Research Article

Comparing the Linear and Quadratic Discriminant Analysis of Diabetes Disease Classification Based on Data Multicollinearity

Autcha Araveeporn [ID]

Department of Statistics, School of Science, King Mongkut's Institute of Technology Ladkrabang, Bangkok, Thailand

Correspondence should be addressed to Autcha Araveeporn; autcha.ar@kmitl.ac.th

Linear and quadratic discriminant analysis are two fundamental classification methods used in statistical learning. Moments (MM), maximum likelihood (ML), minimum volume ellipsoids (MVE), and t-distribution methods are used to estimate the parameter of independent variables on the multivariate normal distribution in order to classify binary dependent variables. The MM and ML methods are popular and effective methods that approximate the distribution parameter and use observed data. However, the MVE and t-distribution methods focus on the resampling algorithm, a reliable tool for high resistance. This paper starts by explaining the concepts of linear and quadratic discriminant analysis and then presents the four other methods used to create the decision boundary. Our simulation study generated the independent variables by setting the coefficient correlation via multivariate normal distribution or multicollinearity, often through basic logistic regression used to construct the binary dependent variable. For application to Pima Indian diabetic dataset, we expressed the classification of diabetes as the dependent variable and used a dataset of eight independent variables. This paper aimed to determine the highest average percentage of accuracy. Our results showed that the MM and ML methods successfully used large independent variables for linear discriminant analysis (LDA). However, the t-distribution method of quadratic discriminant analysis (QDA) performed better when using small independent variables.

1. Introduction

Logistic regression is a standard statistical algorithm used to classify binary dependent Logistic regression is a standard statistical algorithm to classify binary dependent variables based on independent variables. Nevertheless, the critical assumption of logistic regression is that there is a linear decision surface between the dependent and independent variables. Furthermore, logistic regression requires an average amount of or no multicollinearity problems between independent variables. Linearly separable data and multicollinearity in data are often found in real-world situations. It is hard to determine complicated relationships using logistic regression. More powerful and concise algorithms such as discriminant analysis [1] can easily exceed the performance of this algorithm.

Discriminant analysis is used for classification, dimension, and data visualization. The classification was explored in a study of discriminant analysis [2] that has been used in many classification problems [3, 4]. When performing discriminant analysis, users can discuss classification methods in which two or more groups and one or more independent variables are placed into one of the measured characteristics. Medical scientists investigate how groups (characterized by blood pressure, blood glucose levels, and age) differ across independent variables. One study used discriminant analysis to determine which patients had previously suffered a heart attack [5] to classify whether the patient would survive based on other variables. Two discriminant analyses are interesting: linear discriminant analysis (LDA) and quadratic discriminant analysis (QDA).

LDA can be interpreted from decision boundaries at two points. The first is a probabilistic interpretation, and the second is an interpretation of Fisher [6]. The first interpretation helps consider the assumptions of LDA. The

second interpretation provides a better understanding of how LDA performs classification. Fisher [7] introduced an LDA that identified a linear combination of the independent variables with a maximum class separation ratio. LDA's ability to successfully distinguish between tumor classes [8] using gene expression data was an essential feature of a new approach to cancer classification. LDA assumes that the variance between all classes is equal, and the decision boundary is calculated in a linear function. In other words, when each class has an individual covariance matrix, this is QDA, and the decision boundary shows in the form of a quadratic function. The LDA and QDA [9] should be recommended when using NIR data to classify an ill-posed problem. The principal component analysis [10] was applied with LDA and QDA for discriminants between healthy control and cancer samples based on MS data sets. The seven classification methods [11] were tested with binary logistic, probit, and cumulative probit regression, LDA, QDA, artificial neural networks, and naïve Bayes classification, to examine skeletal sex estimation. Consequently, LDA may be more preferred in skeletal sex estimation than other methods. LDA and QDA proposed flow regime identification [12] to combine responses from a non-intrusive optical sensor for air and water's vertical upward gas-liquid flow.

The LDA and QDA methods use the multivariate normal distribution of independent variables as a classification rule. The parameter of prior probability, mean, and covariance matrix of each class also create the discriminant function for the boundary of classes. The two classes are assumed to have the normal distribution, the most common and default distribution in real-world applications. Therefore, if the normal distribution is considered for the two classes where the covariance matrices are assumed to be equal, the decision boundary of classification is in the form of a linear discriminant function. When the covariance matrices are assumed to be unequal, the decision boundary of classification is in the form of a quadratic discriminant function.

