$\Sigma_0 = \begin{pmatrix} 0.2500 & 0.0048 & -0.0120 & 0.0002 & 0.0936 & -0.1631 & 0.0179 \\ 0.0048 & 2.8920 & -1.4424 & -0.0087 & 0.3218 & 0.3472 & -0.2406 \\ -0.0120 & -1.4424 & 11.6826 & 0.0466 & -2.3629 & -1.9465 & 0.0230 \\ 0.0002 & -0.0087 & 0.0466 & 0.2467 & 0.0515 & 0.1651 & 0.0073 \\ 0.0936 & 0.3218 & -2.3629 & 0.0515 & 60.5546 & 28.0952 & 1.0964 \\ -0.1631 & 0.3472 & -1.9465 & 0.1651 & 28.0952 & 49.2620 & 0.7217 \\ 0.0179 & -0.2406 & 0.0230 & 0.0073 & 1.0964 & 0.7217 & 2.8415 \end{pmatrix},$$

$$\Sigma_1 = \begin{pmatrix} 0.2503 & -0.0210 & -0.0389 & 0.0111 & -0.0084 & -0.7241 & -0.0304 \\ -0.0210 & 2.0673 & -1.6133 & 0.0008 & -1.0247 & -1.2006 & -0.6255 \\ -0.0389 & -1.6133 & 11.0467 & 0.1960 & 3.2611 & 4.5973 & 1.1885 \\ 0.0111 & 0.0008 & 0.1960 & 0.2467 & 0.0149 & 0.4264 & 0.1159 \\ -0.0084 & -1.0247 & 3.2611 & 0.0149 & 73.6316 & 59.3241 & 3.7913 \\ -0.7241 & -1.2006 & 4.5973 & 0.4264 & 59.3241 & 97.4734 & 5.0979 \\ -0.0304 & -0.6255 & 1.1885 & 0.1159 & 3.7913 & 5.0979 & 2.9501 \end{pmatrix}.$$

Table 3 provide the coefficients, partial F-values along with significance and the precision of each model.

In simulation 1, the full model with 7 markers is observed to have an insignificant F ratio (0.7883 sig. = 0.579). The stepwise models 1 and 2 are observed to be significant and contain a subset of three significant markers that provide 64.86% accuracy. In simulation 2, the full model is insignificant (F ratio = 0.7748, sig. = 0.589) with an AUC of 63.73%. Stepwise Model 1 is observed be to significant (F ratio = 4.0076, sig. = 0.007) and includes 4 markers with an accuracy of 63.72% while Stepwise Model 2 includes only 3 markers with an accuracy of 63.67% with a loss of 0.1% accuracy compared to the Stepwise Model 1. The simulation show the need for the identification of significant markers to understand true classification ability of the model. The MROC curves are depicted in Figure 3.

4. Discussion

The present paper provides a variable selection procedure for a multivariate extension of ROC curve technique known as MROC curve analysis. The idea behind the concept is to identify a subset of

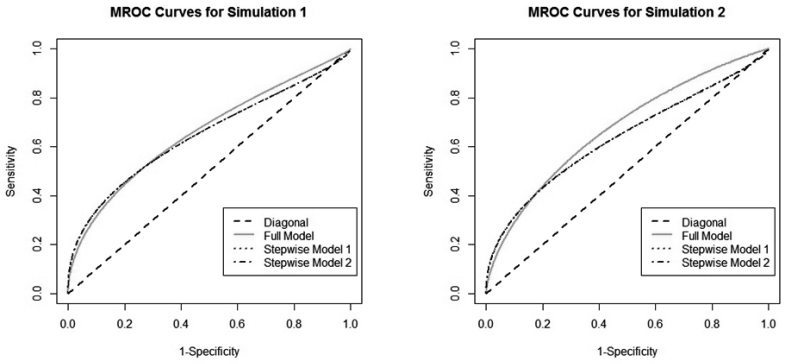
Table 3: Coefficients and partial F-values for full model and stepwise model - Simulation datasets

Dataset	Simulation 1 (n = 2200)				Simulation 2 (n = 1000)			
	Variables	Coefficients	Partial F	sig.	Variables	Coefficients	Partial F	sig.
Full Model	$X1$	0.2665	0.2134	0.973	$X1$	−0.1008	0.2621	0.954
	$X2$	0.0881	2.1755	0.043	$X2$	0.2519	0.7317	0.624
	$X3$	0.0400	0.1913	0.979	$X3$	0.0550	0.6900	0.658
	$X4$	−0.0148	2.2509	0.036	$X4$	0.1015	0.1338	0.992
	$X5$	0.0253	12.8950	0.000	$X5$	0.0299	6.2444	0.000
	$X6$	0.0338	12.9578	0.000	$X6$	0.0152	6.5322	0.000
	$X7$	0.1607	0.7327	0.623	$X7$	0.1947	0.7925	0.576
	Model F (sig.) = 0.7883 (0.579NS)				Model F (sig.) = 0.7748 (0.589NS)			
Stepwise Model 1 (α_{IN} = 0.05 & α_{OUT} = 0.10)	$X5$	0.0269	306.0346	0.000	$X1$	−0.0855	2.9122	0.033
	$X6$	0.0306	302.3586	0.000	$X5$	0.0217	79.2230	0.000
	$X7$	0.1456	9.7466	0.000	$X6$	0.0281	76.0753	0.000
					$X7$	0.1696	7.4475	0.000
	Model F (sig.) = 6.037 (0.002*)				Model F (sig.) = 4.0076 (0.007*)			
Stepwise Model 2 (α_{IN} = 0.01 & α_{OUT} = 0.05)	$X5$	0.0269	306.0346	0.000	$X5$	0.0273	171.8816	0.000
	$X6$	0.0306	302.3586	0.000	$X6$	0.0226	174.2524	0.000
	$X7$	0.1456	9.7466	0.000	$X7$	0.1696	17.4699	0.000
	Model F (sig.) = 6.037 (0.002*)				Model F (sig.) = 5.274 (0.005*)			

