A Generalized Multi-Group Discriminant Function Procedure for Classification: an Application To Ten Groups Of Yam Species

Anyanwu Paul E.
Department of Mathematics & Statistics, Federal Polytechnic, P. M. B. 55, Bida, Niger State, Nigeria.

Ekezie Dan Dan
Department of Statistics, Imo State University, P. M. B. 2000, Owerri, Imo State, Nigeria.

Onyeagu I. Sidney
Department of Statistics, Nnamdi Azikiwe University, Awka, Anambra State, Nigeria.

Nwankwo, I. I.M
National Root Crops Research Institute, Umudike Umuahia, Abia State Nigeria

ABSTRACT

Multivariate Analysis (MVA) is based on the Statistical principle of Multivariate Statistics which involves observation and analysis of more than one Statistical outcome variables at a time. Classification in Multivariate analysis deals with developing a statistical rule for allocating observation to one or more groups. A closely associated multivariate technique is discriminant analysis which predicts group membership for an observation. Fishers (1936) developed a technique (Fishers Linear Discriminant Function) that optimally discriminate only two groups. The challenges of developing a mathematical based procedure with some underlying distribution for multiple groups have remained a task to be accomplished as it only exist in theory but not in practice. Owing to these challenges, this work introduces and suggests a mathematical procedure that is based on combinatorial analysis which gave rise to All Possible Pair of functions and allocation rules for a multiple group case. The developed procedure was generalized and applied to both real and simulated data. The developed procedure gave a higher accuracy rate for the real and simulated data under various sample sizes when compared with other conventional methods. It is therefore recommended that the All Possible Pair procedure could be a better approach in situations of any multivariate data structure.

Key Words: Discriminant, Function, Classification, combination, Accuracy Rate.

1.0 Introduction

Multivariate statistics concerns the different aims and background of each of the different forms of multivariate analysis and how they relate to each other. Multivariate analysis techniques, besides discriminant analysis also includes principal component analysis (Ekezie, 2013) and Canonical correlation (Onyeagu et al., 2014) When two or more measurements are available, which usually yields much more information than does one about the population being studied, the discriminant function proposed by Fisher (1936) affords a procedure for obtaining the best linear function for discriminating the population under study. A close look at the allocation rule associated with the Fisher's Linear Discriminant (FLD) procedure provides reasons to infer that the FLD procedure is important, easy and simple when applied to just two groups. Fisher (1940) pointed out that although the proposed technique have been applied in widely differing field especially for the two group case, considerable work in theory remains to be done for the more than two group case. As Allwein et al (2000) pointed out, in practice, the choice of reduction method from Multi-Class to Binary (Two-Class) is problem-dependent and not a trivial task since each reduction method has its own limitations. In classification generally, solutions to Multi-class (group) problems have been proposed by many researchers. Examples includes the Super Vector Machines (SVMs) (Vapnik, 1998), One-versus-the-rest method (Bottou, et al., 1994), Pairwise Comparison (Hastie & Tibshirani, 1998), Direct Graph Traversal (Platt, et al., 2000), Error Correcting Output Coding (Dietterich & Bakiri, 1995). Solving multi-group classification problems has been improved by overcoming the limit of conventional statistical methods supported by development of artificial intelligence methods yet a number of studies based on various methods are still ongoing in many academic fields (Kyung, et al., 2004). The multiple group problems, however, has very rarely been addressed and most of the methods proposed for two groups do not generalize and the performance of the methods that can be used with several group is not generally reliable (David, H., 1996). It is fair to say that there is probably no multi-class approach generally outperforms the others. For practical problems, the choice of approach will depend on constraints on hand such as required accuracy, the time available for development and training and the nature of the classification problem and data structure. The simple, efficient and accurate discriminant analysis provides a good choice for practical multi-class classification problems. As multi-group classification problem is not confined to specific studies but it is rather faced by overall studies, verifying its general applicability is important. Efficient multi-group classification model would be one which to a significantly large extent correctly classifies objects into their group thereby producing a minimum error rate or a higher rate of correct classification.

The performance of the classification rule can be evaluated by obtaining the optimal error rate (proportion of objects wrongly classified by the allocation rule) associated with the allocation rule where the allocation rule itself is derived from the suggested model.