This study aimed to investigate the parameter estimations for LDA and QDA. We proposed commonly-used methods, namely the moments (MM) and maximum likelihood (ML) methods, and the iterative algorithm, namely the minimum value ellipsoid (MVE) and t-distribution methods, to determine which conditional distribution was from the multivariate normal distribution.

Traditionally, LDA and QDA are estimated using the first and the second theoretical moments, also called the MM method introduced by Bowman and Shenton [13] in their concept of the central limit theorem. This method estimated population parameters that expressed population moments as the parameter of interest in terms of the functions. These functions were then set equal to the sample moments, and the number of equations was the same as the number of parameters. The equations were then used to estimate the parameters of interest before the estimators were solved to approximate these parameters.

The ML method [14] was a technique that was widely used for estimating probability distribution function parameters based on the observed data. The parameters were estimated by maximizing the likelihood function of observation. The logarithm was then based on the likelihood function. Taking the derivative with respect to a parameter was straightforward, and when set to zero, the ML estimators produced a pleasant result.

Rousseeuw [15, 16] studied the resampling algorithm using the minimum volume ellipsoid method. This method generated a robust estimation of multivariate location and scale. It estimated a low bias and expedient for outlier inspection via multivariate data, often through the benefit of robust distances on the MVE estimate. The MVE properties included breakdown value, affine equivarance, and efficiency.

The suitable t-distribution method [17] used the statistical modeling finite mixtures model to classify multivariate data sets. Its advantages were that it was the most comfortable method for modeling, and it obtained increasing determination. Multivariate normal data had been used because of their easy computation. Modeling that uses a mixture of t-distribution methods can be fitted using an ML iteration [18] expectation-maximization (EM) algorithm.

The debate on using machine learning or discriminant analysis (LDA and QDA) has been one of the critical questions in classification. Generally speaking, the discriminative classifiers model directly depends on the probability distribution function and thus needs fewer parameters; this alleviates the difficulty of parameter estimation problems. Machine learning is concerned with the algorithms that learn how to assign a class label to an example from the problem domain. On the other hand, because class probability distributions are known, LDA and QDA have unique advantages in dealing with multi-collinearity data. The ML, MM, MVE, and t-distribution methods are approximate mean and covariance from the multivariate normal distribution that plays important role in classification.

This study compares the MM, ML, MVE, and t-distribution methods depending on the LDA and QDA. These methods were used to classify the binary dependent variables related to the multicollinearity of independent variables. Their percentage of accuracy determined the performance of the four methods. Additionally, we clarified some of the simulation study's theoretical concepts and real data.

2. The Concept of Linear and Quadratic Discriminant Methods

The probability density function for multivariate normal distribution, or $x \sim N(\mu, \Sigma)$, is written as:

$$f(x|\mu, \Sigma) = \frac{1}{\sqrt{(2\pi)^p |\Sigma|}} \exp\left(-\frac{(x-\mu)^T \Sigma^{-1}(x-\mu)}{2}\right), \quad (1)$$

where $x = (x_1, x_2, ..., x_p)$ is the independent variable, $\mu = (\mu_1, \mu_2, ..., \mu_p)$ is the mean of the independent variables, and Σ is the covariance matrix.

2.1. Linear Doscriminant Analysis (LDA). LDA [19] assumes that the binary classification has the equal covariance matrix $\Sigma_1 = \Sigma_2 = \Sigma$. Therefore, the equation (1) becomes:

$$\frac{1}{\sqrt{(2\pi)^p|\Sigma|}}\exp\left(-\frac{(x-\mu_1)^T\Sigma^{-1}(x-\mu_1)}{2}\right)\pi_1$$

$$=\frac{1}{\sqrt{(2\pi)^p|\Sigma|}}\exp\left(-\frac{(x-\mu_2)^T\Sigma^{-1}(x-\mu_2)}{2}\right)\pi_2,$$

$$\exp\left(-\frac{(x-\mu_1)^T\Sigma^{-1}(x-\mu_1)}{2}\right)\pi_1$$

$$=\exp\left(-\frac{(x-\mu_2)^T\Sigma^{-1}(x-\mu_2)}{2}\right)\pi_2, \tag{2}$$

where π_1 and π_2 are the prior probability of the two classes and μ_1 and μ_2 are the mean of the two classes.