Table 4: Measures of MROC curve and optimal cutpoint - simulation datasets

Dataset	Model	Opt c	$(1 - S_p, S_n)$	AUC
Simulation 1	Full	−1.0277	(0.3918, 0.6081)	0.6486
	Stepwise 1	−1.0277	(0.3918, 0.6081)	0.6486
	Stepwise 2	−1.0277	(0.3918, 0.6081)	0.6486
Simulation 2	Full	−1.1141	(0.4007, 0.5992)	0.6373
	Stepwise 1	−1.1784	(0.4006, 0.5993)	0.6372
	Stepwise 2	−1.0570	(0.4010, 0.5989)	0.6367

MROC = multivariate receiver operating characteristic; AUC = area under the curve.



(a) MROC curves - Simulation 1 (b) MROC curves - Simulation 2

Figure 3: *MROC curves for full model and stepwise method - simulation datasets. MROC = multivariate receiver operating characteristic.*

markers that provide true accuracy and a valid classifier rule/linear combination when combined. The proposed stepwise methodology is supported with the help of real datasets and a simulation study. Two cases are discussed, one when the obtained linear combination is insignificant and required to identify

the subset that can provide a significant linear combination (Neonatal dataset and Heart dataset) and the other when the linear combination is significant, but also contains insignificant markers that influence the performance of the linear combination (Spondylolisthesis and Disk Hernia datasets). The sensitivity of α_{IN} and $\alpha_{\text{OUT}}$are observed using simulation studies. The observed results show that the linear combination obtained using a stepwise algorithm is observed to have a better significance than the full model. The algorithm identified significant markers that help classify individuals into two groups.

Acknowledgements

The author, Sameera G, would like to acknowledge Department of Science and Technology for supporting her research through a fellowship under DST-INSPIRE programme (IF130958).

References

Balaswamy S, Vishnu Vardhan R, and Rao MB (2014). A divergence measure for STROC curve in binary classification, *Journal of Advanced Computing*, **3**, 68–81.

Bamber D (1975). The area above the ordinal dominance graph and the area below the receiver operating characteristic graph, *Journal of Mathematical Psychology*, **12**, 387–415.

Faraggi D and Reiser B (2002). Estimation of the area under the ROC curve, *Statistics in Medicine*, **21**, 3093-3106.

Gao F, Xiong C, Yan Y, Yu K, and Zhang Z (2008). Estimating optimum linear combination of multiple correlated diagnostic tests at a fixed specificity with receiver operating characteristic curves, *Journal of Data Science*, **6**, 1–13.

Guilherme de Alencar Barreto and Ajalmar RÃ^{a}go da Rocha Neto (2011). *UCI Machine Learning Repository*. Fortaleza, Ceará, Brazil: Department of Teleinformatics Engineering, Federal University of Ceará. Available from: https://archive.ics.uci.edu/ml/datasets/Vertebral+Column

Hanley JA and McNeil BJ (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve, *Radiology*, **143**, 29–36.

Hanley JA and McNeil BJ (1983). A method of Comparing the Areas under Receiver Operating Characteristic Curves Derived from the Same Cases, *Radiology*, **148**, 839–843.

Liu A, Schisterman EF, and Zhu Y (2005). On linear combinations of Biomarkers to improve diagnostic accuracy, *Statistics in Medicine*, **24**, 37–47

Metz CE (1978). Basic principles of ROC analysis, *Seminars in Nuclear Medicine*, **8**, 283–298.

Michie D, Spiegelhalter DJ, and Taylor CC (1994). *UCI Machine Learning Repository*. Machine Learning, Neural and Statistical Classification, Ellis Horwood Limited. Available from: https://archive.ics.uci.edu/ml/datasets/Statlog+%28Heart%29

Norton SJ, Gorga MP, Widen JE, *et al.* (2000). Identification of neonatal hearing impairment: evaluation of transient evoked ototacoustic emission, distortion product otoacoustic emission, and auditory brain stem response test performance, *Ear Hearing*, **21**, 508–528.

Sameera G and Vishnu Vardhan R (2016). Testing the precision of linear combination of an multivariate ROC (MROC) model. In *Proceedings of National Conference entitled "Statistical Modelling and Analysis Techniques"*, NAROSA Publications, 103–110.

Sameera G, Vishnu Vardhan R, and Sarma KVS (2016). Binary classification using multivariate receiver operating characteristic curve for continuous data, *Journal of Biopharmaceutical Statistics*, **26**, 421–431.

Su JQ and Liu JS (1993). Linear combinations of multiple diagnostic markers, *Journal of American*

Statistical Association, **88**, 1350–1355.

Vishnu Vardhan R and Kiruthika C (2015). Properties and the use of half normal distribution in ROC curve analysis, *IAPQR Transactions*, **39**, 169–179.

Vishnu Vardhan R and Sarma KVS (2010). Estimation of the area under the ROC curve using confidence intervals of mean, *ANU Journal of Physical Sciences*, **2**, 29–39.

Zhang B (2006). A semi parametric hypothesis testing procedure for the ROC curve area under a density ratio model, *Computational Statistics and Data Analytics*, **50**, 1855–1876.

Received August 3, 2018; Revised November 17, 2018; Accepted January 7, 2019