JOURNAL OF INFORMATION AND COMMUNICATION TECHNOLOGY
http://e-journal.uum.edu.my/index.php/jict

How to cite this article:
Okwonu, F. Z., Ahad, N. A., Ogini, N. O., Okoloko, I. E., & Wan Husin, W. Z. (2022). Comparative performance evaluation of efficiency for high dimensional classification methods. *Journal of Information and Communication Technology, 21*(3), 437-464. https://doi.org/10.32890/jict2022.21.3.6

Comparative Performance Evaluation of Efficiency for High Dimensional Classification Methods

[1]Friday Zinzendoff Okwonu,
*[2]Nor Aishah Ahad, [3]Nicholas Oluwole Ogini,
[4]Innocent Ejiro Okoloko & [5]Wan Zakiyatussariroh Wan Husin
[1]Department of Mathematics,
Faculty of Science, Delta State University, Nigeria
[2]School of Quantitative Sciences, College of Arts and Sciences,
Universiti Utara Malaysia, Malaysia
[3]Department of Computer Science, Delta State University, Nigeria
[4]Faculty of Computing, Dennis Osadebay University, Nigeria
[5]Faculty of Computer and Mathematical Science,
Universiti Teknologi MARA, Kelantan Branch, Malaysia

okwonufz@delsu.edu.ng,
*aishah@uum.edu.my, ogini@delsu.edu.ng,
okoloko@ieee.org, wanzh@uitm.edu.my
*Corresponding author

Received: 15/3/2022 Revised: 12/5/2022 Accepted: 25/5/2022 Published: 17/7/2022

ABSTRACT

This paper aimed to determine the efficiency of classifiers for high-dimensional classification methods. It also investigated whether an extreme minimum misclassification rate translates into robust efficiency. To ensure an acceptable procedure, a benchmark evaluation threshold (BETH) was proposed as a metric to analyze the comparative performance for high-dimensional classification methods. A simplified performance metric (ω) was derived to show the efficiency of different classification methods. To achieve the objectives, the existing probability of correct classification (PCC) or classification accuracy reported in five different articles was used to generate the BETH value. Then, a comparative analysis was performed between the application of BETH value $(\emptyset)$and the well-established PCC value (∂) ,derived from the confusion matrix. The analysis indicated that the BETH procedure had a minimum misclassification rate, unlike the Optimal method. The results also revealed that as the PCC inclined toward unity value, the misclassification rate between the two methods (BETH and PCC) became extremely irrelevant. The study revealed that the BETH method was invariant to the performance established by the classifiers using the PCC criterion but demonstrated more relevant aspects of robustness and minimum misclassification rate as compared to the PCC method. In addition, the comparative analysis affirmed that the BETH method exhibited more robust efficiency than the Optimal method. The study concluded that a minimum misclassification rate yields robust performance efficiency.

Keywords: Classification, confusion matrix, efficiency, high-dimensional data, robustness.

INTRODUCTION

Recent studies have reported a rapid evolution in high-dimensional data acquisition due to advancements in technology and the outbreak of the Covid-19 pandemic. As a result, new classification techniques have evolved (Vidaurre, 2020; Saadati & Benner, 2014; Ferraty, 2010; Johnstone & Titterington, 2009). The huge acquisition and availability of high-dimensional data have rendered many conventional classification methods impracticable to apply due to dimensionality problems (Casanova et al., 2011). This development has invalidated the concept of $\frac{n}{p} \geq 5$ (Hamilton, 1970), which has metamorphized to $\frac{n}{p} < 5.$ This has given rise to alternative procedures for evaluating $p > n$ classification methods. Some of the shortfalls of the alternative procedures have been discussed (Johnstone & Titterington, 2009) without much emphasis on robustness and efficiency.

The conventional technique to evaluate the performance of a large sample of small dimensional $(n > p)$ classification methods is the optimal probability of correct classification (OPC) as derived in Equation 1 (Blagus & Lusa, 2013; Johnson & Wichern, 1992):

$$OPC = \Phi\left(\frac{mq}{2}\right) \tag{1}$$

where $mq = \sqrt{(\bar{X}_1 - \bar{X}_2)^T S_P^{-1}(\bar{X}_1 - \bar{X}_2)}, \bar{X}_i(i = 1,2)$ is the sample mean and S_P^{-1} denotes the inverse pooled covariance matrix. Thereafter, the corresponding OPC value is obtained by checking the standard normal cumulative probability distribution table $(\Phi(.))$. On the other hand, the confusion matrix technique compares the classification score (w) for each group with the classifier's benchmark (Ω). The classifier's benchmark for several $n > p$ classification problems (Okwonu & Othman, 2013a) is derived from Equation 2:

$$\Omega = \frac{(\bar{X}_1 - \bar{X}_2)^T S_P^{-1}(\bar{X}_1 + \bar{X}_2)}{2} \tag{2}$$

where $\bar{X}_i(i = 1,2)$ is the sample mean and S_P^{-1} denotes the inverse pooled covariance matrix.

The numerical value obtained from the confusion matrix is compared with the OPC value to determine classification accuracy, which is a standard practice in the classification domain (Ghosh et al., 2021; Yan et al., 2021; Kranenburg et al., 2020; Penenberg, 2016; Bickel & Doksum, 2015; Okwonu & Othman, 2013b; Croux et al., 2008, 2011; Kim & Kittler, 2005). A drawback of this method is that it overestimates the misclassification rate, possibly because the testing sample is drawn from different distributions from the training sample. It also underestimates the misclassification rate if the training set is used to validate the model. This is a two-way problem in the classification literature (Okwonu, 2013; Johnson & Wichern, 1992).

The benchmark probability method, which is based on the squared Mahalanobis distance, relies heavily on the sample covariance, which makes it unsuitable for high-dimensional problems due to the curse of singularity or dimensionality (Ghosh, 2012; Bühlmann & Geer, 2011). For $p > n$ classification problems, the classification performance depends strictly on the classifier with the highest probability of correct classification obtained via the confusion matrix. For this type of problem, the effect of data redundancy and correlation on the classifier may not appear noticeable to the analyst (Wang et al., 2018; Guo et al., 2008; Yu & Liu, 2003). Several studies on the application of classification accuracy to determine classifiers' robustness or performance have been criticized. Some researchers have suggested the use of the receiver operating characteristic curve (Lin & Chen, 2013; Provost et al., 1998) to determine the performance of classifiers. Other studies considered the area under the receiver operating characteristic curve to determine the robustness of high-dimensional classification algorithms (Bradley, 1997).

LITERATURE REVIEW

Previous studies have shown that classification accuracy is often applied to determine the robustness of any classifier (Lin & Chen, 2013). This approach has received criticisms from the classification literature such that precision, sensitivity, geometric mean, and -score have been suggested as better alternatives.