Using the natural logarithm from two sides of the equation, we get the simplified term:

$$-\frac{1}{2}(x-\mu_1)^T\overset{-1}{\Sigma}(x-\mu_1)+\ln(\pi_1)=$$

$$-\frac{1}{2}(x-\mu_2)^T\overset{-1}{\Sigma}(x-\mu_2)+\ln(\pi_2),$$

$$-\frac{1}{2}x^T\overset{-1}{\Sigma}x-\frac{1}{2}\mu_1^T\overset{-1}{\Sigma}\mu_1+\mu_1^T\overset{-1}{\Sigma}x+\ln(\pi_1)$$

$$=-\frac{1}{2}x^T\overset{-1}{\Sigma}x-\frac{1}{2}\mu_2^T\overset{-1}{\Sigma}\mu_2+\mu_2^T\overset{-1}{\Sigma}x+\ln(\pi_2), \tag{3}$$

where equation (3) is $x^T\Sigma^{-1}\mu_1 = \mu_1^T\Sigma^{-1}x$, and when each side is multiplied by 2, we get:

$$2\left(\overset{-1}{\Sigma}(\mu_2-\mu_1)\right)^T x+\left((\mu_2-\mu_1)^T\overset{-1}{\Sigma}(\mu_2-\mu_1)\right)$$

$$+2\ln\left(\frac{\pi_2}{\pi_1}\right)=0. \tag{4}$$

Equation (4) shows the form of a linear term. LDA discriminant function the two classes as:

$$\delta(x)=2\left(\overset{-1}{\Sigma}(\mu_2-\mu_1)\right)^T x+\left((\mu_2-\mu_1)^T\overset{-1}{\Sigma}(\mu_2-\mu_1)\right)$$

$$+2\ln\left(\frac{\pi_2}{\pi_1}\right)=0. \tag{5}$$

The class of instance x is defined by:

$$\hat{c}(x)=\begin{cases}1, \text{if } \delta(x)<0,\\ 2, \text{if } \delta(x)>0.\end{cases} \tag{6}$$

The LDA Algorithm.

(1) Input independent variables $x = (x_1, x_2, ..., x_p)$ of n sample.

(2) Compute the mean of each class μ_1 and μ_2.

(3) Calculate the prior probability of each class π_1 and π_2.

(4) Compute the covariance matrix for each class Σ_1 and Σ_2.

(5) Approximate the pooled covariance matrix as $\hat{\Sigma} = (n_1 - 1)\hat{\Sigma}_1 + (n_2 - 1)\hat{\Sigma}_2/n - 2, n = n_1 + n_2$.

(6) Calculate the LDA discriminant function as in equation (5).

(7) Assign the class label as in as in equation (6).

2.2. Quadratic Doscriminant Analysis (QDA). QDA [19] for binary classification is defined as the unequal covariance matrix $\Sigma_1 \neq \Sigma_2$. Therefore, equation (2) can be adjusted in term of unequal covariance as:

$$\frac{1}{\sqrt{(2\pi)^p|\Sigma_1|}}\exp\left(-\frac{(x-\mu_1)^T\Sigma_1^{-1}(x-\mu_1)}{2}\right)\pi_1$$

$$=\frac{1}{\sqrt{(2\pi)^p|\Sigma_2|}}\exp\left(-\frac{(x-\mu_2)^T\Sigma_2^{-1}(x-\mu_2)}{2}\right)\pi_2. \tag{7}$$

The natural logarithm taken from equation (7) is:

$$-\frac{p}{2}\ln(2\pi)-\frac{1}{2}\ln\left(\left|\Sigma_1\right|\right)-\frac{1}{2}(x-\mu_1)^T\overset{-1}{\Sigma_1}(x-\mu_1)+\ln(\pi_1)=-\frac{p}{2}\ln(2\pi)-\frac{1}{2}\ln\left(\left|\Sigma_2\right|\right)-\frac{1}{2}(x-\mu_2)^T\overset{-1}{\Sigma_2}(x-\mu_2)+\ln(\pi_2),$$

$$-\frac{1}{2}\ln\left(\left|\Sigma_1\right|\right)-\frac{1}{2}x^T\overset{-1}{\Sigma_1}x-\frac{1}{2}\mu_1^T\overset{-1}{\Sigma_1}\mu_1+\mu_1^T\overset{-1}{\Sigma_1}x+\ln(\pi_1)=-\frac{1}{2}\ln\left(\left|\Sigma_2\right|\right)-\frac{1}{2}x^T\overset{-1}{\Sigma_2}x-\frac{1}{2}\mu_2^T\overset{-1}{\Sigma_2}\mu_2+\mu_2^T\overset{-1}{\Sigma_2}x+\ln(\pi_2). \tag{8}$$