Although practical evidences have shown that discriminant analysis is effective, it should be pointed out that a significant separation does not necessarily imply a good classification (Tao, et el., 2006). Owing to the demand for more theoretical work to be done due to lack of a reliable and dependable tool/procedure for multi-group classification in discriminant analysis (Fishers, 1940, David, 1996, Vark et al., 1982, Ramaswamy, et al., 2011, Kyung, et al., 2004, Xu et al., 2009, Daniela and Tibshirani, 2011), methods that exist have limitations ranging from bias, inconsistency, weak statistical based assumptions, cost and time inefficiency.

This work is aimed at developing an efficient procedure for discrimination and classification when we have multiple groups. Compare the result of this work with the conventional method using their accuracy rate as a criterion.

Several methods were reviewed, few of which includes; Linear Discriminant Analysis (Fukunaga, 1990), Two-class linear discriminant analysis (Fishers, 1936), Pairwise Comparison (Kwon et al., 1997), Nearest Neighbour classifier (Sandrine et al, 2000), Aggregate Classifiers (Breiman, 1996, 1998), Boosting (Freud and Schapire, 1997), Multiple Group Logistic Model (Lesaffre and Albert, 1989). However, emphasis was much on Fishers Linear Discriminant procedure since it forms the basis of most of the part of this work.

2.0 Methodology

Assuming we have a set of observation with attributes represented by variables $x_1, x_2, \dots, x_k$ coming from m-population (groups). Group I has n_1 observations, group II has n_2 observations and so on up to group m having n_m observations where $n_1 + n_2 + \cdots + n_m = N$. Our interest is to classify a future (or new) observation whose origin is unknown with same attributes as $x_1, x_2, \dots, x_k$ to the correct group. We desire to do this with so much caution so as to minimize the cost of misclassification. Fisher's procedure obtains a set of m-1 linear functions which represents the functional relationship between the discriminating attributes (or variables).

3.0 The New Procedure

In an attempt, to obtain such linear functions of the discriminating variables, we start by considering the concept of combinatorial analysis. The possible ways of arranging n objects in r ways (considering the order and without repetition) is given by

$$\binom{n}{r} = \frac{n!}{r!\,(n-r)!} \dots \dots \dots \dots (1)$$

Applying (1) into our set up, where n would be replaced by m, the number of groups and r the possible pairs of group combination, thus

$$\binom{m}{r} = \frac{m!}{r!\,(m-r)!} = \lambda \dots \dots \dots \dots (2)$$

λ is the number of functions arising from all possible pairs of combination with m-groups. λ would certainly be a non-negative integer. Given m-groups, evaluation of the number of all possible pairs would result to λ number of functions in the form of a linear functions arising from combining possible pairs of groups without repetition. Clearly, we would have a set Discriminant Functions (DF) representing every possible pairs of group combination thus

$$DF_{1,2}, DF_{1,3}, \dots, DF_{1,m-1}, DF_{1,m} \dots \dots \dots \dots (3)$$

$$DF_{2,1}, DF_{2,3}, \dots, DF_{2,m-1}, DF_{2,m} \dots \dots \dots \dots (4)$$

$$\vdots$$

$$DF_{m,1}, DF_{m,2}, \dots, DF_{m,m-2}, DF_{m,m-1} \dots \dots (5)$$

Equations (3), (4) & (5) are now the discriminant functions arising from all possible pairs of group combination from m-groups(in general).

From Fishers (1936), an allocation rule for classifying future observation is given as: allocate to group I if

$$X > \frac{\overline{X}_1 + \overline{X}_2}{2} \text{ (in univariate case)}$$

Otherwise Allocate to group II.

And:

Allocate to group I if

$$\alpha^T X > D\text{(in multivariate case)}$$

Otherwise allocate to group II

Where

$$Z = a^T X = a_1 x_1 + a_2 x_2 + \dots + a_k x_k \dots \dots \dots \dots (6)$$

$$D = \frac{1}{2}(\overline{X}_1 + \overline{X}_2)^T \mathbf{S}^{-1}(\overline{X}_1 - \overline{X}_2) \dots \dots \dots \dots \dots (7)$$

Equation (7) can readily be obtained as above but with respect to available number of pairs of groups.