Various performance evaluation procedures for high-dimensional unequal class sample size problems have been discussed extensively (Gil-Begue et al., 2021; Bielza et al., 2011). An alternative measure to determine classifiers' performance was formulated by Gibaja (2013). Zhu et al. (2005) also developed a rigid method to compute classification accuracy based on a zero-one subset. The evaluation methods enumerated so far are based on the final output from the various classification models. The macro and micro concepts have been extended to the receiver operating characteristic curve based on the area under the curve to determine robustness. Most of these evaluation methods, such as pairwise techniques (Fürnkranz et al., 2008; Hüllermeier et al., 2008), are not universally applied to other aspects of high-dimensional classification problems (Gil-Begue et al., 2021). Bielza et al. (2011) proposed another evaluation metric called joint and mean accuracy to evaluate the performance of high-dimensional classification methods. The Brier score was also coined as a metric to measure the robustness of classifiers. This concept was popularized to high-dimensional classification by Fernandes et al. (2013). The aforementioned metric has an optimal performance benchmark as one excluding the Brier score (Gil-Begue et al., 2021). Chuang et al. (2008) and Hu et al. (2018) advanced a method that was a reformulation and the equivalent of the confusion matrix approach in which weights were assigned based on selected relevant variables. The problem with the method is that it depended on the number of selected variables and the total number of variables in the data set. This restricts its application to other high-dimensional classification problems that may apply data point elimination, equal class sample sizes, or direct classification analyses. Another weakness of this method is that it would overshoot the misclassification rate when the testing data was derived from different distributions from the training sample. The third weakness is that it underestimated the misclassification rate when the training sample was used to test the model.

Based on the aforementioned weaknesses, the present study proposes a unifying benchmark evaluation threshold (BETH) that depends strictly on the output of any high-dimensional classifier to generate the BETH value, which is used as an optimal value for the classifier. The objectives of this study are: i) to utilize the probability of correct classification (PCC) to develop the optimal BETH; ii) to compare the extreme minimum misclassification rate between the BETH method and the existing method; and iii) to determine the performance efficiency and robustness of the classification methods such as variants of Fisher linear classification methods, random forest, support vector machine, K-nearest neighbor, etc.

The next part of this paper describes the optimal value method for high-dimensional classification methods. This section also presents some of the notable evaluation methods and the BETH method. Data presentation, results, and discussions are reported in the subsequent section, followed by conclusions in the last section.

METHODOLOGY

The data set used in this study was obtained from several classification results reported in five different articles. Based on the axiom of probability and the binomial probability distribution, the sum of the probability of correct classification and misclassification is equal to one as shown in Equation 3:

$$P = p_i + q_i = 1 \tag{3}$$

where p_i denotes the probability of correct classification and q_i is the probability of misclassification associated to each group of the object of investigation. The optimal value method was formulated based on the axiom of probability and the binomial probability distribution. Accordingly, the sum of the probability of the correct classification and misclassification of the classifiers is equal to one. Therefore, $P = p_i + q_i = 1$ was generalized as the optimal probability benchmark value in classification domain. On the other hand, the BETH method derived its probability benchmark value from the classification accuracy of the classifiers. This section also describes several evaluation methods for high-dimensional data sets.

The Optimal Value For $n > p$

This method simply computed the optimal value based on the given data set. This criterion of Equation 4 is described in detail in Johnson and Wichern (1992) as:

$$\delta = \Phi\left(\frac{\Delta}{2}\right) \tag{4}$$

where Φ denotes the standard normal cumulative probability distribution, and Δ is the Mahalanobis distance, as defined in Equation 5:

$$\Delta = \sqrt{(\bar{X}_1 - \bar{X}_2)' S_p^{-1} (\bar{X}_1 - \bar{X}_2)} \tag{5}$$

The sample mean and pooled sample variance are defined as follows:

$$\bar{X}_I = \frac{\sum_{I=1}^{n_k} X_I}{n_k}, \quad k = 1,2 \tag{6}$$

$$S_p = \frac{(n_1-1)S_1 + (n_2-1)S_2}{(n_1+n_2)-2} \tag{7}$$

where $k = 1,2$ and $S_k = \frac{\sum_{j=1}^{n_k}(X_j - \bar{X}_k)^2}{n_k}$. Equation (4) gives the optimal probability of correct classification for $n > p$ problems. This optimal procedure cannot be applied to $p > n$ problems due to the curse of dimensionality resulting from the computation of the Mahalanobis distance. Therefore, several other methods such as performance evaluation function, classification accuracy, precision, sensitivity, macro, micro, and geometric mean have been applied to determine the classifier's performance.

High-Dimensional Evaluation Methods (HDEM)

Performance Evaluation Function (PEF) of HDEM

The performance evaluation function (PEF) is designed such that exact classification results are derived from selected variables for high-dimensional data. The classification accuracy measure of PEF (Hu et al., 2018; Blagus & Lusa, 2010; Chuang et al., 2008) is described in Equations 8–10 as follows:

$$\partial = \widehat{\emptyset}_1 * \nabla + \widehat{\emptyset}_2[1-q], q = \frac{m}{M} < 1, \quad (8)$$

$$\nabla = \frac{\mathrm{C}}{\tau} = \frac{\mathrm{C}}{\mathrm{C}+\vartheta}, \tau = \mathrm{C} + \vartheta \quad (9)$$

$$\varepsilon = \cap - \partial = \cap - \nabla, (\cap = 1) \quad (10)$$

where ∂ is the classification accuracy, $[\widehat{\emptyset}_1, \widehat{\emptyset}_2] = [0.999, 0.001]$ are the assigned weights, M is the total number of variables in the data set, m is the number of selected variables, C is the proportion of correct classification from the different collections of classifiers, $\vartheta = (1 - \mathrm{C})$ is the proportion of misclassification from the different collections of classifiers, and ε is the optimal misclassification error. If $M = m$, the rightmost part of Equation 8 reduces to zero, thus the classification accuracy is determined by Equation 11:

$$\partial = \widehat{\emptyset}_1 * \nabla \quad (11)$$

This is a special weakness of this method, where $\nabla > \partial$ because the weight $\widehat{\emptyset}_1$ would reduce the value of ∇. On the other hand, if $M > m$, then $\nabla = \partial$.

Therefore, PEF cannot be applied as a metric for an equal sample size problem. Otherwise, it would underestimate the classification accuracy. Equation 9 is generally applicable, but may overshoot the misclassification rate because of the benchmark of $\cap = 1$.

Evaluation Procedures for HDEM

In multi-classification problems, the concept of macro and micro-precision, macro and micro recall, and macro and micro-F1 have been advanced to evaluate the performance of high-dimensional unequal class sample sizes (Zhang et al., 2019; Luo & Li, 2014; Lin & Chen, 2013; Michiels et al., 2005). These methods are described with the aid of the following equations, and the confusion matrix shown in Table 1.

Table 1

Confusion Matrix

		Classified as	
		A	B
Actual groups	A	TP	FP
	B	FN	TN

Macro-recall: $MAC-R=\frac{\sum_{J=1}^{K}R_J}{K}$ (12)

Macro-F_1: $MAC-F_1=\frac{MAC-P\times MAC-R\times 2}{MAC-P+MAC-R}$ (13)

Micro-precision: $MIC-P=\frac{\sum_{J=1}^{K}TP_J}{\sum_{J=1}^{K}TP_J+FP_J}$ (14)

Micro-recall: $MIC-R=\frac{\sum_{J=1}^{K}TP_J}{\sum_{J=1}^{K}TP_J+FN_J}$ (15)

Micro-F_1: $MIC-F_1=\frac{MIC-P\times MIC-R\times 2}{MIC-P+MIC-R}$ (16)

Classification accuracy: $CA=\frac{TP+TN}{TP+FP+FN+TN}$ (17)

g-mean=$\sqrt{\text{Micro}-\text{precision}\times\text{Micro}-\text{recall}}$ (18)

The g-mean (Equation 18) simply balances micro-recall and micro-precision (Lin & Chen, 2013). A recent study revealed that the application of macro or micro to determine performance is based on the rule of thumb (Jimoh et al., 2022; Gil-Begue et al., 2021; Gibaja, 2013).