We multiplied the sides of the previous equation by 2 and got:

$$x^T \left(\sum_1 - \sum_2 \right)^{-1} x + 2 \left(\sum_2^{-1} \mu_2 - \sum_1^{-1} \mu_1 \right)^T x + \left(\mu_1^t \sum_1^{-1} \mu_1 - \mu_2^t \sum_2^{-1} \mu_2 \right) + \ln\left(\frac{|\Sigma_1|}{|\Sigma_2|} \right) + 2\ln\left(\frac{\pi_2}{\pi_1} \right) = 0. \tag{9}$$

To obtain Equation (8) in the quadratic form $x^T A x + b^T x + c = 0$, we brought the expression to the QDA discriminant function as:

$$
\begin{aligned}
\delta(x) := \; & x^T \left(\sum_1 - \sum_2 \right)^{-1} x + 2 \left(\sum_2^{-1} \mu_2 - \sum_1^{-1} \mu_1 \right)^T x \\
& + \left(\mu_1^t \sum_1^{-1} \mu_1 - \mu_2^t \sum_2^{-1} \mu_2 \right), \\
& + \ln\left(\frac{|\Sigma_1|}{|\Sigma_2|} \right) + 2\ln\left(\frac{\pi_2}{\pi_1} \right) = 0.
\end{aligned}
\tag{10}
$$

The classification term is shown in equation (6).
The QDA Algorithm.

(1) Input independent variables $\mathbf{x} = (x_1, x_2, ..., x_p)$ of n sample.
(2) Compute the mean of each class μ_1 and μ_2.
(3) Calculate the prior probability of each class π_1 and π_2.
(4) Compute the covariance matrix for each class Σ_1 and Σ_2.
(5) Calculate the QDA discriminant function as in equation (10).
(6) Assign the class label as in as in equation (6).

3. The Estimation Methods

The classification of these methods focused on the probability density function for multivariate normal distribution from equation (1), $x \sim N(\mu, \sum)$, and is written as:

$$f(x|\mu, \Sigma) = \frac{1}{\sqrt{(2\pi)^p |\sum|}} \exp\left(-\frac{1}{2}(x-\mu)^T \sum^{-1} (x-\mu) \right). \tag{11}$$

This section estimated parameters from multivariate normal distribution across the four methods.

3.1. Moments (MM) Method. The MM method relates the equivalent of sample and theoretical moments. Sample moments are denoted where $E(X^k)$ is the kth moment of the distribution for $k = 1, 2, ...,$ and $M_k = 1/n \sum_{i=1}^n X_i^k$ is the kth sample moment for $k = 1, 2, ...$

For multivariate normal distribution, the first sample moment approximates the origin by.
$M_1 = 1/n \sum_{i=1}^n x_i = \overline{x}$ to the first theoretical moment $E(X)$. Then, the second sample moment approximates the origin using $M_2 = 1/n \sum_{i=1}^n x_i^2 = 1/n \sum_{i=1}^n x^i x_i^T$ to the second theoretical moment $E(X^2)$. The first and the second theoretical moments approximate the origin by $E(X_i) = \mu$ and. $E(X_i^2) = \Sigma + \mu\mu^T$.

The first theoretical moments approximate the origin with the sample moment and the moment estimator of μ is:

$$E(X_i) = \hat{\mu}_{mm} = \frac{1}{n} \sum_{i=1}^n x_i. \tag{12}$$

The second theoretical moment approximates the origin with the sample moment and we obtain:

$$
\begin{aligned}
E(X_i^2) &= \Sigma + \mu\mu^T \\
&= \frac{1}{n} \sum_{i=1}^n x_i^2 \\
&= \frac{1}{n} \sum_{i=1}^n x_i x_i^T,
\end{aligned}
\tag{13}
$$

where $\hat{\mu}_{mm} = 1/n \sum_{i=1}^n x_i$, and the moments estimator of Σ is approximated as:

$$
\begin{aligned}
\hat{\Sigma}_{MM} &= \frac{1}{n} \sum_{i=1}^n x_i x_i^T - \hat{\mu}_{MM} \hat{\mu}_{MM}^T \\
&= \frac{1}{n} \sum_{i=1}^n (x_i - \hat{\mu}_{MM})(x_i - \hat{\mu}_{MM})^T.
\end{aligned}
\tag{14}
$$

