Pertanika J. Sci. & Technol. 30 (4): 2831 - 2850 (2022)

SCIENCE & TECHNOLOGY

Journal homepage: http://www.pertanika.upm.edu.my/

Robust Hybrid Classification Methods and Applications

Friday Zinzendoff Okwonu[1,2], Nor Aishah Ahad[1]*, Innocent Ejiro Okoloko[3], Joshua Sarduana Apanapudor[4], Saadi Ahmad Kamaruddin[1,2,5] and Festus Irimisose Arunaye[4]

[1]*School of Quantitative Sciences, College of Arts and Sciences, Universiti Utara Malaysia, 06010 Sintok, Kedah, Malaysia*
[2]*Institute of Strategic Industrial Decision Modeling, School of Quantitative Sciences, College of Arts and Sciences, Universiti Utara Malaysia, 06010 Sintok, Kedah, Malaysia*
[3]*Faculty of Computing, Dennis Osadebay University, Anwai Rd., Asaba, Nigeria*
[4]*Department of Mathematics, Faculty of Science, Delta State University, P.M.B.1, Abraka, Nigeria*
[5]*Centre for Testing, Measurement, and Appraisal (CeTMA), Universiti Utara Malaysia, 06010 Sintok, Kedah, Malaysia*

ABSTRACT

The sample mean classifier, such as the nearest mean classifier (NMC) and the Bayes classifier, is not robust due to the influence of outliers. Enhancing the robust performance of these methods may result in vital information loss due to weighting or data deletion. The focus of this study is to develop robust hybrid univariate classifiers that do not rely on data weighting or deletion. The following data transformation methods, such as the least square approach (LSA) and linear prediction approach (LPA), are applied to estimate the parameters of interest to achieve the objectives of this study. The LSA and LPA estimates are applied to develop two groups of univariate classifiers. We further applied the predicted estimates from the LSA and LPA methods to develop four hybrid classifiers. These classifiers are applied to investigate whether cattle horn and base width length could be used to determine cattle gender. We also used these classification methods to determine whether shapes could classify banana variety. The NMC, LSA, LPA, and hybrid classifiers showed that cattle gender could be determined using horn length and base width measurement. The analysis further revealed that shapes could determine banana variety. The comparative results using the two data

ARTICLE INFO

Article history:
Received: 22 November 2021
Accepted: 31 January 2022
Published: 23 September 2022

DOI: https://doi.org/10.47836/pjst.30.4.29

E-mail addresess:
o.friday.zinzendoff@uum.edu.my (Friday Zinzendoff Okwonu)
aishah@uum.edu.my (Nor Aishah Ahad)
Okolokoinno@yahoo.com (Innocent Ejiro Okoloko)
japanpudor@yahoo.com (Joshua Sarduana Apanapudor)
s.ahmad.kamaruddin@uum.edu.my (Saadi Ahmad Kamaruddin)
ifarunaye@yahoo.com (Festus Arunaye Irimisose)
*Corresponding author

ISSN: 0128-7680
e-ISSN: 2231-8526

sets demonstrated that all the methods have over 90% performance prediction accuracy. The findings affirmed that the performance of the NMC, LSA, LPA, and the hybrid classifiers satisfy the data-dependent theory and are suitable for classifying agricultural products. Therefore, the proposed methods could be applied to perform classification tasks efficiently in many fields of study.

Keywords: Classification, least squares, linear prediction, prediction errors, robust

INTRODUCTION

This article focuses on univariate classification methods. Classification methods often assign an object to the actual groups based on certain rules (Tang et al., 2014). Univariate classification methods have been discussed extensively in different fora (Gupta & Govindarajulu, 1973; Huberty & Holmes, 1983). The most frequently applied classifiers are based on the group mean estimates, univariate time series (Karimi-Bidhendi et al., 2018; Song et al., 2020), and the Bayes probability rule (Harianto et al., 2020; Ye, 2020). Unfortunately, these classifiers are influenced by outliers, thereby resulting in a high misclassification rate. The outliers are weighted or deleted, thereby resulting in significant information loss to minimize the misclassification rate. Other robust estimates used as a plug-in to robustify the mean classifiers are the minimum covariance determinant (MCD) (Hubert et al., 2018; Hubert & Debruyne, 2010; Leys et al., 2019), the S and M estimators (Almetwally & Almongy, 2018; Campbell et al., 1999; Croux et al., 1994; Kordestani et al., 2020; Verardi & McCathie, 2012). However, these estimators applied to develop robust classifiers often result in vital information loss. Thus, the Bayes rule is a unique univariate classifier that does not depend on the mean and covariance methods but may perform poorly if the data set in one group is significantly larger than the data set in the other. To avert the above problems, we proposed robustifying the above methods by estimating the parameters of interest using the least square approach (LSA) and linear prediction approach (LPA). The LSA and the LPA estimate drastically minimize the loss of information and hence are better estimates to be applied as a plug-in to learn or train the classical classifiers.

The concept and applications of the linear prediction approach (LPA) have been discussed in detail (Atal, 2006; Manolakis & Proakis, 1996). Linear prediction is based on the theory of estimation (Marple & Carey, 1989). It is a robust and dependable predictive estimator (Srivastava, 2017). Prediction based on linear or multivariate methods applies information on linear or multivariate variables. For example, let δ be the dependent random variable and x_i, $i = 1,2,3, \ldots, k$ be the independent random variable, otherwise called the "predictor random variable." Useful information can be obtained if the Borel functions are defined (Bickel & Doksum, 2015; Dobler, 2002; Lindley, 1999; Penenberg, 2015). The variables used in defining the Borel functions are random variables that generate a subspace

w of the Hilbert space H_{sp}.This concept produces the most tractable prediction variable δ, that relies on x_i. The assumption is that the tractable predictor and the independent variables, x_i have a normal joint distribution (Jaeger, 2006).

The least-square approach (LSA), like other estimation procedures and its variant, has received extensive coverage (He et al., 2021; Drygas, 2021; Yao et al., 2020; Kern, 2016; Miller, 2006). This unique technique can be traced to Galton (1886) though coined by Legendre in 1800s. Pearson and Fisher expanded the work of Galton in diverse ways. The main objective of applying this procedure is to estimate and fit the given data set into the function to obtain numerical value (Miller, 2006). It can be done by considering pairs of observations (Y_n, X_n), n = 1,2, .. , K, which consist of the dependent random variables Y_n and the independent random variables X_n. To perform prediction involves the linear combination of these variables (Y_n, X_n).