The aforementioned evaluation criteria are useful for unequal sample sizes based on class categorization. The following subsection considers the BETH criterion that is indifferent to class sample size but relies on the classification accuracy to determine the optimal value of classification. The study further applies it to demonstrate the robustness and efficiency of any classifier. This method addresses the shortfall of Equation 8, and it is assumed to give the exact performance benchmark and enhance the robust analysis for high-dimensional classification problems.

Benchmark Evaluation Threshold (BETH)

The BETH method is described as follows: let π be the BETH proportion of correct classification from different classifiers defined in Equation 19:

$$\pi = \left(\frac{\sum_{i=1}^{n_1}\hat{\Delta}_1+\sum_{i=1}^{n_2}\hat{\Delta}_2}{n_1+n_2}\right) \tag{19}$$

where $\hat{\Delta}_i$ are the unit objects correctly assigned in each group. From Equation 19, Equation 20 is derived as:

$$\rho = \left[\frac{(1-\pi)}{2\times\pi}\right] \times \pi \tag{20}$$

Equation 20 describes the optimal misclassification rate (ρ)based on Equation 19, and BETH $(\emptyset)$ is defined in Equation 21:

$$\emptyset = \cap - \rho \tag{21}$$

For different classifiers on $p > n$ data set, the average BETH value (τ), as shown in Equation 22, could be used to analyze the performance. However, this process has some drawbacks such as overfitting and underfitting associated with Equation 22:

$$\tau = \frac{\sum_{i=1}^{k}\emptyset_i}{k} \tag{22}$$

Equation 22 has a unique drawback that may not be recommended for all time performance analyses. Another drawback is that if the numerical values of π fluctuate with a large positive deviation, thus resulting in $\tau > 1$, this value exceeds the optimal performance benchmark $(\delta = 1)$. If the following occurs, $\varphi = \delta - \tau = -\theta$,where $\theta \leq -0.1$, then Equation 22 is not suitable for the analysis, in which case an alternative procedure should be considered. Therefore, it is expected that a perfect model performance is attained when $\varphi = \delta - \tau = 0$.

Based on the BETH method, the following formulas can also be applied to determine the performance and efficiency of any classifier. Equation 23 denotes the deviation between the average BETH value for any method and the probability of exact classification (Equation 8). The misclassification due to BETH (Equations 19, 21) is defined in Equation 24, which indicates the minimum misclassification error as compared to Equation 10. The proportion of misclassification error associated with BETH and the conventional method is defined in Equation 25:

$$\beta = \tau - \partial_i \tag{23}$$

$$\epsilon = \emptyset - \pi \tag{24}$$

$$\ddot{\Delta} = \frac{\epsilon}{\varepsilon} \tag{25}$$

The performance efficiency of BETH and Optimal methods are described in Equations 26–27 as follows:

$$\omega = \frac{\pi}{\emptyset} \times 100\% \tag{26}$$

$$t = \frac{\partial}{\cap} \times 100\% \tag{27}$$

The following summary shows that both methods satisfy the axiomatic definition of probability, which implies that the sum of the probability of correct classification and misclassification is one as shown in Equations 28 and 29, respectively:

$$\gamma = \emptyset + \epsilon = 1 \tag{28}$$

$$\sigma = \partial + \varepsilon = 1 \tag{29}$$

From the above analysis, it is established that the sum of the probability of correct classification and misclassification satisfies Equation 3.

Data Set Description

In this subsection, the different data sets and classifiers are discussed as follows. The results reported in Table 8 are based on the data sets in Tables 2 and 3. The three data sets for this study were culled from Table 4 in Zhang and Cao (2019). Table 2 contains detailed information from the original data source. The data sets were reported on disease diagnosis for high-dimensional biomedical studies.

Table 2

Biomedical Data Sets (Zhang & Cao, 2019)

Data Set Name	Dimension	Sample Size	Class	Reference
A: Arcene	10,000	200	2	Dua and Graff (2019)
CT: Colon Tumor	2,000	62	2	Li and Liu (2004)
NS: Nervous System	7,129	60	2	Li and Liu (2004)

Table 3 contains the algorithm names used in Zhang and Cao (2019) to enhance the classification performance of random forest (RF), support vector machine (SVM), and K-nearest neighbor (KNN). For ease of discussion, B, C, D, E, and F were used to represent the names of the algorithms reported in Table 3 of this paper.

Table 3

Algorithm Names (Zhang & Cao, 2019)

Algorithms	Full Set	FSBRR	Relief	mRmR	GA
New names	B	C	D	E	F

The algorithms in Table 3 were feature selection algorithms applied to the methods to perform classifications. 'Full set' refers to conventional classifiers such as RF, SVM, and KNN. The results in Table 9 were based on the data sets derived from Dua and Graff (2019), and the biomedical data sets as depicted in Table 2 (Zhang et al., 2019), which consisted of six data sets denoted by A to F, as replicated in Table 4.

Table 4

Different Data Sets (Zhang et al., 2019; Dua & Graff, 2019)

Name of Data Set	Sample Size	Dimension
Detect Malicious Executable Anti-Virus Data Set	373	513
SECOM	1,567	591
Heart Disease	267	44
Musk	6,598	168
Urban Land Cover	675	147
Multiple Features	1,608	649

The results in Table 10 consisted of different data sets reported in Table 5, which comprised nine real biomedical data sets described in Table 2 based on Hu et al.'s (2018) study.

Table 5

Data Sets in Table 2 (Hu et al., 2018)

Data name	Sample size	Dimension
Colon Tumur	62	2,000
DLBCL Harvard	58	7,129
Nervous System	60	71,29
Lung Cancer Harvard 1	203	12,600
All-AML-Leukemia	106	7,130
Lung Cancer Ontario	39	2,880
DLBCL Standford	47	4,026
DLBCL NIH	160	7,400
Lung Cancer	181	12,534

The data sets in Table 6 were culled from the report in Table 5 in Lin and Chen (2013) with different classifiers, namely the diagonal linear discriminant analysis (DLDA), RF, and SVM classifiers, and distinct modifications applied to enhance the classical classifiers. The outputs are reported in Table 11.

Table 6

Names of Data Sets Reported in Table 5 in Lin and Chen (2013)

Name of data set	Sample size	Dimension
Lymphoma	58	6,817
Colon Cancer	40	2,000
Breast Cancer	65	7,650
Gene Imprint	88	1,446
Lung Cancer	58	6,817

The data sets reported in Table 7 consisted of Emotions, Yeast, and Scene, as reported in Table 3 in Bielza et al. (2011). Eight different types of algorithms were used to perform classification as reported in Table 4 in Bielza et al. (2011). In this presentation, the algorithm names were abbreviated as TT: Tree Tree, PP: polytree–polytree, PF: pure filter, PW: pure wrapper, HYB: hybrid, WBN: wrapper bn, K2-BN and ML-KMM, respectively. The outputs are reported in Table 13.