3.2. Maximum Likelihood (ML) Method. From the probability density function, the likelihood function is created by

$$L(\mu, \Sigma) = (2\pi)^{-np/2} |\sum|^{-n/2} \exp\left(-\frac{1}{2} \sum_{i=1}^n (x_i - \mu)^T \sum^{-1} (x_i - \mu) \right), \tag{15}$$

where $x = (x_{i1}, x_{i2}, ..., x_{ip})$ and $\mu = (\mu_1, \mu_2)$.

The likelihood function is calculated using the logarithm to the log likelihood function:

$$l(\mu, \Sigma) = -\frac{np}{2} \log(2\pi) - \frac{n}{2} \log|\sum| - \frac{1}{2} \sum_{i=1}^n (x_i - \mu)^T \sum^{-1} (x_i - \mu). \tag{16}$$

Taking the derivative with respect to μ is straightforward:

$$\frac{\partial l(\mu, \Sigma)}{\partial \mu} = \sum_{i=1}^n (x_i - \mu)^T \sum^{-1}, \tag{17}$$

and setting to zero we obtain a pleasant result:

$$\hat{\mu}_{ML} = \frac{1}{n} \sum_{i=1}^n x_i. \tag{18}$$

The ML estimator is the mean in the sample mean. The covariance matrix is approximated by taking the derivative with respect to the matrix from Σ^{-1} using equation (16).

$$\frac{\partial l(\mu, \Sigma)}{\partial \sum^{-1}} = \frac{n}{2}\Sigma - \frac{1}{2}\sum_{i=1}^{n}(x_i - \mu)^T(x_i - \mu). \tag{19}$$

Finally, setting to zero yields the ML estimator:

$$\widehat{\Sigma}_{ML} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \widehat{\mu}_{ML})(x_i - \widehat{\mu}_{ML})^T. \tag{20}$$

It can be concluded that the ML and the MM estimators used the same equation and values.

The MM and ML Algorithm.

(1) Input independent variables $x = (x_1, x_2, ..., x_p)$ of n sample.

(2) Select the sample sizes of each class n_1 and n_2.

(3) Calculate the mean of each class $\widehat{\mu}_1 = \sum_{i=1}^{n_1} x_i$ and $\widehat{\mu}_2 = 1/n \sum_{i=1}^{n_2} x_i$

(4) Compute the covariance matrix for each class $\widehat{\Sigma}_1 = 1/n \sum_{i=1}^{n_1}(x_i - \widehat{\mu}_1)(x_i - \widehat{\mu}_1)^T$ and $\widehat{\Sigma}_2 = 1/n \sum_{i=1}^{n_2}(x_i - \widehat{\mu}_2)(x_i - \widehat{\mu}_2)^T$

3.3. Minimum Value Ellipsoid (MVE) Method. This method [20] depends on the MVE, which covers the least point h of the n observations. The multivariate data is defined as $x = (x_1, x_2, ..., x_p)$ with independent variables p and $x_i = (x_{i1}, ..., x_{ip})^T; i = 1, 2, ..., n$. Computing the MVE for a set of data x requires the examination of all $\binom{n}{h}$ ellipsoids that contain the observed $h = n + p + 1/2$ of x for the smallest ellipsoid volume. The MVE algorithms defined ellipsoids by including the subsets of observed $p + 1$ on data x. For each subset of size $p + 1$, indicated by $J = \{i_1, ..., i_{p+1}\} \subset \{1, ..., n\}$, the sample mean and sample covariance matrix were computed by:

$$\widehat{\mu}_{\text{MVE}(J)} = \frac{1}{p+1}\sum_{j=1}^{p+1} x_{i_j} \quad \widehat{\Sigma}_{\text{MVE}(J)} = \frac{1}{p+1}\sum_{j=1}^{p+1}\left(x_{i_j} - \overline{x}_J\right)\left(x_{i_j} - \overline{x}_J\right)^T. \tag{21}$$

The covariance matrix Σ_J is a nonsingular covariance matrix when the subset of $p + 1$ is in general rank. If the subset of $p + 1$ was not in general rank, then the observed data from x were included until a subset with the sample of nonsingular covariance matrix was achieved.