The LSA tends to minimize the parameters of interest by estimation based on the sum of squares deviation (Kern, 2016). It is also applied to determine the line of best fit of the given data set. The LSA has been applied to provide solutions to a power line (Girshin et al., 2016), pressure detection (Sun et al., 2015), motor induction (Koubaa, 2006), data fitting (Chen & Liu, 2012), and identification of groundwater pollution (He et al., 2021). On the other hand, the LPA has been applied to solve different problems, including the travel time and modeling the prediction of Covid-19 outbreak (Ogundokun et al., 2020; Olarenwaju & Harrison, 2020), climate change (Hasselman & Barnett, 1981) and data forecasting (Vaseghi, 2008).

In applied research, measurement errors or data imputation errors frequently occur if the process is not properly monitored or equipment calibration failure, which may result in data point differential often called influential observations or outliers. Influential observation is described as a data point that is far away from most of the data points. Influential observations often alter the performance of the classical methods, such as the nearest mean classifiers (NMC) (Okwonu & Othman, 2012; Skurichina & Duin, 2000) that depends on the sample mean. Hence robust methods are applied to overcome this problem by weighting the data set or deleting the influential observations. However, these procedures often result in information loss. This paper applies the LSA and LPA to obtain robust prediction estimates (Srivastava, 2017) without information loss. We propose two robust classification rules based on the LSA and the LPA estimates. We further apply the predicted estimates from the LSA and LPA to develop four hybrid classifiers. Finally, we compare these proposed methods with the classical nearest mean classifier (NMC) (Okwonu & Othman, 2012; Skurichina & Duin, 2000) and the Bayes classifier.

The comparative classification performance of these methods is also investigated based on the probability of correct classification (PCC) and the percentage performance prediction accuracy (PPPA). These classifiers were adopted to investigate whether horn measurement can be used to determine cattle gender. We also applied these classifiers to investigate whether their shapes can determine banana variety. This study proposes robust hybrid univariate classifiers that do not expunge outliers, thereby minimizing the loss of vital information. Therefore, the objectives are (1) to minimize the loss of vital information, (2) to minimize the misclassification rate, (3) to derive new hybrid classifiers with robust classification accuracy, (4) to investigate the comparative classification performance of the conventional univariate classifiers and the proposed hybrid classifiers and (5) to investigate the validity of the data dependency theory which states that the performance of any classifier strictly depends on the data structure and sign direction.

The rest of this paper is structured as follows. First, the LSA, LPA, and hybrid methods are explained in Section 2. Then, data collection and analysis are presented in Section 3. Finally, the conclusion follows in Section 4.

MATERIALS AND METHODS

The univariate classifier (UC) and the Bayes classifier (BC) are known classification methods for univariate applications. The UC is based on mean computation, while the BC is designed using the probability concept. In this paper, we will skip the rigors of the formulations and focus on the estimate and plug-in methods.

Linear Prediction Approach (LPA)

The LPA has gained wide coverage to the extent that its coefficients are termed backward and forward autoregression (Eriksson et al., 2019; Mello, 2006; Engle, 1982; Jones, 1978). It has applications in digital signal processing, economics, and many other disciplines (Randall et al., 2020; Tan & Jiang, 2018; Srivastava, 2017; Manolakis & Proakis, 1996; Bultheel & van Barel, 1994). In addition, the LPA produces robust predictive estimates (Srivastava, 2017). We start by defining the dependent and independent random variables to develop the LPA classifier. Let δ denote the dependent random variable and x_i, $_i = 1,2,3, \ldots, k$ be the independent predictor random variables. Suppose:

$$\delta = \theta_0 x^0 + \sum_{i=1}^{k} \theta_i x_i \,.$$

The expression can be written as:

$$\delta - \theta_0 - \theta_1 x_1 - \theta_2 x_2 - \cdots - \theta_k x_k = \delta - \theta_0 - \theta_1 x_1 - \theta_2 x_2 - \cdots - \theta_k x_k = 0.$$

Taking the expectation of the last expression and squaring it, we obtain

$$E|\delta - \theta_0 - \theta_1 x_1 - \theta_2 x_2 - \cdots - \theta_k x_k|^2 = 0.$$

It can be expressed as Equation 1:

$$E|\delta|^2 = E|\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_k x_k|^2 = |\theta_0|^2 + \left|\sum_{i=i}^{k} \theta_i x_i\right|^2. \tag{1}$$

It implies that the mean squared error (MSE) is minimized. The focus is to determine the dependent variable δ based on x_i.Then, iterative steps can be introduced on the dependent variable to comply with the predictor variables x_i.

Let $\vartheta = 1$ be a constant, define β as the Hilbert space and φ be the subspace of β, then $\langle \vartheta, x_i, i = 1,2,3,..,k \rangle$ is the random variable defined on $\varphi \epsilon \beta$ (Bickel & Doksum, 2015; Lindley, 1999). This process can be viewed as a "minimization problem." Further analysis revealed that $\delta = <\delta_i>$ could be paired with $\langle \vartheta, x_i, i = 1,2,3,..,k \rangle$. The first pairing based on expectation property yields

$$\gamma_{i\delta} = \frac{|\delta, \vartheta|}{|\vartheta|^2} \times |\vartheta| = |\delta, \vartheta| = E(\delta) = \bar{\delta}.$$

It satisfies $E|\delta - \vartheta|^2 = \bar{\delta}$ and $\alpha = \delta - E(\delta)$. The variance of δ, that is δ^2 can be computed in a similar procedure. Suppose there exists a random variable x such that where $\hat{\hat{x}} = \mu = E(x)$ and the variance of x defined as $var(x) = E|x - \mu|^2$. Based on the projection concept we have

$$\gamma_{i\delta} = \frac{|\delta, \vartheta|}{|\vartheta|^2} \times |\vartheta| + \frac{|\delta, \varepsilon|}{|\varepsilon|^2} \times |\varepsilon| = \bar{\delta} + \frac{|\alpha, \varepsilon|}{|\varepsilon|^2} \times |\varepsilon| + \frac{|\bar{\delta}, \varepsilon|}{|\varepsilon|^2} \times |\varepsilon|$$

$$= \bar{\delta} + \frac{|\alpha, \varepsilon|}{|\varepsilon|^2} \times |\varepsilon| + \frac{0}{|\varepsilon|^2} \times |\varepsilon|$$

$$= \bar{\delta} + \frac{|\alpha, \varepsilon|}{|\varepsilon|^2} \times |\varepsilon|, \left(|\bar{\delta}, \varepsilon| = 0\right).$$

Recall that the covariance between δ and x is denoted as $cov(\delta, x) = (\alpha, \varepsilon)$; it indicates the relationships between the two variables. From the last expression, we obtain $\frac{|\alpha,\varepsilon|}{|\varepsilon|^2} = \rho_{\delta x}\sqrt{\frac{\sigma_\delta^2}{\sigma_x^2}}$. It implies that the prediction can be performed as Equation 2:

$$\widehat{\gamma_{i\delta}} = \bar{\delta} + \rho_{\delta x}\sqrt{\frac{\sigma_\delta^2}{\sigma_x^2}} \times |\varepsilon|, \tag{2}$$

where $\rho_{\delta x}$ denotes the correlation between δ and x_i, this implies that σ_δ^2 and σ_x^2 are the variance of δ and x, respectively. The analysis indicates that δ can be estimated based on x.