It is worthy to note that D can only be computed for two groups at a time. Since our derivations are in pairs, it is also possible to obtain for each possible pair, a corresponding and appropriate D-value. Thus, for m-groups and λ number of discriminant functions, we would have a set of D-values in the form;

$$D_{1,2}, D_{1,3}, \dots, D_{1,m-1}, D_{1,m} \dots \dots \dots \dots \dots (8)$$

$$D_{2,1}, D_{2,3}, \dots, D_{2,m-1}, D_{2,m} \dots \dots \dots \dots (9)$$

$$\vdots$$

$$D_{m,1}, D_{m,2}, \dots, D_{m,m-2}, D_{m,m-1} \dots \dots \dots \dots (10)$$

Equations (8), (9) & (10) are the D-values corresponding to discriminant functions of all possible pairs of group combination.

Table 1: Summary of Discriminating values for all Groups

Group I	Group II	Group III	…	Group M
$D_{1,2}$	$D_{2,1}$	$D_{3,1}$	…	$D_{m,1}$
$D_{1,3}$	$D_{2,3}$	$D_{3,2}$	…	$D_{m,2}$
$D_{1,4}$	$D_{2,4}$	$D_{3,4}$	…	$D_{m,3}$
…	…	…	…	…
$D_{1,m-1}$	$D_{2,m-1}$	$D_{3,m-1}$		$D_{m,m-2}$
$D_{1,m}$	$D_{2,m}$	$D_{3,m}$	…	$D_{m,m-1}$

Since the combinatorial analysis so far has given us λ number of discriminant functions and λ number of D-values. It follows that λ number of rules would be required to conveniently allocate observations. Fishers procedure for allocation provides an only convenient way of combining just a pair (with regards to order) at a time for classification (i.e evaluation can only be done by comparing two groups at a time). Having stated this, the λ number of rules that can allocate future observation derived on the basis of our initial combinatorial concept would be such that each possible pair would have a corresponding allocation rule. That is, for pairs of group 1&2, a rule would exist independently for its classification, for 1&3, another rule would also exist independently for its classification, also for 1&4 up to and including every available possible pairs of functions. Clearly put this way;

For $DF_{1,2}$, a corresponding $D_{1,2}$ would exist,

For $DF_{1,3}$, a corresponding $D_{1,3}$ would exist,

For $DF_{m-1,m}$, a corresponding $D_{m-1,m}$ would exist,

up to

For $DF_{m,m-1}$, a corresponding $D_{m,m-1}$ would exist.

The above evaluations would clearly give rise to a set of λ independent rules for every possible pairs. Thus, we can set up a rule as:

Allocate to

$$G_1 \ if \ DF_{1,2} > D_{1,2} \dots\dots\dots\dots (11)$$

else

$$G_2 \ if \ DF_{1,3} > D_{1,3} \dots\dots\dots\dots (12)$$

else

$$\vdots$$

$$G_{m-1} \ if \ DF_{1,m-1} > D_{1,m-1} \dots\dots\dots\dots (13)$$

Otherwise

$$Allocate \ to \ G_m \dots\dots\dots\dots (14)$$

(14), (15), (16), and (17) are set of rules existing when allocating any observation coming (or assumed) from group I. It therefore follows that, another set of rules would exist for allocating observations coming from group II. Thus we have;

Allocate to

$$G_1 \ if \ DF_{2,1} > D_{2,1} \dots\dots\dots\dots (15)$$

else

$$G_2 \ if \ DF_{2,3} > D_{2,3} \dots\dots\dots\dots (16)$$

else

$$\vdots$$

$$G_{m-1} \ if \ DF_{2,m-1} > D_{2,m-1} \dots\dots\dots\dots (17)$$

Otherwise

$$Allocate \ to \ G_m \dots\dots\dots\dots (18)$$

The above continues in the same way until the mth group such that we would have a set of rules for classifying observations coming (or assumed) from the mth group. Thus we would