The MVE Algorithm.

(1) Input independent variables $x = (x_1, x_2, ..., x_p)$ of n sample.

(2) Compute the standard choice $h = n + p + 1/2$.

(3) Require examination of all $\binom{n}{h}$ ellipsoids.

(4) Search to ellipsoids determined by subsets consisting of $p + 1$ obsavations.

(5) Compute mean $\widehat{\mu}_{MVE(J)}$ and covariance $\widehat{\Sigma}_{MVE(J)}$.

3.4. The T-distribution Method. The modeling of a mixture of t-distributions [17] is considered a robust estimator that fits the data set with the ML via the expectation-maximization (EM) algorithm [18]. The t-distribution proposes a longer-tailed distribution that is closed to the normal distribution for large sample sizes.

The data set is denoted by $x = (x_1, x_2, ..., x_p)$ with p independent variables from sample sizes n. In a normal mixture model based on these data, each data set is assumed to be a comprehension of the random variables of dimensional vector p with the g-component normal mixture probability density function on data set x:

$$f(x; \Psi) = \sum_{j=1}^{g} \pi_j \phi(x; \mu_j, \Sigma_j), \tag{22}$$

where the proportions π_j are nonnegative and all proportions sum to one. Then

$$\phi(x; \mu_j, \Sigma_j) = (2\pi)^{-p/2}|\sum|^{-1/2} \times \exp\left\{-\frac{1}{2}(x - \mu_j)^T\sum_{j}^{-1}(x - \mu_j)\right\}, \tag{23}$$

denotes the p-variate multivariate normal probability density function with mean μ_i and covariance matrix $\sum_j (j = 1, ... g)$. Here $\Psi = (\pi_1, ..., \pi_{g-1}, \theta^T)^T$ where θ is composed of the components of the μ_j and the distinct components of the $\Sigma_j (j = 1, ..., g)$.

The t mixture model depending on the normal distribution is obtained with the location parameter μ, positive covariance matrix Σ, and v degrees of freedom,

$$f(x; \mu, \Sigma, v) = \frac{\Gamma(v + p/2)|\Sigma|^{-(1/2)}}{(\pi v)^{1/2}\Gamma(v/2)\{1 + \delta(x, \mu, \Sigma)/v\}^{1/2}(v + p)}, \tag{24}$$

where $\delta(x, \mu, \Sigma) = (x - \mu)^T\Sigma^{-1}(x - \mu)$.

The implementation of the EM algorithm for ML estimation in the case of a single component t-distribution has been discussed in MaLachlan et al. [21]. For the EM algorithm at the iterations $(k + 1)$ th, $\pi^{(k+1)}, \theta^{(k+1)}$, and $v^{(k+1)}$ can be computed independently of each other. To compute the weight sample mean and sample covariance matrix, we used:

$$\widehat{\mu}_j^{(k+1)} = \frac{\sum_{i=1}^{n} \tau_{ij}^{(k)} u_{ij}^{(k)} x_i}{\sum_{i=1}^{n} \tau_{ij}^{(k)} u_{ij}^{(k)}},$$

$$\widehat{\Sigma}_j^{(k+1)} = \frac{\sum_{i=1}^{n} \tau_{ij}^{(k)} u_{ij}^{(k)}\left(x_i - \mu_j^{(k+1)}\right)\left(x_i - \mu_j^{(k+1)}\right)^T}{\sum_{i=1}^{n} \tau_{ij}^{(k)}}. \tag{25}$$

Where $u_{ij}^{(k)} = v_j^k + p/v_j^k + \delta(x_i, \mu_j^{(k)}; \Sigma_j^{(k)})$ and $\tau_{ij}^{(k)} = \pi_j^k f(x_i; \mu_j^{(k+1)}, v_j^{(k+1)})/f(x_i; \Psi^{(k+1)})$.

It can be seen that $\widehat{\mu}_i^{(k+1)}$ and $\widehat{\Sigma}_j^{(k+1)}$ were selected after weighted least squares estimation. The E-step drew the

weights $u_{ij}^{(k)}$ while the M-step successfully chose $\hat{\mu}_i^{(k+1)}$ and $\hat{\Sigma}_j^{(k+1)}$ using the weighted least squares method, which we called the EM algorithm.