From Equation 2, we can derive the LPA classifier as follows. First, we obtain each group predicted estimate; therefore, Equation 2 can be defined as group predictors such that $k = 1,2$; then, we restate Equation 2 as Equation 3:

$$\widehat{\gamma_{i\delta}}_k = \bar{\delta}_k + \rho_{\delta x_k}\sqrt{\frac{\sigma^2_{\delta\ k}}{\sigma^2_{x\ k}}} \times |\varepsilon_k|. \tag{3}$$

The cutoff mark of Equation 3 is defined in Equation 4:

$$\mathfrak{y} = \frac{\sum_{k=1}^{2} \widehat{\gamma_{i\delta}}_k}{2}. \tag{4}$$

The LPA classifier assigns an object to group one (G_1) if $\widehat{\gamma_{i\delta}}_1 < \mathfrak{y}$. otherwise to group two (G_2) if $\widehat{\gamma_{i\delta}}_1 < \mathfrak{y}$.

The Least Square Approach (LSA)

Prediction by the LSA has received detailed attention in the literature. However, this subsection adopts a brief discussion on its prediction approach. The LSA can be stated as Equation 5:

$$\hat{y} = a + bx, \tag{5}$$

where x denotes the predictor variable, $\hat{y}$ denotes the estimate of the response variable, and b denotes the slope,

$$b = \frac{SS_{xy}}{SS_{xx}}, SS_{xy} = \sum xy - \frac{\sum x \sum y}{k}, SS_{xx} = \sum x^2 - \frac{(\sum x)^2}{k},\ \bar{x} = \frac{\sum_{i=1}^{k} x_i}{k}, \bar{y} = \frac{\sum_{i=1}^{k} y_i}{k},$$

and the intercept is given as $a = \bar{y} - b\bar{x}.$ We apply Equation 5 to derive the LSA classifier by restating Equation 5 as Equation 6:

$$\hat{y}_k = a_k + b_k x_k. \tag{6}$$

Then the cutoff mark of Equation 6 is defined as Equation 7:

$$\beta = \frac{\sum_{k=1}^{2} \overline{\hat{y}_k}}{2}. \tag{7}$$

The LSA classifies an object $(\hat{y}_1)$ to group one (G_1) if $\hat{y}_1 \geq \beta$, otherwise assigns $\hat{y}_1$ to group two (G_2) if $\hat{y}_1 < \beta$.

Hybrid Linear Prediction Classifier (HLPC)

From the above discussions, we have given a detailed description of the LPA concerning prediction and classification. The output from Equation 2 is assumed to be a robust estimate to perform group classification for univariate cases. In this subsection, we will invoke the univariate classifier (Huberty & Holmes, 1983), Bayes classifier (Theodoridis

& Koutroumbas, 2009; Ma et al., 2011), the smart univariate classifier (SUC), and smart univariate Bayes classifier (SUBC) to form the four hybrid classifiers categorized as the hybrid linear prediction classifiers (HLPC). We will not undergo derivational details but apply the LPA and LSA predictive values as a plug-in to train these classifiers.

Univariate Linear Predictive Classifier (ULPC)

We apply the estimates from Equation 2 as input to construct the ULPC model and decision boundary as Equations 8 and 9:

$$\overline{q_k} = \frac{\sum_{k=1}^{2}(\widehat{\gamma_{i\delta k}})}{n_k}, \tag{8}$$

$$\overline{\gamma_{i\delta k}} > \frac{\overline{q_1}+\overline{q_2}}{2}. \tag{9}$$

Therefore Equations 8 and 9 are the univariate linear predictive classifier based on the decision rule from Huberty and Holmes (1983).

Linear Predictive Bayes Classifier (LPBC)

Based on Equation 2 and the concept discussed in Theodoridis and Koutroumbas (2009) and Ma et al. (2011), the LPBC is stated as Equations 10 and 11:

$$P\left(G_1 \middle| \gamma_{\widehat{i\delta k,p\times 1}}\right) = \frac{w_1 t_1\left(\gamma_{\widehat{i\delta k,p\times 1}}\right)}{\sum_{k=1}^{2} w_k t_k\left(\gamma_{\widehat{i\delta k,p\times 1}}\right)} = \left(\sum_{k=1}^{2} w_k t_k\left(\gamma_{\widehat{i\delta k,p\times 1}}\right)\right)^{-1} w_1 t_1\left(\gamma_{\widehat{i\delta k,p\times 1}}\right) \tag{10}$$

$$P\left(G_2 \middle| \gamma_{\widehat{i\delta k,p\times 1}}\right) = \frac{w_2 t_2\left(\gamma_{\widehat{i\delta k,p\times 1}}\right)}{\sum_{k=1}^{2} w_k t_k\left(\gamma_{\widehat{i\delta k,p\times 1}}\right)} = 1 - \left(\sum_{k=1}^{2} w_k t_k\left(\gamma_{\widehat{i\delta k,p\times 1}}\right)\right)^{-1} w_1 t_1\left(\gamma_{\widehat{i\delta k,p\times 1}}\right) \tag{11}$$

Hence, assign $\widehat{\gamma_{i\delta 1}}$ to G_1 if $P\left(G_1 \middle| \gamma_{\widehat{i\delta 1,p\times 1}}\right) \geq P\left(G_2 \middle| \gamma_{\widehat{i\delta 2,p\times 1}}\right)$ otherwise, allocate $\widehat{\gamma_{i\delta 1}}$ to G_2 if

$$P\left(G_1 \middle| \gamma_{\widehat{i\delta 1,p\times 1}}\right) < P\left(G_2 \middle| \gamma_{\widehat{i\delta 2,p\times 1}}\right).$$

The LPBC applies the Bayes classifier rule to assign an object to the actual group.