Table 7

Data Sets in Table 3 in Bielza et al. (2011)

Name of Data Set	Sample size	Dimension	Reference
Scene	2,407	6	Elisseeff and Beston, (2002)
Yeast	2,417	14	Boutell et al. (2004)
Emotions	593	6	Trohidis et al. (2008)

RESULTS AND DISCUSSION

The results (∂) reported in Table 8 were generated from the classification output for three high-dimensional data sets on three classifiers (RF, SVM, and KNN) and five algorithms as reported in Table 4 (Zhang & Cao, 2019). From Table 8, the BETH method had a general performance remark with the results reported in the original paper. Based on Equation 11, the BETH method had a comparative misclassification rate of 50 percent over the conventional Optimal method. It is observed that the classifier with a minimum misclassification rate had a correspondingly higher efficiency value. The analysis in Figure 1 clearly indicated how the classifiers performed on the data sets.

Table 8

Comparative Misclassification Rate between Optimal and BETH

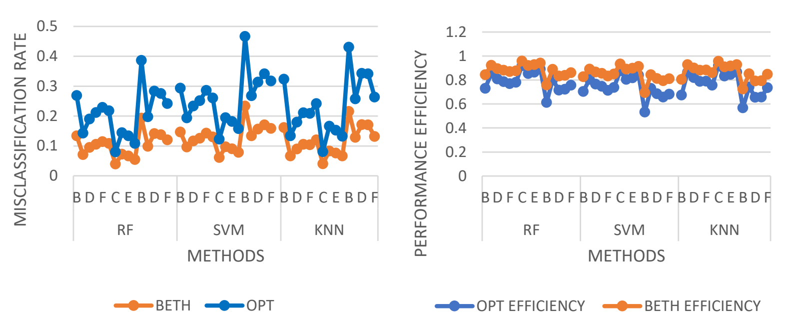
Classifiers	Data set	Algorithms	∂	τ	$\emptyset$	ϵ	β	ε	Δ	ω
Random	A	B	0.73	0.90	0.87	0.13	0.16	0.27	0.5	0.84
Forest		C	0.86		0.93	0.07	0.04	0.14	0.5	0.92
		D	0.81		0.91	0.09	0.09	0.19	0.5	0.90
		E	0.79		0.89	0.11	0.11	0.21	0.5	0.88
		F	0.77		0.89	0.11	0.13	0.23	0.5	0.87
	CT	B	0.78	0.93	0.89	0.11	0.15	0.22	0.5	0.88
		C	0.92		0.96	0.04	0.01	0.08	0.5	0.96
		D	0.86		0.93	0.07	0.08	0.15	0.5	0.92
		E	0.87		0.93	0.07	0.07	0.13	0.5	0.93
		F	0.89		0.95	0.05	0.04	0.11	0.5	0.94
	NS	B	0.61	0.86	0.81	0.19	0.25	0.39	0.5	0.76
		C	0.80		0.90	0.09	0.06	0.19	0.5	0.89
		D	0.72		0.89	0.14	0.15	0.28	0.5	0.83
		E	0.72		0.86	0.14	0.14	0.28	0.5	0.84
		F	0.76		0.88	0.12	0.10	0.24	0.5	0.86
SVM	A	B	0.71	0.87	0.85	0.15	0.17	0.29	0.5	0.83
		C	0.81		0.90	0.09	0.07	0.19	0.5	0.89
		D	0.77		0.88	0.12	0.11	0.23	0.5	0.87
		E	0.75		0.87	0.13	0.13	0.25	0.5	0.86
		F	0.71		0.86	0.14	0.16	0.29	0.5	0.83
	CT	B	0.74	0.91	0.87	0.13	0.17	0.26	0.5	0.85
		C	0.88		0.94	0.06	0.03	0.12	0.5	0.93
		D	0.81		0.90	0.09	0.10	0.19	0.5	0.89
		E	0.82		0.91	0.09	0.09	0.18	0.5	0.89
		F	0.84		0.92	0.08	0.07	0.16	0.5	0.91
	NS	B	0.53	0.83	0.77	0.23	0.29	0.47	0.5	0.70
		C	0.73		0.87	0.13	0.09	0.27	0.5	0.85
		D	0.69		0.84	0.16	0.14	0.31	0.5	0.81
		E	0.66		0.83	0.17	0.17	0.34	0.5	0.79
		F	0.68		0.84	0.16	0.15	0.32	0.5	0.81

(continued)

Classifiers	Data set	Algorithms	∂	τ	∅	ϵ	β	ε	Δ	ω
KNN	A	B	0.68	0.89	0.84	0.16	0.22	0.32	0.5	0.81
		C	0.87		0.93	0.07	0.03	0.13	0.5	0.93
		D	0.82		0.91	0.09	0.07	0.18	0.5	0.90
		E	0.79		0.89	0.11	0.11	0.21	0.5	0.88
		F	0.79		0.89	0.10	0.10	0.21	0.5	0.88
	CT	B	0.76	0.89	0.88	0.12	0.14	0.24	0.5	0.86
		C	0.92		0.96	0.04	-0.02	0.08	0.5	0.96
		D	0.83		0.92	0.08	0.07	0.17	0.5	0.91
		E	0.85		0.92	0.08	0.05	0.15	0.5	0.92
		F	0.87		0.93	0.07	0.03	0.13	0.5	0.93
	NS	B	0.57	0.84	0.78	0.22	0.27	0.43	0.5	0.73
		C	0.74		0.87	0.13	0.09	0.26	0.5	0.85
		D	0.66		0.83	0.17	0.18	0.34	0.5	0.79
		E	0.66		0.83	0.17	0.18	0.34	0.5	0.79
		F	0.74		0.87	0.13	0.10	0.26	0.5	0.85

Figure 1

Comparative Misclassification Rate and Performance Efficiency



a) Comparative misclassification rate b) Comparative performance efficiency

The classifiers in Table 9 consisted of different algorithms applied to robustify the performance of SVM. Therefore, Table 9 contained variants of algorithm such as synthetic minority oversampling, feature selection, and data over sampling techniques. The result reported in Table 9 was based on the data sets reported in Table 4 and the output of the classifiers in Table 5 in Zhang et al. (2019) for four different classification methods denoted by the upper-case letters of each word. This study replicated the results to generate the BETH value in determining the best classifier with the minimum misclassification rate and performance efficiency.