Allocate to

$$G_1 \; if \; DF_{m,1} > D_{m,1} \dots\dots\dots\dots (19)$$

else

$$G_2 \; if \; DF_{m,2} > D_{m,2} \dots\dots\dots\dots (20)$$

else

$$\vdots$$

$$G_{m-1} \; if \; DF_{m,m-1} > D_{m,m-1} \dots\dots\dots\dots (21)$$

Otherwise

$$Allocate \; to \; G_m \dots\dots\dots\dots (22)$$

Table 2: Summary of the Allocation Rules

All Original G_1	**All Original G_2**	**All Original G_3**	**…**	**All original G_m**
$G_1 \; if \; DF_{1,2} > D_{1,2}$	$G_1 \; if \; DF_{2,1} > D_{2,1}$	$G_1 \; if \; \mathrm{DF}_{3,1} > D_{3,1}$	…	$G_1 \; if \; DF_{m,m-1} > D_{m,m-1}$
$G_2 \; if \; DF_{1,3} > D_{1,3}$	$G_2 \; if \; DF_{2,3} > D_{2,3}$	$G_2 \; if \; DF_{3,2} > D_{3,2}$	…	$G_2 \; if \; DF_{m,m-1} > D_{m,m-1}$
$G_3 \; if \; DF_{1,4} > D_{1,4}$	$G_3 \; if \; DF_{2,4} > D_{2,4}$	$G_3 \; if \; DF_{3,4} > D_{3,4}$	…	$G_3 \; if \; DF_{m,m-1} > D_{m,m-1}$
…	…	…	…	…
$G_{m-2} \; if \; DF_{1,m-1} > D_{1,m-1}$	$G_{m-2} \; if \; DF_{2,m-1} > D_{2,m-1}$	$G_{m-2} \; if \; DF_{3,m-1} > D_{3,m-1}$		$G_{m-2} \; if \; DF_{m,m-1} > D_{m,m-1}$
$G_{m-1} \; if \; DF_{1,m} > D_{1,m}$	$G_{m-1} \; if \; DF_{1,m} > D_{1,m}$	$G_{m-1} \; if \; DF_{1,m} > D_{1,m}$	…	$G_{m-1} \; if \; DF_{1,m} > D_{1,m}$
Otherwise G_m	Otherwise G_m	Otherwise G_m	…	Otherwise G_m

4.0 Application

The developed procedure can be applied in any field; however, data from the Agricultural field were used to evaluate the performance of the APPS procedure. Data of ten groups , represented by different yam varieties (Adaka, 99/Amo/95a, 99/Amo040, 99/Amo/03, 99/Amo/080, 99/Amo/056, 99/Amo/114, Ame, 99/Amo/060, 99/Amo/064) were collected from National Root Crops Research Institute (NRCRI) Umudike. Each yam has four attributes (discriminating variables) represented as X_1=Weight of Yam tuber, X_2=Weight of standing yam tuber, X_3=Tube Length and X_4=Tube girth. It comprises of N=30 observed yam tuber with 3 of each variety. Data used is presented in Appendix. The basic assumptions of normality and equality of variance among the ten groups still holds.

For the ten groups applied, the number of functions is obtained by combinatorial analysis

$$\binom{m}{r} = \frac{m!}{r!\,(m-r)!} = \binom{10}{2} = \frac{10!}{2!\,(10-2)!} = 45$$

This implies that we would have 45 discriminat functions and there corresponding discriminating values representing all possible pairs of group combinations. These functions were not included here due to space.

Table 3a: The Allocation Rules are given below

All Original G1	All Original G2	All Original G3	All Original G4	All Original G5
$G_1\ if\ DF_{1,2} > D_{1,2}$	$DF_{2,1} > D_{2,1}$	$DF_{3,1} > D_{3,1}$	$DF_{4,1} > D_{4,1}$	$DF_{5,1} > D_{5,1}$
$G_2\ if\ DF_{1,3} > D_{1,3}$	$DF_{2,3} > D_{2,3}$	$DF_{3,2} > D_{3,2}$	$DF_{4,2} > D_{4,2}$	$DF_{5,2} > D_{5,2}$
$G_3\ if\ DF_{1,4} > D_{1,4}$	$DF_{2,4} > D_{2,4}$	$DF_{3,4} > D_{3,4}$	$DF_{4,3} > D_{4,3}$	$DF_{5,3} > D_{5,3}$
$G_4\ if\ DF_{1,5} > D_{1,5}$	$DF_{2,5} > D_{2,5}$	$DF_{3,5} > D_{3,5}$	$DF_{4,5} > D_{4,5}$	$DF_{5,4} > D_{5,4}$
$G_5\ if\ DF_{1,6} > D_{1,6}$	$DF_{2,6} > D_{2,6}$	$DF_{3,6} > D_{3,6}$	$DF_{4,6} > D_{4,6}$	$DF_{5,6} > D_{5,6}$
$G_6\ if\ DF_{1,7} > D_{1,7}$	$DF_{2,7} > D_{2,7}$	$DF_{3,7} > D_{3,7}$	$DF_{4,7} > D_{4,7}$	$DF_{5,7} > D_{5,7}$
$G_7\ if\ DF_{1,8} > D_{1,8}$	$DF_{2,8} > D_{2,8}$	$DF_{3,8} > D_{3,8}$	$DF_{4,8} > D_{4,8}$	$DF_{5,8} > D_{5,8}$
$G_8\ if\ DF_{1,9} > D_{1,9}$	$DF_{2,9} > D_{2,9}$	$DF_{3,9} > D_{3,9}$	$DF_{4,9} > D_{4,9}$	$DF_{5,9} > D_{5,9}$
$G_9\ if\ DF_{1,10} > D_{1,10}$	$DF_{2,10} > D_{2,10}$	$DF_{3,10} > D_{3,10}$	$DF_{4,10} > D_{4,10}$	$DF_{5,10} > D_{5,10}$
Otherwise G_{10}	Otherwise G_{10}	Otherwise G_{10}	Otherwise G_{10}	Otherwise G_{10}