Smart Univariate Linear Predictive Classifier (SULPC)

We apply the input from Equation 2 as a plug-in to train the classifier as Equations 12 and 13

$$T_1 = \frac{(\bar{q}_1 - \bar{q}_2)^{`}}{S^2_{\gamma_{\widehat{i\delta,p\times 1}}}} w_1 \gamma_{\widehat{i\delta 1,p\times 1}}, \tag{12}$$

$$T_2 = \frac{(\bar{q}_1 - \bar{q}_2)^{`}}{S^2_{\gamma_{\widehat{i\delta,p\times 1}}}} w_2 \gamma_{\widehat{i\delta 2,p\times 1}}. \tag{13}$$

Then the F-weight (w_k) and the pooled variance $\left(S^2_{\gamma_{\widehat{i\delta,p\times 1}}}\right)$ based on Equation 2 are stated as

$$w_1 = \frac{\gamma_{\widehat{i\delta 1,p\times 1}}}{Z}, w_2 = \frac{\gamma_{\widehat{i\delta 2,p\times 1}}}{Z}, Z = \gamma_{\widehat{i\delta 1,p\times 1}} \gamma_{\widehat{i\delta 2,p\times 1}}^{\ T}, S^2_{G_1} = \frac{\sum_{i=1}^{G_1} \left(w_1 \gamma_{\widehat{i\delta 1},p\times 1} - \bar{q}_1\right)^2}{n_{G_1} - 1},$$

$$S^2_{G_2} = \frac{\sum_{i=1}^{G_2} \left(w_2 \gamma_{\widehat{i\delta 2},p\times 1} - \bar{q}_2\right)^2}{n_{G_2} - 1}, S^2_{\gamma_{\widehat{i\delta,p\times 1}}} = \frac{\left(n_{G_k} - 1\right) \sum_{k=}^{2} S^2_{G_k}}{\sum_{k=}^{2} n_{G_k} - 2}.$$

The group evaluation criteria are obtained as Equation 14 to evaluate the performance of this method:

$$T_{t1} = \frac{(\bar{q}_1 - \bar{q}_2)^{`}}{S^2_{\gamma_{\widehat{i\delta,p\times 1}}}} \overline{q_1}, \quad T_{t2} = \frac{(\bar{q}_1 - \bar{q}_2)^{`}}{S^2_{\gamma_{\widehat{i\delta,p\times 1}}}} \overline{q_2}, T_{t1t2} = \frac{T_{t1} + T_{t2}}{2}. \tag{14}$$

Therefore, an object $w_1 \gamma_{\widehat{i\delta 1,p\times}1}$ is allocated to G_1 if $T_1 \geq T_{t1t2}$; otherwise, assign $w_1 \gamma_{\widehat{i\delta 1,p\times}1}$ to G_2 if $T_1 \geq T_{t1t2}$.

LPBC/ SULPC

The LPBC/SULPC combines LPBC and SULPC to produce unbiased robust classification results. This combination averts the overfitting problem and upward bias. Overfitting is a process whereby the model predicted value exceeds the given optimal probability of correct classification (PCC).

Hybrid Least Square Classifier (HLSC)

The four methods, i.e., the univariate classifier, Bayes classifier, smart univariate classifier, and smart univariate Bayes classifier discussed in the last subsection, utilize the LSA estimate (Equation 5) to train the different methods. The HLSC consists of the univariate least square classifier (ULSC), least square Bayes classifier (LSBC), the least square smart univariate classifier (LSSUC), and LSBC/LSSUC. Similar plug-in procedures discussed in Equations 8 to 14 are implemented by replacing the LPA estimates with the LSA estimates.

Evaluation Criteria

The evaluation criteria (Huberty & Holmes, 1983) applied in this study are based on Equation 15:

$$C_\sigma = 0.5\Phi\left(\frac{\alpha}{2}\right) + 0.5\Phi\left(\frac{\alpha}{2}\right) = \Phi\left(\frac{\alpha}{2}\right) \quad (15)$$

Where Φ (.) is the cumulative distribution function of the standard normal distribution, and α is the Mahalanobis distance (Johnson & Wichern, 1992). The probability of correct classification (*PCC*) denoted by ω, derived from the various methods, is compared with the optimal value (C_σ) to investigate the best method. The error of misclassification ($\epsilon = C_\sigma - \omega$)can also be applied to determine the robustness of the methods. Equation 16 is called the percentage of performance prediction accuracy (PPPA),

$$\Omega=\left(\frac{\omega}{C_\sigma}\right) \times 100 \quad (16)$$

The last expression determines the overall percentage of a correct group membership. It is useful to analyze the performance of the methods at a glance. Equation 16 will be adopted to analyze the comparative performance analysis of these methods.

Data Collection and Analysis

These data sets were collected to investigate the comparative classification performance for the above classifiers. The applications of this study focus on two real data sets from the agricultural sector to determine whether these classifiers could be used to maximally separate different species and varieties of agricultural products. The first data consist of cattle horns measurement in Appendices 1 and 2. The second consists of artificial data on varieties of banana shapes (https://www.openml.org/d/1460). The first data set consists of collections of cattle horns for ten months in an abattoir in Abraka, Delta State, Nigeria. This data set consists of two features: horn length and width measured in centimeters (cm) for bull and cow, with 100 instances categorized into two groups. The first group consists of features measured on a bull, while the second group consists of features measured on a cow. This data uses to determine whether the classifiers can accurately predict cattle gender. Appendix 1 consists of the bull data set, categorized as Group one (G_1), and Appendix 2 consists of the cow data set, categorized as Group two (G_2). The banana variety data set originally contains n = 5,300 with two attributes. The second group (G_2) consist of n_2 = 2,376, we selected n_1 = 2,376, hence $n_1 = n_2$ = 2,376, $n = n_1 + n_2$. The details for this data set are contained in (https://www.openml.org/d/1460 (KEEL, 2015). The data set was reshuffled into a training set (60%) and a validation set (40%). The mean probability of correct classifications is based on 1000 replications. Both data sets will be analyzed based on the percentage of performance prediction accuracy (PPPA).

RESULTS AND DISCUSSIONS

This section consists of the applications of agricultural production data. The focus is to investigate if these classifiers could be applied to separate animal species or gender and plant varieties into different groups based on measured attributes. This study mimics the univariate measurement for the classification of nanoelectronics and spectroscopy (Leys et al., 2019) and the study by Huberty and Holmes (1983).

Application 1: Cattle Horns Data Set

The results in Table 1 and Figure 1 are based on Equation 16, demonstrating the comparative performance analysis between the classical univariate nearest mean classifier (NMC) and the hybrid NMC based on LSA and LPA predicted estimates. From the analysis, we observed that the LSANMC and LPANMC have higher PPPA values (>95%) than the classical NMC. Furthermore, it shows that the robust predictive estimates of LSA and LPA enhance classification accuracy better than the classical NMC method. The comparative performance analysis is depicted in Figure 1. The findings revealed that the NMC method based on the LSA and LPA is more resistant to influential observations than the classical NMC.