Table 9

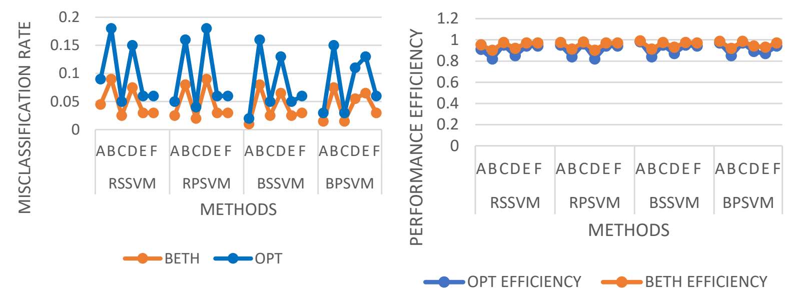
Comparative Misclassification Rate between Optimal and BETH

Classifier	Data set	∂	τ	$\emptyset$	ϵ	β	ε	Δ	ω
RSSVM	A	0.91	0.95	0.96	0.05	0.04	0.09	0.5	0.95
	B	0.82		0.91	0.09	0.13	0.18	0.5	0.91
	C	**0.95**		**0.98**	**0.03**	**0.01**	**0.05**	**0.5**	**0.97**
	D	0.85		0.93	0.08	0.10	0.15	0.5	0.92
	E	0.94		0.97	0.03	0.01	0.06	0.5	0.97
	F	0.94		0.97	0.03	0.01	0.06	0.5	0.97
RPSVM	A	**0.95**	**0.95**	**0.98**	**0.03**	**0.01**	**0.05**	**0.5**	**0.97**
	B	0.84		0.92	0.08	0.11	0.16	0.5	0.91
	C	**0.96**		**0.98**	**0.02**	**-0.01**	**0.04**	**0.5**	**0.98**
	D	0.82		0.91	0.09	0.13	0.18	0.5	0.90
	E	0.94		0.97	0.03	0.01	0.06	0.5	0.97
	F	0.94		0.97	0.03	0.01	0.06	0.5	0.97
BSSVM	**A**	**0.98**	**0.96**	**0.99**	**0.01**	**-0.02**	**0.02**	**0.5**	**0.99**
	B	0.84		0.92	0.08	0.12	0.16	0.5	0.91
	C	**0.95**		**0.98**	**0.03**	**0.01**	**0.05**	**0.5**	**0.97**
	D	0.87		0.94	0.07	0.09	0.13	0.5	0.93
	E	**0.95**		**0.98**	**0.03**	**0.01**	**0.05**	**0.5**	**0.97**
	F	0.94		0.97	0.03	0.02	0.06	0.5	0.97
BPSVM	**A**	**0.97**	**0.96**	**0.99**	**0.02**	**-0.01**	**0.03**	**0.5**	**0.98**
	B	0.85		0.93	0.08	0.11	0.15	0.5	0.92
	C	**0.97**		**0.99**	**0.02**	**-0.01**	**0.03**	**0.5**	**0.98**
	D	0.89		0.95	0.06	0.07	0.11	0.5	0.94
	E	0.87		0.94	0.07	0.09	0.13	0.5	0.93
	F	0.94		0.97	0.03	0.02	0.06	0.5	0.97

From Figure 2, it is observed that the comparative misclassification rate and the comparative performance efficiency were the direct opposite of each other. This indicated that a minimum misclassification rate gave better performance efficiency.

Figure 2

Comparative Misclassification Rate and Performance Efficiency



a) Comparative misclassification rate b) Comparative performance efficiency

The result in Table 10 was based on the nine real biomedical data set description in Table 5 and the classification output reported in Table 2 in Hu et al. (2018). The results reported in Table 2 in Hu et al. (2018) were centered on the KNN(1-NN) classifier. The results indicated that the BETH method had a minimum misclassification rate as compared to the Optimal method. The analysis in Figure 3 revealed that the BETH method reduced the misclassification rate by 50 percent when compared to the unique Optimal method.

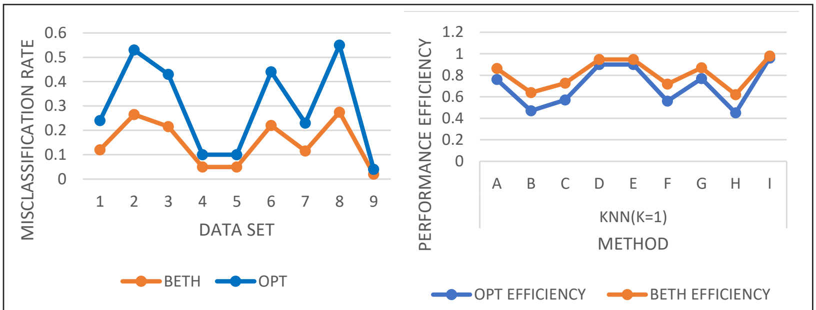
Table 10

Comparative Misclassification Rate between Optimal and BETH

Classifier	Data set	∂	τ	$\emptyset$	ϵ	β	ε	Δ	ω
KNN(K=1)	1	0.76	0.85	0.88	0.12	0.09	0.24	0.5	0.86
	2	0.47		0.74	0.27	0.38	0.53	0.5	0.64
	3	0.57		0.79	0.22	0.28	0.43	0.5	0.73
	4	0.9		0.95	0.05	-0.05	0.10	0.5	0.95
	5	0.9		0.95	0.05	-0.05	0.10	0.5	0.95
	6	0.56		0.78	0.22	0.29	0.44	0.5	0.72
	7	0.77		0.89	0.12	0.08	0.23	0.5	0.87
	8	0.45		0.73	0.28	0.40	0.55	0.5	0.62
	9	**0.96**		**0.98**	**0.02**	**-0.11**	**0.04**	**0.5**	**0.98**

Figure 3

Comparative Misclassification Rate and Performance Efficiency



a) Comparative misclassification rate b) Comparative performance efficiency

The data sets used to generate the output in Table 11 were obtained from Table 5 in Lin and Chen (2013) and replicated as in Table 6. From the output, it is concluded that the results for the different classifiers measured by the BETH performance metric were very robust with an extreme minimum misclassification rate as compared to the rate from the Optimal method.

Table 11

Comparative Misclassification Rate between Optimal and BETH

Classifier	Data set	∂	τ	$\emptyset$	ϵ	β	ε	Δ	ω
SVM	A	0.84	0.946	0.92	0.08	0.11	0.16	0.5	0.91
	B	0.85		0.93	0.08	0.09	0.15	0.5	0.92
	C	0.83		0.92	0.09	0.12	0.17	0.5	0.91
	D	0.79		0.89	0.11	0.16	0.21	0.5	0.88
	E	0.86		0.93	0.07	0.09	0.14	0.5	0.92
	F	0.85		0.93	0.08	0.09	0.15	0.5	0.92
	G	**0.99**		**0.99**	**0.01**	**-0.04**	**0.01**	**0.5**	**0.99**
	H	**0.98**		**0.99**	**0.01**	**-0.03**	**0.02**	**0.5**	**0.99**
	I	**0.98**		**0.99**	**0.01**	**-0.03**	**0.02**	**0.5**	**0.99**
	J	0.95		0.98	0.03	-0.01	0.05	0.5	0.97
DLDA	A	0.67	0.9245	0.84	0.17	0.25	0.33	0.5	0.80
	B	0.85		0.93	0.08	0.07	0.15	0.5	0.92