Table 3b: The Allocation Rules are given below Cont'd

All Original G6	All Original G7	All Original G8	All Original G9	All Original G10
$G_1\ if\ DF_{6,1} > D_{6,1}$	$DF_{7,1} > D_{7,1}$	$DF_{8,1} > D_{8,1}$	$DF_{9,1} > D_{9,1}$	$DF_{10,1} > D_{10,1}$
$G_2\ if\ DF_{6,2} > D_{6,2}$	$DF_{7,2} > D_{7,2}$	$DF_{8,2} > D_{8,2}$	$DF_{9,2} > D_{9,2}$	$DF_{10,2} > D_{10,2}$
$G_3\ if\ DF_{6,3} > D_{6,3}$	$DF_{7,3} > D_{7,3}$	$DF_{8,3} > D_{8,3}$	$DF_{9,3} > D_{9,3}$	$DF_{10,3} > D_{10,3}$
$G_4\ if\ DF_{6,4} > D_{6,4}$	$DF_{7,4} > D_{7,4}$	$DF_{8,4} > D_{8,4}$	$DF_{9,4} > D_{9,4}$	$DF_{10,4} > D_{10,4}$
$G_5\ if\ DF_{6,5} > D_{6,5}$	$DF_{7,5} > D_{7,5}$	$DF_{8,5} > D_{8,5}$	$DF_{9,5} > D_{9,5}$	$DF_{10,5} > D_{10,5}$
G_6 if $DF_{6,7} > D_{6,7}$	$DF_{7,6} > D_{7,6}$	$DF_{8,6} > D_{8,6}$	$DF_{9,6} > D_{9,6}$	$DF_{10,6} > D_{10,6}$
G_7 if $DF_{6,8} > D_{6,8}$	$DF_{7,8} > D_{7,8}$	$DF_{8,7} > D_{8,7}$	$DF_{9,7} > D_{9,7}$	$DF_{10,7} > D_{10,7}$
G_8 if $DF_{6,9} > D_{6,9}$	$DF_{7,9} > D_{7,9}$	$DF_{8,9} > D_{8,9}$	$DF_{9,8} > D_{9,8}$	$DF_{10,8} > D_{10,8}$
G_9 if $DF_{6,10} > D_{6,10}$	$DF_{7,10} > D_{7,10}$	$DF_{8,10} > D_{8,10}$	$DF_{9,10} > D_{9,10}$	$DF_{10,9} > D_{10,9}$
Otherwise G_{10}	Otherwise G_{10}	Otherwise G_{10}	Otherwise G_{10}	Otherwise G_{10}

Table 4: The classification Table

	Predicted Group									
	G1	**G2**	**G3**	**G4**	**G5**	**G6**	**G7**	**G8**	**G9**	**G10**
G1	1	0	0	0	1	0	0	0	0	1
G2	0	2	0	0	0	0	1	0	0	0
G3	0	0	3	0	0	0	0	0	0	0
G4	0	0	2	1	0	0	0	0	0	0
G5	0	0	0	0	2	0	0	0	0	0
G6	0	0	0	0	0	3	0	0	0	0
G7	0	0	0	0	0	0	3	0	0	0
G8	0	0	0	0	0	0	0	3	0	0
G9	0	0	0	0	0	0	0	0	3	0
G10	0	0	0	0	0	0	2	0	0	1

$$\text{Accuracy Rate (AR)} = \frac{1+2+3+1+2+3+3+3+3+1}{30} = \frac{22}{30} = 0.733 * 100 = 73.33\%$$

The Accuracy Rate (AR) calculated above from table 8, gave 73.33% correct classification.