Table 1

Comparative performance of NMC and the hybrid NMC for cattle gender

Classical Methods	PPPA
NMC	64.20
LSANMC	97.00
LPANMC	99.00

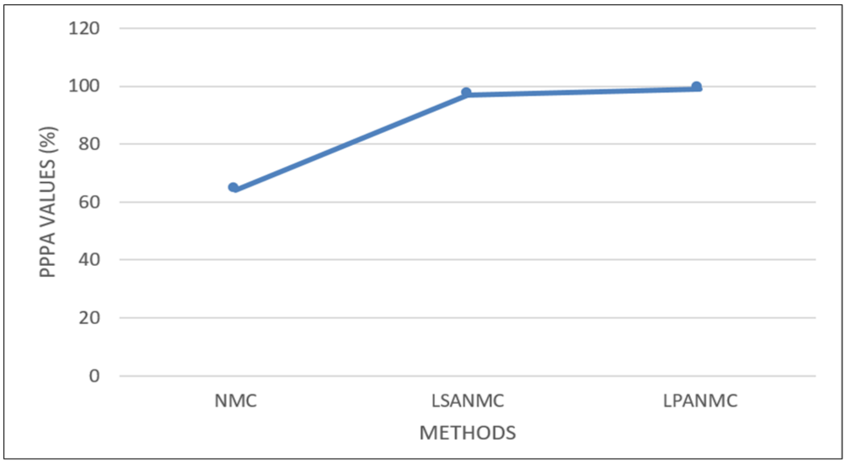

Figure 1. Comparative analysis of the classical NMC and the hybrid NMC for cattle gender

The results in Table 2 are based on Equation 16, demonstrating the comparative performance analysis between the conventional and hybrid methods. The result revealed that all the methods have a very high percentage performance prediction accuracy (PPPA) (Ω). The conventional methods have an average of 95.14% PPPA, while the HLPC and HLSPC have 98.75% and 99.25% PPPA for classifying cattle by gender. Based on the average PPPA for all the methods, the hybrid methods showed more robust classification accuracy than the conventional methods. From this analysis, we remark that the PPPA adopted to analyze the performance of these methods is suitable for the classification task. In Figure 2, the hybrid methods have more robust PPPA values than conventional ones. However, the probability methods (Bayes classifier (BC) and smart univariate Bayes classifier (SUCBC)) demonstrated comparable performance to some of the hybrid methods.

Table 2

Comparative analysis of percentage performance prediction accuracy (PPPA)

Conventional Methods (Average %)		HLPC (%)		HLSPC (%)	
UC	89.54	ULPC	97.00	ULSC	99.00
BC	98.12	LPBC	100.00	LSBC	100.00
SUC	89.05	SULPC	98.00	LSSUC	98.00
SUCBC	98.12	LPBC/ SULPC	100.00	LSBC/ LSSUC	100.00
LPA	99.00				
LSA	97.00				

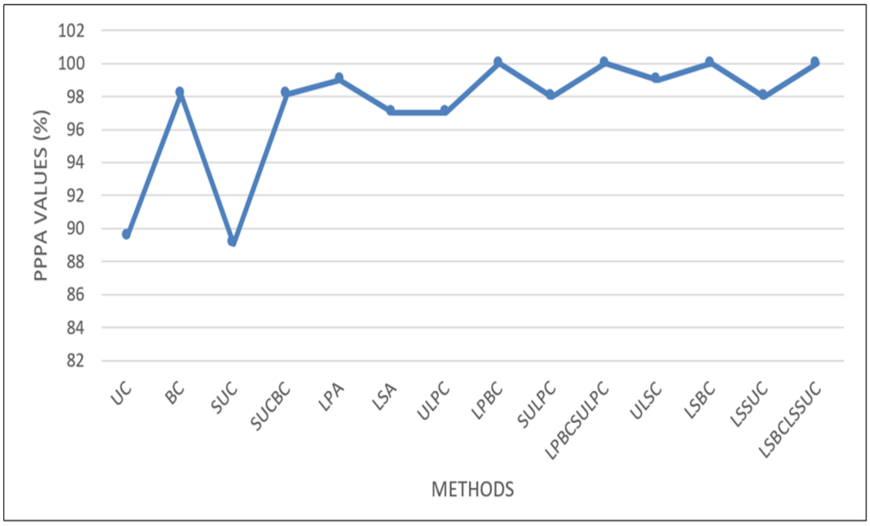

Figure 2. Comparative analysis of PPPA values for cattle gender classification

Application 2: Banana Variety Data Set

The classification results reported in Table 3 show that the classical NMC has betterclassification performance than the hybrid NMC. However, all the methods have over 80% PPPA, as illustrated in Figure 3.

In comparison to the results reported in Table 1, the analysis in Table 3 showed that the performance of these methods is data-dependent. The former (Table 1) showed an upward trend, while the latter (Table 3) showed vice versa. The implication of the comparative analysis indicated that the performance of any classifier depends on the data set. It affirmed the data dependency theory, which states that the performance of any classification method depends strictly on the data structure and sign direction. The data structure can be continuous or discrete.

Table 3

Comparative performance of NMC and the hybrid NMC for a banana variety

Classical Methods	PPPA
NMC	98.14
LSANMC	84.89
LPANMC	84.89

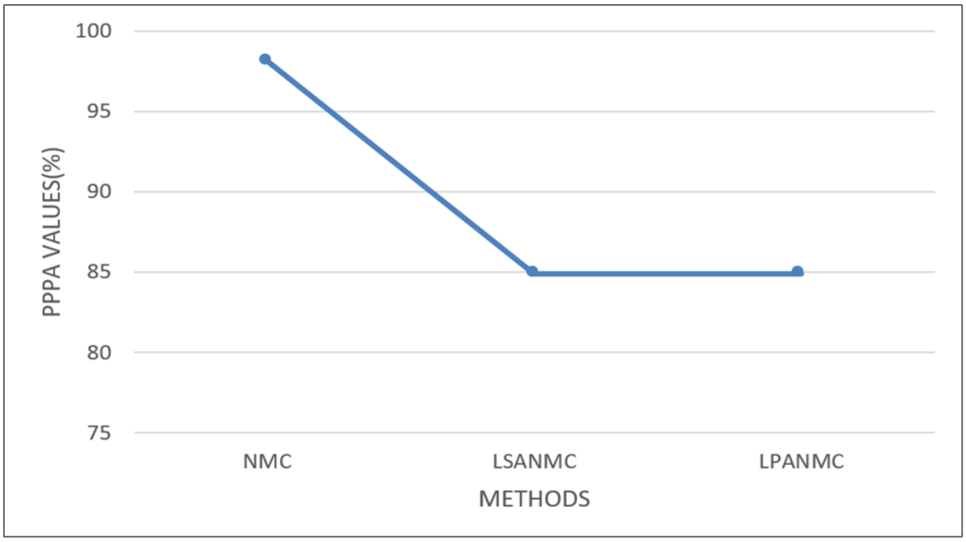

Figure 3. Comparative analysis of the classical NMC and the hybrid NMC for a banana variety

Table 4
Comparative analysis of percentage performance prediction accuracy (PPPA)

Conventional Methods (%)		HLPC (%)		HLSPC (%)	
UC	97.85	ULPC	88.81	ULSC	88.81
BC	99.93	LPBC	93.51	LSBC	93.51
SUC	99.98	SULPC	89.85	LSSUC	89.85
SUCBC	99.96	LPBC/SULPC	93.51	LSBC/LSSUC	93.51
LPA	93.51				
LSA	93.51				