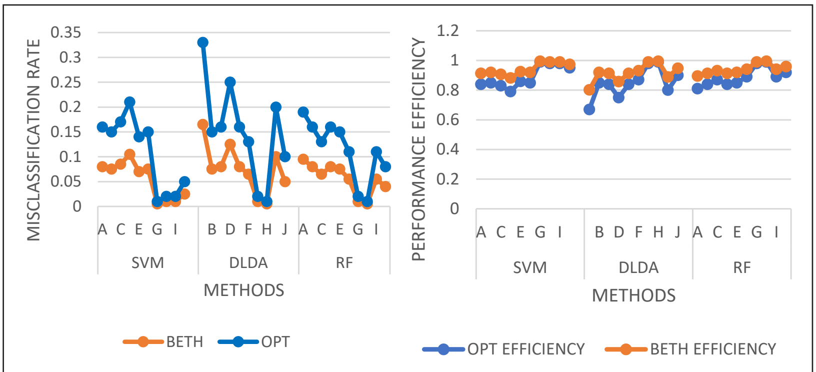
(continued)

Classifier	Data set	∂	τ	∅	ϵ	β	ε	Δ	ω
	C	0.84		0.92	0.08	0.08	0.16	0.5	0.91
	D	0.75		0.88	0.13	0.17	0.25	0.5	0.86
	E	0.84		0.92	0.08	0.08	0.16	0.5	0.91
	F	0.87		0.94	0.07	0.05	0.13	0.5	0.93
	G	**0.98**		**0.99**	**0.01**	**-0.06**	**0.02**	**0.5**	**0.99**
	H	**0.99**		**0.99**	**0.01**	**-0.07**	**0.01**	**0.5**	**0.99**
	I	0.8		0.90	0.10	0.12	0.2	0.5	0.89
	J	0.9		0.95	0.05	0.02	0.1	0.5	0.95
RF	A	0.81	0.944	0.91	0.09	0.13	0.19	0.5	0.90
	B	0.84		0.92	0.08	0.10	0.16	0.5	0.91
	C	0.87		0.94	0.07	0.07	0.13	0.5	0.93
	D	0.84		0.92	0.08	0.10	0.16	0.5	0.91
	E	0.85		0.93	0.08	0.09	0.15	0.5	0.92
	F	0.89		0.95	0.06	0.05	0.11	0.5	0.94
	G	**0.98**		**0.99**	**0.01**	**-0.04**	**0.02**	**0.5**	**0.99**
	H	**0.99**		**0.99**	**0.01**	**-0.05**	**0.01**	**0.5**	**0.99**
	I	0.89		0.95	0.06	0.05	0.11	0.5	0.94
	J	0.92		0.96	0.04	0.02	0.08	0.5	0.96

The analysis in Figure 4 affirmed the robust performance of the BETH method as a suitable evaluation metric for high-dimensional classifiers.

Figure 4

Comparative Misclassification Rate and Performance Efficiency



a) Comparative misclassification rate b) Comparative performance efficiency

Table 12 and Figure 5 include the comparative performance analysis of the evaluation criteria in terms of classification accuracy and g-mean as reported in Table 6 (Lin & Chen, 2013). From Table 12, as the classification output leaned toward unity, the misclassification rate between the two methods became immaterial. Similar results were shown in Tables 9 to 12 marked in bold. This demonstrated that the BETH method was suitable to determine the robust performance of any classifier. From the analysis in Figure 5, it is observed that the BETH performance was superior to the other two methods based on the data sets investigated.

Table 12

Comparative Performance Analysis

Accuracy	g-mean	BETH performance
0.81	0.78	0.90
0.84	0.82	0.91
0.87	0.79	0.93
0.84	0.79	0.91
0.85	0.81	0.92
0.89	0.87	0.94
0.98	0.91	**0.99**
0.99	0.94	**0.99**
0.89	0.78	0.94
0.92	0.86	0.96

Figure 5

Comparative Performance Analysis

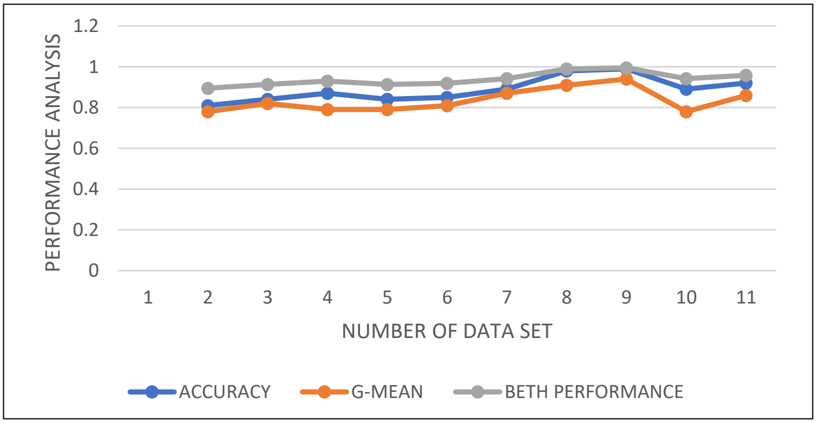


Table 13 contains the output from the mean performance measures on Scene, Yeast, and Emotions data sets with eight classifiers reported in Table 7 (Table 3 in Bielza et al., 2011). The performance efficiency in Table 13 shows that the best classifier had 90 percent and the lowest was 75 percent based on the BETH method, whereas the Optimal method achieved 83 percent and 61 percent, respectively. In Figure 6, the performance demonstrated that the BETH method had an extreme minimum misclassification rate as compared to the Optimal method.