Table 5: Result from the Conventional Fishers Method

Classification Results

			Predicted Group Membership										
		Group	1	2	3	4	5	6	7	8	9	10	Total
Original	Count	1	2	0	0	0	0	1	0	0	0	0	3
		2	0	3	0	0	0	0	0	0	0	0	3
		3	1	1	1	0	0	0	0	0	0	0	3
		4	1	0	0	1	0	0	0	0	0	1	3
		5	0	0	0	1	1	0	0	0	1	0	3
		6	0	0	0	0	0	3	0	0	0	0	3
		7	0	0	0	0	1	0	0	0	1	1	3
		8	0	0	0	0	0	0	0	3	0	0	3
		9	0	1	0	1	1	0	0	0	0	0	3
		10	0	0	1	0	0	0	0	0	0	2	3
	%	1	66.7	.0	.0	.0	.0	33.3	.0	.0	.0	.0	100.0
		2	.0	100.0	.0	.0	.0	.0	.0	.0	.0	.0	100.0
		3	33.3	33.3	33.3	.0	.0	.0	.0	.0	.0	.0	100.0
		4	33.3	.0	.0	33.3	.0	.0	.0	.0	.0	33.3	100.0
		5	.0	.0	.0	33.3	33.3	.0	.0	.0	33.3	.0	100.0
		6	.0	.0	.0	.0	.0	100.0	.0	.0	.0	.0	100.0
		7	.0	.0	.0	.0	33.3	.0	.0	.0	33.3	33.3	100.0
		8	.0	.0	.0	.0	.0	.0	.0	100.0	.0	.0	100.0
		9	.0	33.3	.0	33.3	33.3	.0	.0	.0	.0	.0	100.0
		10	.0	.0	33.3	.0	.0	.0	.0	.0	.0	66.7	100.0

a. 53.3% of original grouped cases correctly classified.

Interpretation: Table 9 above gave 53.3% correct classification.

5.0 Findings

The new procedure developed as applied in the classification of ten groups of yam species, the result gave 73.33% correct classification. Furthermore, when it was compared with the result of Fisher's method which gave 53.3% correct classification, it was found to be better. Having carefully considered and implemented the procedure suggested in this work, it could be observed that when available groups are many, it is better to consider and carry out evaluation in pairs. The "pairs" ensures that possible error resulting from combining the many groups simultaneously is avoided. It also ensures that every possible pairs are considered appropriately since statistically accepted allocation rule makes provision for accommodating only two groups at a time.

6.0 Summary

The procedure suggested and presented in this work undoubtedly has shown considerable and better performance when compared with its conventional Fishers procedure. The suggested procedure would be a dependable and alternative tool in situation where we have multiple groups and intends to discriminate and allocate. It would also assuredly overcome the problem of sample size because even with a small sample, its performance was fairly outstanding. The procedure is based on mathematical acceptable concepts and has in no way violated or deviated from known and important statistical principles. The procedure though may look cumbersome but carefully written computer programs would make the procedure more appreciable in terms of speed and accuracy.

7.0 Conclusion

When we have multiple groups, the conventional procedure only provides a method that exists in theory but contradictory in practice. It has been observed and hence suggested that with multiple groups, higher accuracy in discrimination and allocation of observation can be enhanced by adopting the procedure suggested in this work. Application of this procedure is not only limited to the Agricultural settings but to every area where discrimination and allocation is desired.

8.0 References

Allwein, E. L., Schapire, R. E. & Singer, Y. (2000). 'Reducing Multiclass to Binary: A Unifying Approach for Margin Classifiers'. JMLR 1:113–141.

Anderson, J. A. (1972) Separate Sample Logistic Discrimination. Biometrika, 59, 19-35.

Breiman, L. (1996) Bagging Predictors. Machine Learning, 24:123{140, 1996.