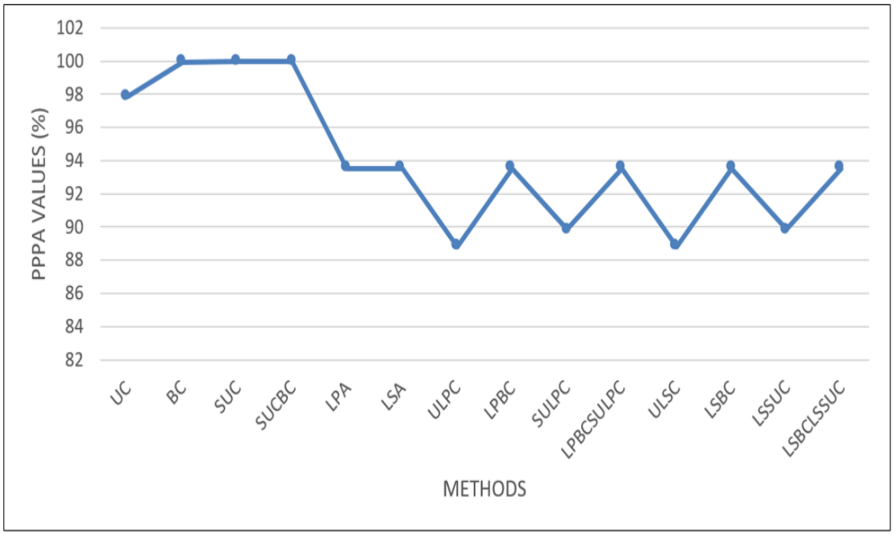

Figure 4. Comparative analysis of PPPA values for banana variety classification by shapes

The comparative percentage performance prediction accuracy based on this data set is reported in Table 4. The result revealed that all the methods have over 88% correct group membership prediction. This result demonstrated that these methods are capable of classifying banana varieties based on shapes with a minimum PPPA of 88.81% and a maximum PPPA of 99.98%. The conventional methods have an average PPPA of 97.46% banana variety classification by shapes, while the hybrid methods HLPC and HLSPC have 91.42% average PPPA, respectively. The hybrid methods based on LPA (HLPC) and LSA (HLSPC) have similar results for this data set. This unique performance showed that the conventional classifiers demonstrated more robust classification accuracy than the hybrid methods. Figure 4 shows that the conventional methods have superior performance over the hybrid methods. However, the hybrid probability methods (LPBC, LSBC, LPBC/SULPC, LSBC/LSSUC) have the same PPPA values as the LPA and LSA methods.

The cattle data set and the classifiers demonstrated that cattle gender could be determined by horn length and base width measurements. The study also indicated that shapes could classify the banana variety. The 90% average PPPA classification accuracy for all the methods based on the two data sets showed that these methods are robust. The analysis in Figures 1 and 2 imply that the hybrid methods showed superior performance over the conventional methods, while in Figures 3 and 4, the conventional methods altered the superior performance of the hybrid methods except for the probability-based hybrid methods. From this analysis, we observed the upward and downward trends as depicted in Figures 1 to 4, which affirmed the data dependency theory. We also observed that the LSA and LPA showed consistent performance in classifying cattle gender and banana variety. This consistent robust performance was also observed in the hybrid classifiers. The strength of the data dependency theory on the classification methods was obvious in the two data sets, as shown in the upward and downward trends depicted in Figures 1 to 4. The limitations of these classifiers are based on the validity of the data dependency theory.

CONCLUSION

We have proposed the LPA and LSA classification rules and four hybrid classifiers. The evaluation criteria were also established and reinforced into PPPA for easy and fast analysis. The performance of these classifiers is demonstrated using agricultural produce data. The first data set applied to test the performance comparison of these methods was obtained by measuring the length and base width of cattle horns. The second data set consists of two classes of banana variety. The result affirmed that we could apply the proposed LPA, LSA, and the hybrid classifiers to robustly classify cattle into gender based on horn and width length measurement and classify banana variety based on shapes. The investigated LPA and LSA techniques showed comparable classification accuracy with PPPA of over 90%. The analysis revealed that the proposed and hybrid classifiers are robust enough to perform classification based on these data sets. These techniques generally showed varying high percentage performance prediction accuracy based on the data sets. The results demonstrated that these methods could be applied as alternative classifiers to perform classification tasks. We remark that the results reported in this paper affirmed the effects of data dependency theory on classification methods. We look forward to extending these classifiers to multi-dimensional applications in the future.

ACKNOWLEDGEMENT

The authors want to express our gratitude to Universiti Utara Malaysia and Delta State University for supporting this research work. In addition, the authors acknowledged Dr. Okwonu and Ms. Blessing. for contributing the cattle data set for this study. However, this research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

REFERENCES

Almetwally, E. M., & Almongy, H. M. (2018). Comparison between M-estimation, S-estimation, and MM estimation methods of robust estimation with application and simulation censoring. *International Journal of Mathematical Archive*, *9*(11), 1-9.

Atal, B. S. (2006). The history of linear prediction. *IEEE Signal Processing Magazine*, *23*(2), 154-161.

Bickel, P. J., & Doksum, K. A. (2015). *Mathematical Statistics: Basic Ideas and Selected Topics* (Vol. 1, 2nd Ed). Chapman and Hall/CRC.

Bultheel, A., & van Barel, M. (1994). Linear prediction: Mathematics and engineering. *Bulletin of the Belgian Mathematical Society*, *1*, 1-58.

Campbell, N. A., Lopuhaä, H. P., & Rousseeuw, P. J. (1999). On the calculation of a robust S-estimator of a covariance matrix. *Statistics in Medicine, 17*(23), 2685-2695. https://doi.org/10.1002/(SICI)1097-0258(19981215)17:23<2685::AID-SIM35>3.0.CO;2-W

Chen, C., & Liu, C. J. (2012). The application of total least squares method in data fitting of speed radar. *Applied Mechanics and Materials*, *203*, 69-75. https://doi.org/10.4028/www.scientific.net/amm.203.69

Croux, C., Rousseeuw, P. J., & Hossjer, O. (1994). Generalized S-estimators. *Journal of the American Statistical Association*, *89*(428), 1271-1281. https://doi.org/10.2307/2290990

Dobler, P. C. (2002). Mathematical statistics: Basic ideas and selected topics. *The American Statistician*, *56*(4), 332-332. https://doi.org/10.1198/tas.2002.s204

Drygas, H. (2011). On the relationship between the method of least squares and Gram-Schmidt orthogonalization. *Acta et Commentationes Universitatis Tartuensis de Mathematica*, *15*(1), 3-13. https://doi.org/10.12697/ACUTM.2011.15.01

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, *50*(4), 987-1007. https://doi.org/10.2307/1912773

Eriksson, A., Preve, D., & Yu, J. (2019). Forecasting realized volatility using a nonnegative semiparametric model. *Journal of Risk and Financial Management*, *12*(3), Article 139. https://doi.org/10.3390/jrfm12030139

Girshin, S. S., Kuznetsov, E. A., & Petrova, E. V. (2016, May). Application of least square method for heat balance equation solving of overhead line conductors in case of natural convection. In *2016 2nd International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM)* (pp. 1-5). IEEE Publishing. https://doi.org/10.1109/ICIEAM.2016.7911417.