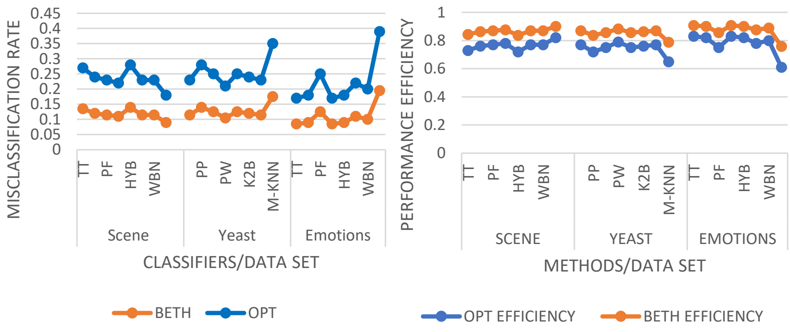
Table 13

Comparative Misclassification Rate between Optimal and BETH

Classifier	Data Set	∂	τ	$\emptyset$	ϵ	β	ε	Δ	ω
TT	Scene	0.73	0.88	0.87	0.13	0.15	0.27	0.5	0.84
PP		0.76		0.88	0.12	0.12	0.24	0.5	0.86
PF		0.77		0.89	0.12	0.11	0.23	0.5	0.87
PW		0.78		0.89	0.11	0.10	0.22	0.5	0.88
HYB		0.72		0.86	0.14	0.16	0.28	0.5	0.84
K2B		0.77		0.89	0.12	0.11	0.23	0.5	0.87
WBN		0.77		0.89	0.12	0.11	0.23	0.5	0.87
M-KNN		0.82		0.91	0.09	0.06	0.18	0.5	0.90
TT	Yeast	0.77	0.87	0.89	0.12	0.10	0.23	0.5	0.87
PP		0.72		0.86	0.14	0.15	0.28	0.5	0.84
PF		0.75		0.88	0.13	0.12	0.25	0.5	0.86
PW		0.79		0.89	0.12	0.08	0.21	0.5	0.88
HYB		0.75		0.88	0.13	0.12	0.25	0.5	0.86
K2B		0.76		0.88	0.12	0.11	0.24	0.5	0.86
WBN		0.77		0.89	0.13	0.10	0.23	0.5	0.87
M-KNN		0.65		0.83	0.18	0.22	0.35	0.5	0.79
TT	Emotions	0.83	0.89	0.92	0.09	0.06	0.17	0.5	0.91
PP		0.82		0.91	0.09	0.07	0.18	0.5	0.90
PF		0.75		0.88	0.13	0.14	0.25	0.5	0.86
PW		0.83		0.92	0.09	0.06	0.17	0.5	0.91
HYB		0.82		0.91	0.09	0.07	0.18	0.5	0.90
K2B		0.78		0.89	0.11	0.11	0.22	0.5	0.88
WBN		0.8		0.90	0.10	0.09	0.20	0.5	0.89
M-KNN		0.61		0.81	0.20	0.28	0.39	0.5	0.76

Figure 6

Comparative Misclassification Rate and Performance Analysis



a) Comparative misclassification rate b) Comparative performance efficiency

From the analysis above, it is observed that both methods were consistent with the general performance of each classifier. Nevertheless, the BETH performance metric was more robust with an extreme minimum misclassification rate as compared to the Optimal method. The outcome of this study demonstrated that if the probability of correct classification from the classifiers was 100 percent, which implied zero misclassification, then the misclassification rate from the BETH method would be 0.00 percent, which corresponded to 100 percent performance efficiency. The study showed that as the Optimal method's value inclined toward unity, the performance efficiency of both methods seemed to be similar. It is revealed that a robust alternative threshold method has been advanced to achieve the best classification efficiency.

CONCLUSION

This study has shown that apart from the confusion matrix, classification accuracy, precision, sensitivity, g-mean, and F_1-score, other robust performance metrics, such as the BETH method, could be applied to determine the performance of different classifiers for high-dimensional data sets for equal and unequal class sample sizes. In many instances, the performance of different algorithms varies based on the group sample sizes. As such, the group varying sample sizes play a similar role in classification performance analyses like robust methods and classical methods. In this case, large sample sizes tend to have larger sample means as compared to groups with small sample sizes, which may also likely produce small sample means. To solve these problems of unequal sample sizes, various procedures have been proposed to select data sets in large groups to correspond with the number of sample sizes selected in smaller groups. These variants of algorithm help to improve the classification performance for high-dimensional data sets. The analysis revealed that unlike the Optimal method, the proposed method had an extreme minimum misclassification rate. This study indicated that as the classifiers' results from the Optimal method tended to achieve 100 percent classification accuracy, in which the misclassification rate between the two methods became extremely irrelevant (zero misclassification). The analysis affirmed that as the misclassification rate reduces, the performance efficiency increases. From the aforesaid analysis, it is concluded that both methods could be applied to analyze performance efficiency effectively.

ACKNOWLEDGMENT

The authors acknowledge Universiti Utara Malaysia and Delta State University for their support for this research. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

REFERENCES

Bickel, P. J., & Doksum, K. A. (2015). Mathematical statistics: Basic ideas and selected topics, second edition. In *Mathematical Statistics: Basic Ideas and Selected Topics, Second Edition* (Vol. 1). https://doi.org/10.1201/b18312

Bielza, C., Li, G., & Larrañaga, P. (2011). Multi-dimensional classification with Bayesian networks. *International Journal of Approximate Reasoning*, *52*(6), 705–727. https://doi.org/10.1016/j.ijar.2011.01.007

Blagus, R., & Lusa, L. (2010). Class prediction for high-dimensional class-imbalanced data. *BMC Bioinformatics*, *11*(1), 1–17. https://doi.org/10.1186/1471-2105-11-523

Blagus, R., & Lusa, L. (2013). SMOTE for high-dimensional class-imbalanced data. *BMC Bioinformatics*, *14*(1), 1–13. https://doi.org/10.1186/1471-2105-14-106

Boutell, M. R., Luo, J., Shen, X., & Brown, C. M. (2004). Learning multi-label scene classification. *Pattern Recognition*, *37*(9), 1757–1771. doi:10.1016/j.patcog.2004.03.009

Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition*, *30*(7), 1145-1159. https://doi.org/10.1016/S0031-3203(96)00142-2

Bühlmann, P., & Geer, S. van de. (2011). Statistics for high-dimensional data: Methods, theory, and applications. In *Springer Series in Statistics*.

Casanova, R., Whitlow, C. T., Wagner, B., Williamson, J., Shumaker, S. A., Maldjian, J. A., & Espeland, M. A. (2011). High dimensional classification of structural MRI Alzheimer's disease data based on large-scale regularization. *Frontiers in Neuroinformatics*, *5*, 22. https://doi.org/10.3389/fninf.2011.00022

Chuang, L. Y., Chang, H. W., Tu, C. J., & Yang, C. H. (2008). Improved binary PSO for feature selection using gene expression data. *Computational Biology and Chemistry*, *32*(1), 29–38. https://doi.org/10.1016/j.compbiolchem.2007.09.005

Croux, C., Filzmoser, P., & Joossens, K. (2008). Classification efficiencies for robust linear discriminant analysis. *Statistica Sinica*, *18*(2), 581–599. http://www.jstor.org/stable/24308496

Dua, D., & Graff, C. (2019). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.

Elisseeff, A., & Weston, J. (2002). A kernel method for multi-labelled classification. In T. Dietterich, S. Becker, Z. Ghahramani (Eds.), *Advances in Neural Information Processing Systems*, vol. 14, MIT Press, pp. 681–687. ttps://proceedings.neurips.cc/paper/2001/file/39dcaf7a053dc372fbc391d4e6b5d693-Paper.pdf

Fernandes, J. A., Lozano, J. A., Inza, I., Irigoien, X., Pérez, A., & Rodríguez, J. D. (2013). Supervised pre-processing approaches in multiple class variables classification for fish recruitment forecasting. *Environmental Modelling and Software*, *40*, 245–254. https://doi.org/10.1016/j.envsoft.2012.10.001

Ferraty, F. (2010). High-dimensional data: A fascinating statistical challenge. *Journal of Multivariate Analysis, 101*(2), 305–306. https://doi.org/10.1016/j.jmva.2009.10.012

Fürnkranz, J., Hüllermeier, E., Loza Mencía, E., & Brinker, K. (2008). Multilabel classification via calibrated label ranking. *Machine Learning*, *73*(2). 133–153. https://doi.org/10.1007/s10994-008-5064-8

Ghosh, A., SahaRay, R., Chakrabarty, S., & Bhadra, S. (2021). Robust generalized quadratic discriminant analysis. *Pattern Recognition*, *117*, 107981. https://doi.org/10.1016/j.patcog.2021.107981

Ghosh, J. K. (2012). Statistics for high-dimensional data: Methods, theory, and applications by Peter Bühlmann, Sara van de Geer. *International Statistical Review*, *80*(3), 486-487. https://doi.org/10.1111/j.1751-5823.2012.00196_18.x