Breiman, L. (1998) Arcing classifiers. Annals of Statistics, 26:801{824, 1998.

David, H. (1996). Error-Rate Estimation in Multiple-Group Linear Discriminant Analysis. Technometrics, 4(38), 389-399

Daniela, M. & Tibshirani, R. (2011) Penalized Classification using Fishers' Linear Discriminant. Journal of the American Royal Statistical Society, Series B (Statistical Methodology) Vol. 73, No. 5 (November, 2011) pp. 753-772

Dietterich, T. G. & Bakiri, G. (1995). 'Solving Multiclass Learning Problems via Error- Correcting Output Codes'. Journal of Artificial Intelligence Research 2:263–286.

Ekezie, D. D.(2013). Principal Component Analysis, an Aid to Interpretation of Data. A Case Study of Oil Palm (Elaeis guineensis jacq). Journal of Emerging Trends in Engineering and Applied Sciences, 4(2), 237-241.

Fisher, R. A. (1936). The Use of Multiple Measurements in Taxonomic Problems. Ann. Eugenics, 7 (1936): 179-188

Fisher, R. A. (1940). The Precision of Discriminant Functions. Ann. Eugenics, 10, (1940): 422- 429

Freund, Y. & Schapire, R. E. (1997). A decision-theoretic generalization of on-line learning and an application to boosting. Journal of computer and system sciences, 55:119-139, 1997.

Fukunaga, F. (1990). Introduction to statistical pattern recognition. Academic Press.

Hastie, T. & Tibshirani, R. (1998). 'Classification by Pairwise Coupling'. In M. I. Jordan, M. J. Kearns, & S. A. Solla (eds.), Advances in Neural Information Processing Systems, vol. 10.The MIT Press.

Kwon, Y. S., Han, I. G. & Lee, K. C. (1997). Ordinal Pairwise Partitioning approach to Neural Networks Training in Bond Rating. Intelligent Systems in Accounting Finance and Management, Vol. 6, pp. 23-40, 1997.

Kyung, S. L., Sung, W. C. & Dong, W. L.(2004). A Study on Solving Multi-Group Classification Problems. Proceedings of ICAD2004, The Third International Conference on Axiomatic Design, Seoul-June 21-24, 2004.

Lesaffre, E. & Albert, A. (1989). Multiple-Group Logistic Regression Diagnostics. Journal of the Royal Statistical Society, Series C (Applied Statistics), Vol. 38, No. 3 (1989), pp. 425-440

Onyeagu, S. I., Osuji, G. A. and Onyia, O. M. (2014). Comparison of Canonical Correlation Analysis and the Generalized Canonical Correlation Analysis using the Lognormal and Cauchy Distributed Data. Journal of Mathematical Theory and Modelling, 5(4), 177-185

Ramaswamy, S., Tamayo, P., Rifkin, R., Mukherjee, S., Yeang, C., Angelo, M., Ladd, M., Lander, E. and Golub, T. (2011) Multiclass Cancer diagnosis using Tumor gene expression signature. Proc. Natn. Acad. Sci. USA, 98, 15149-15154

Tao, L., Shenghou, Z. & Mitsunori,O. (2006). Using Discriminant Analysis for Multi-class Classification: An Experimental Investigation. Knowledge and Information Systems, Vol. 10, Issue 4, (November, 2006) pp. 453-472

Platt, J. C., Cristianini, N. & Taylor, S. J. (2000). 'Large Margin DAGs for Multiclass Classification'. In S. Solla, T. Leen, & K. R. Muller (eds.), Advances in Neural Information Processing Systems, vol. 12. MIT Press.

Sandrine, D., Jane, F., & Terence, P. S. (2000). Comparison of Discrimination Methods for the Classification of Tumors Using Gene Expression Data. Technical Report #576, Department of Statistics, University of California, Berkely, June, 2000.

Vark, G. N. & Sman, P. G. M. (1982). New Discrimination and Classification Techniques in Anthropological Practice. Zietschrift fur Morphologie und Anthropologie, Bd. 73, H. 1 (January 1982), pp. 21-36

Vapnik, V. N. (1998). Statistical Learning Theory. Wiley, New York.

Xu, P., Brock, G. & Parrish, R. (2009). Modified Linear Discriminant Analysis approaches for Classification of high-dimensional microarray data. Computnl. Statist. Data Anal.; 53, 1674-1687