Gupta, A. K., & Govindarajulu, Z. (1973). Some new classification rules for *c* univariate normal populations. *Canadian Journal of Statistics*, *1*(1-2), 139-157. https://doi.org/10.2307/3314996

Harianto, H., Sunyoto, A., & Sudarmawan, S. (2020). Optimasi algoritma Naïve Bayes classifier untuk mendeteksi anomaly dengan univariate fitur selection [Naïve Bayes algorithm optimization classifier to detect anomalies with univariate feature selection]. *Edumatic: Jurnal Pendidikan Informatika*, *4*(2), 40-49. https://doi.org/10.29408/edumatic.v4i2.2433

Hasselmann, K., & Barnett, T. P. (1981). Techniques of linear prediction for systems with periodic statistics. *Journal of Atmospheric Sciences*, *38*(10), 2275-2283. https://doi.org/10.1175/1520-0469(1981)038<2275:TOLPFS>2.0.CO;2

He, Z., Zuo, R., Zhang, D., Ni, P., Han, K., Xue, Z., Wang, J., & Xu, D. (2021). A least squares method for identification of unknown groundwater pollution source. *Hydrology Research*, *52*(2), 450-460. https://doi.org/10.2166/nh.2021.088

Hubert, M., & Debruyne, M. (2010). Minimum covariance determinant. *WIREs Computational Statistics*, *2*(1), 36-43. https://doi.org/10.1002/wics.61

Hubert, M., Debruyne, M., & Rousseeuw, P. J. (2018). Minimum covariance determinant and extensions. *WIREs Computational Statistics*, *10*(3), Article e1421. https://doi.org/10.1002/wics.1421

Huberty, C. J., & Holmes, S. E. (1983). Two-group comparisons and univariate classification. *Educational and Psychological Measurement*, *43*(1), 15-26. https://doi.org/10.1177/001316448304300103

Jaeger, B. (2006). The method of least squares. In A. A. Lazakidou (Ed.), *Handbook of Research on Informatics in Healthcare and Biomedicine* (pp. 181-185). IGI Global. https://doi.org/10.4018/978-1-59140-982-3.ch023

Johnson, R. A., & Wichern, D. W. (1992). *Applied Multivariate Statistical Analysis* (3rd Ed.). Prentice-Hall.

Jones, R. H. (1978). Multivariate autoregression estimation using residuals. In *Applied Time Series Analysis I,* (pp. 139-162). Academic Press. https://doi.org/10.1016/B978-0-12-257250-0.50009-X

Karimi-Bidhendi, S., Munshi, F., & Munshi, A. (2018). Scalable classification of univariate and multivariate time series. In *2018 IEEE International Conference on Big Data (Big Data)* (pp. 1598-1605). IEEE Publishing. https://doi.org/10.1109/BigData.2018.8621889

KEEL. (2015). *Banana*. OpenML. https://www.openml.org/d/1460

Kern, M. (2016). *Numerical Methods for Inverse Problems*. John Wiley & Sons.

Kordestani, M., Hassanvand, F., Samimi, Y., & Shahriari, H. (2020). Monitoring multivariate simple linear profiles using robust estimators. *Communications in Statistics - Theory and Methods*, *49*(12), 2964-2989. https://doi.org/10.1080/03610926.2019.1584314

Koubaa, Y. (2006). Application of least-squares techniques for induction motor parameters estimation. *Mathematical and Computer Modelling of Dynamical Systems*, *12*(4), 363-375. https://doi.org/10.1080/13873950500064103

Leys, C., Delacre, M., Mora, Y. L., Lakens, D., & Ley, C. (2019). How to classify, detect, and manage univariate and multivariate outliers, with emphasis on pre-registration. *International Review of Social Psychology*, *32*(1), 1-10. https://doi.org/10.5334/irsp.289

Lindley, D. V. (1999). Introduction to the practice of statistics, by David S. Moore and George P. McCabe. Pp. 825 (with appendices and CD-ROM).£ 27.95. 1999. ISBN 0 7167 3502 4 (WH Freeman). *The Mathematical Gazette, 83*(497), 374-375. https://doi.org/10.2307/3619120

Ma, D., Wei, W., Hu, H., & Guan, J. (2011). The application of Bayesian classification theories in distance education system. *International Journal of Modern Education and Computer Science*, *3*(4), 9-16.

Manolakis, D. G., & Proakis, J. G. (1996). *Digital Signal Processing. Principles, Algorithm, and Applications* (4th Ed.). Prentice-Hall International Inc.

Marple, S. L., & Carey, W. M. (1989). Digital spectral analysis with applications. *The Journal of the Acoustical Society of America*, *86*, Article 2043. https://doi.org/10.1121/1.398548

Mello, L. (2006). *Linear Predictive Coding as an Estimator of Volatility*. arXiv e-Print. https://doi.org/10.48550/arXiv.cs/0607107

Miller, S. J. (2006). *The method of least squares*. Williams College. https://web.williams.edu/Mathematics/sjmiller/public_html/probabilitylifesaver/MethodLeastSquares.pdf

Ogundokun, R. O., Lukman, A. F., Kibria, G. B., Awotunde, J. B., & Aladeitan, B. B. (2020). Predictive modelling of COVID-19 confirmed cases in Nigeria. *Infectious Disease Modelling*, *5*, 543-548. https://doi.org/10.1016/j.idm.2020.08.003

Okwonu F. Z., & Othman, A. R. (2012). A model classification technique for linear discriminant analysis for two groups. *International Journal of Computer Science Issues (IJCSI)*, *9*(3), 125-128.

Olarenwaju, B. A., & Harrison, I. U. (2020). Modeling of COVID-19 cases of selected states in Nigeria using linear and non-linear prediction models. *Journal of Computer Sciences Institute*, *17*, 390-395.

Penenberg, D. N. (2015). Mathematical Statistics: Basic Ideas and Selected Topics, 2nd edn, vols I and II PJ Bickel and KA 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 Serires, 179*(4), 1128-1129.