Gibaja, E. (2013). A tutorial on multi-label learning. *ACM Computing Surveys, 9*, 1–38. https://doi.org/10.1145/2716262

Gil-Begue, S., Bielza, C., & Larrañaga, P. (2021). Multi-dimensional Bayesian network classifiers: A survey. *Artificial Intelligence Review*, *54*(1), 519–559. https://doi.org/10.1007/s10462-020-09858-x

Guo, B., Damper, R. I., Gunn, S. R., & Nelson, J. D. B. (2008). A fast separability-based feature-selection method for high-dimensional remotely sensed image classification. *Pattern Recognition*, *41*(5), 1653–1662. https://doi.org/10.1016/j.patcog.2007.11.007

Hamilton, W. C. (1970). The revolution in crystallography. *Science*, *169*(3941), 133–141. https://doi.org/10.1126/science.169.3941.133

Hu, B., Dai, Y., Su, Y., Moore, P., Zhang, X., Mao, C., Chen, J., & Xu, L. (2018). Feature selection for optimized high-dimensional biomedical data using an improved shuffled frog leaping algorithm. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *15*(6), 1765–1773. https://doi.org/10.1109/TCBB.2016.2602263

Hüllermeier, E., Fürnkranz, J., Cheng, W., & Brinker, K. (2008). Label ranking by learning pairwise preferences. *Artificial Intelligence*, *172*(16–17), 1897–1916. https://doi.org/10.1016/j.artint.2008.08.002

Jimoh, R. G., Abisoye, O. A., & Uthman, M. M. B. (2022). Ensemble feed-forward neural network and support vector machine for prediction of multiclass malaria infection. *Journal of Information and Communication Technology*, *21*(1), 117–148. https://doi.org/10.32890/jict2022.21.1.6.

Johnson, R. A., & Wichern, D. W. (1992). *Applied multivariate statistical analysis* (3rd ed.). Prentice-Hall, Inc, Englewood Cliffs.

Johnstone, I. M., & Titterington, D. M. (2009). Statistical challenges of high-dimensional data. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *367*(1906), 4237–4253. https://doi.org/10.1098/rsta.2009.0159

Kim, T. K., & Kittler, J. (2005). Locally linear discriminant analysis for multimodally distributed classes for face recognition with a single model image. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *27*(3), 318–327. https://doi.org/10.1109/TPAMI.2005.58

Kranenburg, R. F., Peroni, D., Affourtit, S., Westerhuis, J. A., Smilde, A. K., & van Asten, A. C. (2020). Revealing hidden information in GC–MS spectra from isomeric drugs: Chemometrics-based identification from 15 eV and 70 eV EI mass spectra. *Forensic Chemistry*, *18*, 100225. https://doi.org/10.1016/j.forc.2020.100225

Li, J., & Liu, H. (2004). Kent Ridge Biomedical Data Set Repository. School of Computer Engineering, Nanyang Technological University, Singapore. http://datam.i2r.astar.edu.sg/datasets/krbd/ index.html.

Lin, W. J., & Chen, J. J. (2013). Class-imbalanced classifiers for high-dimensional data. *Briefings in Bioinformatics, 14*(1), 13–26. https://doi.org/10.1093/bib/bbs006

Luo, L., & Li, L. (2014). Defining and evaluating classification algorithms for high-dimensional data based on latent topics. *PLoS ONE*, *9*(1), e82119. https://doi.org/10.1371/journal.pone.0082119

Michiels, S., Koscielny, S., & Hill, C. (2005). Prediction of cancer outcome with microarrays: A multiple random validation strategy. *Lancet*, *365*(9458), 488–492. https://doi.org/10.1016/S0140-6736(05)17866-0

Okwonu, F. Z. (2013). *Several robust techniques in two-groups unbiased linear classification* (Unpublished Doctoral Dissertation). Universiti Sains Malaysia, Malaysia.

Okwonu, F. Z., & Othman, A. R. (2013a). Heteroscedastic variance-covariance matrices for unbiased two groups linear classification methods. *Applied Mathematical Sciences*, *7*(138), 6855–6865. https://doi.org/10.12988/ams.2013.39486

Okwonu, F. Z., & Othman, A. R. (2013b). Robust fisher linear classification technique for two groups. *World Applied Sciences Journal*, *21*(Special Issue 1). https://doi.org/10.5829/idosi.

wasj.2013.21.mae.99939

Penenberg, D. N. (2016). Mathematical statistics: Basic ideas and selected topics (2nd ed., Vols. I and II). P. J. Bickel and K. A. Doksum, 2015 Boca Raton, Chapman and Hall-CRC xxii + 548 pp., $99.95 (vol. I); 438 pp., $99.95 (vol. II) ISBN 978-1-498-72380-0. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, *179*(4), 1128–1129. https://doi.org/10.1111/rssa.12217

Provost, F., Fawcett, T., & Kohavi, R. (1998, July). The case against accuracy estimation for comparing induction algorithms. In *International Conference on Machine Learning (ICML)* (Vol. 98, pp. 445–453).

Saadati, M., & Benner, A. (2014). Statistical challenges of high-dimensional methylation data. *Statistics in Medicine*, *33*(30), 5347–5357. https://doi.org/10.1002/sim.6251

Trohidis, K., Tsoumakas, G., Kalliris, G., & Vlahavas, I. P. (2008, September). Multi-label classification of music into emotions. In *Proceedings of the Ninth International Society for Music Information Retrieval Conference (ISMIR)* (Vol. 8, pp. 325–330).

Vidaurre, D. (2020). The statistical challenge of finding spontaneous changes in functional connectivity in high-dimensional fMRI data. In *bioRxiv*. https://doi.org/10.1101/2020.12.15.422845

Wang, X. Z., Wang, R., & Xu, C. (2018). Discovering the relationship between generalization and uncertainty by incorporating complexity of classification. *IEEE Transactions on Cybernetics*, *48*(2), 703–715. https://doi.org/10.1109/TCYB.2017.2653223

Yan, C., Chang, X., Luo, M., Zheng, Q., Zhang, X., Li, Z., & Nie, F. (2021). Self-weighted robust LDA for multiclass classification with edge classes. *ACM Transactions on Intelligent Systems and Technology*, *12*(1), 1–19. https://doi.org/10.1145/3418284

Yu, L., & Liu, H. (2003, August). Efficiently handling feature redundancy in high-dimensional data. In *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 685–690). https://doi.org/10.1145/956750.956840

Zhang, B., & Cao, P. (2019). Classification of high-dimensional biomedical data based on feature selection using redundant removal. *PLoS ONE*, *14*(4), e0214406. https://doi.org/10.1371/journal.pone.0214406

Zhang, C., Zhou, Y., Guo, J., Wang, G., & Wang, X. (2019). Research on classification method of high-dimensional class-imbalanced datasets based on SVM. *International Journal of Machine*

Learning and Cybernetics, *10*(7), 1765–1778. https://doi.org/10.1007/s13042-018-0853-2

Zhu, S., Ji, X., Xu, W., & Gong, Y. (2005, August). Multi-labeled classification using the maximum entropy method. In *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 274–281). https://doi.org/10.1145/1076034.1076082