Randall, R. B., Antoni, J., & Borghesani, P. (2020). Applied digital signal processing. In R. Allemang & P. Avitabile (Eds), *Handbook of Experimental Structural Dynamics* (pp.1-81). Springer. https://doi.org/10.1007/978-1-4939-6503-8_6-1

Skurichina, M., & Duin, R. P. (2000). Boosting in linear discriminant analysis. In *International Workshop on Multiple Classifier Systems. MCS 2000. Lecture Notes in Computer Science,* (Vol. 1857, pp. 190-199). Springer. https://doi.org/10.1007/3-540-45014-9_18

Song, K., Wang, N., & Wang, H. (2020). A metric learning-based univariate time series classification method. *Information*, *11*(6), Article 288. https://doi.org/10.3390/info11060288

Srivastava, S. (2017). *Fundamentals of linear prediction*. The Institute for Signal and Information Processing. https://www.isip.piconepress.com/courses/msstate/ece_7000_speech/lectures/current/lecture_03/paper/paper.pdf

Sun, W., Zuo, F., Dong, A., & Zhou, L. (2015). Application of least square curve fitting algorithm based on LabVIEW in pressure detection system. In *2015 International Conference on Applied Science and Engineering Innovation* (pp. 39-43). Atlantis Press.

Tan, L., & Jiang, J. (2018). *Digital Signal Processing: Fundamentals and Applications*. Academic Press.

Tang, J., Alelyani, S., & Liu, H. (2014). Feature selection for classification: A review. In C. C. Aggarwal (Ed.), *Data Classification: Algorithms and Applications* (pp. 1-28). Chapman and Hall/CRC. https://doi.org/10.1201/b17320

Theodoridis, S., & Koutroumbas, K. (2009). Classifiers based on Bayes decision theory. In *Pattern recognition* (pp. 13-28). Academia Press.

Vaseghi, S. V. (2008). *Advanced Digital Signal Processing and Noise Reduction*. John Wiley & Sons.

Verardi, V., & McCathie, A. (2012). The S-estimator of multivariate location and scatter in stata. *The Stata Journal: Promoting Communications on Statistics and Stata*, *12*(2), 299-307. https://doi.org/10.1177/1536867X1201200208

Yao, T. T., Bai, Z. J., Jin, X. Q., & Zhao, Z. (2020). A geometric Gauss-Newton method for

least squares inverse eigenvalue problems. *BIT Numerical Mathematics*, *60*(3), 825-852. https://doi.org/10.1007/s10543-019-00798-9

Ye, N. (2020). Naïve Bayes classifier. In *Data Mining* (pp. 1-6). CRC Press. https://doi.org/10.1201/b15288-5

Appendix 1

Length and base width measurements of bull horns (group one)

Horn length (*x*) (cm)	Horn width (*y*) (cm)	Horn length (*x*) (cm)	Horn width (*y*) (cm)	Horn length (*x*) (cm)	Horn width (*y*) (cm)	Horn length (*x*) (cm)	Horn width (*y*) (cm)
17.6	8.6	23.6	10.4	17.6	8.6	23.6	10.4
21.5	9.5	22.7	10.9	21.5	9.5	22.7	10.9
21	7.8	24.2	11.9	21	7.8	24.2	11.9
17.2	6.4	22.4	11	17.2	6.4	22.4	11
16.5	8.2	19.8	8.6	16.5	8.2	19.8	8.6
21.9	8.7	21.2	7.4	21.9	8.7	21.2	7.4
18.1	9.8	20.3	6	18.1	9.8	20.3	6
11.7	8.1	19	5.7	11.7	8.1	19	5.7
15.4	7.2	18.7	5	15.4	7.2	18.7	5
18.2	8.5	19.4	5.3	18.2	8.5	19.4	5.3
17.8	7.4	17.8	4.7	17.8	7.4	17.8	4.7
16.5	6.3	19.8	5.4	16.5	6.3	19.8	5.4
21.5	8.8	17.6	4.2	21.5	8.8	17.6	4.2
22	10.1	18.6	5.1	22	10.1	18.6	5.1
20.1	9	21	6.5	20.1	9	21	6.5
21	10.1	17.7	4.8	21	10.1	17.7	4.8
17.3	7.2	20.2	6	17.3	7.2	20.2	6
19.2	9.4	19.6	5.6	19.2	9.4	19.6	5.6
20.1	10	19.1	5.1	20.1	10	19.1	5.1
18.2	7.4	17.9	4.6	18.2	7.4	17.9	4.6
18	8.2	18.3	5	18	8.2	18.3	5
17.6	7	17.9	4.6	17.6	7	17.9	4.6
18.2	7.3	17.7	4.8	18.2	7.3	17.7	4.8
16.5	9.2	20.2	6	16.5	9.2	20.2	6
21.5	11.6	19.6	5.6	21.5	11.6	19.6	5.6

Appendix 2

Length and base width measurements of cow horns (group two)

Horn length (x) (cm)	Horn width (y) (cm)	Horn length (x) (cm)	Horn width (y) (cm)	Horn length (x) (cm)	Horn width (y) (cm)	Horn length (x) (cm)	Horn width (y) (cm)
23.3	10.3	29	12.7	23.3	10.3	29	12.7
24.5	8.8	30.1	10.9	24.5	8.8	30.1	10.9
25.6	10.4	30.6	17.1	25.6	10.4	30.6	17.1
24.9	9.2	34	11.2	24.9	9.2	34	11.2
29.3	13.1	26.2	12	29.3	13.1	26.2	12
26.2	9.2	30.1	11.6	26.2	9.2	30.1	11.6
25.9	9	29.5	11.3	25.9	9	29.5	11.3
25	8.9	27.4	10	25	8.9	27.4	10
26.5	10.5	25.9	9.8	26.5	10.5	25.9	9.8
20.3	9.1	28.3	11	20.3	9.1	28.3	11
23.7	10	29	11.2	23.7	10	29	11.2
21.9	9.5	30.2	12	21.9	9.5	30.2	12
27.2	10.7	31	12.4	27.2	10.7	31	12.4
28.1	12.2	27.5	10.8	28.1	12.2	27.5	10.8
26.7	11.9	26.8	10.2	26.7	11.9	26.8	10.2
27.4	11.3	29.6	11.6	27.4	11.3	29.6	11.6
29.1	17.4	28.1	10.8	29.1	17.4	28.1	10.8
27.5	12.6	30	11.8	27.5	12.6	30	11.8
29	12.8	27.4	10.2	29	12.8	27.4	10.2
27.3	11.2	28.8	9.8	27.3	11.2	28.8	9.8
21.9	11.1	28.6	10.3	21.9	11.1	28.6	10.3
23.7	10.4	31.2	12.6	23.7	10.4	31.2	12.6
22.8	13.8	29.6	11.6	22.8	13.8	29.6	11.6
24.9	11	28.1	10.8	24.9	11	28.1	10.8
26.2	12.3	30	11.8	26.2	12.3	30	11.8