The t-distribution Algorithm.

(1) Input independent variables $x = (x_1, x_2, ..., x_p)$ of n sample.

(2) Assume $x = (x_1, x_2, ..., x_p)$ to be the g-component normal mixture probability density function.

(3) Obtain the parameter mean and covariance to t-distribution.

(4) Implement EM algorithm at the iterations $(k + 1)$ th, $\pi^{(k+1)}, \theta^{(k+1)}$, and $\nu^{(k+1)}$.

(5) Compute the weight sample mean and sample covariance matrix as $\hat{\mu}_j^{(k+1)}$ and $\hat{\Sigma}_j^{(k+1)}$.

4. Simulation and Results

The objective of this study was to classify the binary dependent variables using LDA and QDA via MM, ML, MVE, and t-distribution methods. The independent variables (x) were simulated from the multivariate normal distribution in two, four, six, and eight sets of independent variables and constant correlation (ρ) values of 0.1, 0.5, and 0.9. The multivariate normal distribution function of the independent variables (x) consisted of mean (μ) and covariance matrix (Σ):

$$f(x|\mu, \Sigma) = \frac{1}{\sqrt{(2\pi)^p |\Sigma|}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)\right), \quad (26)$$

where

$$x = \begin{bmatrix} x_{i1} \\ x_{i2} \\ ... \\ x_{ip} \end{bmatrix}, \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \\ ... \\ \mu_p \end{bmatrix}, \Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 & ... & \rho\sigma_1\sigma_p \\ \rho\sigma_2\sigma_1 & \sigma_2^2 & ... & \rho\sigma_2\sigma_p \\ ... & ... & ... & ... \\ \rho\sigma_p\sigma_1 & \rho\sigma_p\sigma_2 & ... & \sigma_p^2 \end{bmatrix},$$

$$p = 2, 4, 6, 8, i = 1, ..., n. \quad (27)$$

The mean (μ) was fixed to generate at zero and the standard deviation $(\sigma_i, i = 1, 2, ..., p)$ was fixed to generate as 6. The regression coefficients were denoted by $\beta = (\beta_0, \beta_1, ..., \beta_p)^T$ for the set of two, four, six, and eight independent variables. Finally, the dependent variables (y) were approximated by the logit function $\pi_i = 1/1 + e^{-x\beta}$ from the logistic regression model.

If $\pi_i \geq 0.5$, the dependent variables were defined as $y_i = 1$, and $y_i = 0$, when $\pi_i < 0.5$. The R program employed simulation data at 1,000 replications with sample sizes of 200, 300, 400, and 500. The MM, ML, MVE, and t-distribution methods approximated LDA and QDA parameters to predict the dependent variable. The confusion matrix was used to report the classification performance on a set of estimated data to compare with the real data using the percentage of accuracy (Table 1).

Table 1: The confusion matrix of real data and estimated data.

Estimated data	Real data	
	$y_i = 1$	$y_i = 0$
$\hat{y}_{ii} = 1$	True positive (TP)	False positive (FP)
$\hat{y}_i = 0$	False negative (FN)	True negative (TN)

The percentage of accuracy was computed as (Table 1): Percentage of Accuracy = TP + TN/TP + TN + FP + FN × 100.

The average percentage of accuracy of the MM, ML, MVE, and t-distribution methods for two, four, six and, eight variables are presented in Tables 2–5 according to their correlation coefficient (0.1, 0.5, or 0.9). The mean of over 1,000 replications was used to compute the average percentage of accuracy. The maximum average percentage of accuracy values are shown in bold letters.

Tables 2–5 present the average accuracy of all methods for several correlation coefficients. Based on the results, it could be seen that the average percentage of accuracy for the MM and ML methods had both the highest and equal values across all tables. The MM and ML methods in QDA also showed the highest average percentage of accuracy in Table 2. Additionally, the MM and ML methods in LDA had the highest average percentage in Tables 3–5. Meanwhile, the t-distribution method had the second highest average percentage of accuracy and a minor difference from the MM and ML methods. When the independent variable increased, the average percentage of accuracy decreased. When the sample sizes were large, the average percentage of accuracy was good. However, when the correlation coefficients increased, the average percentage of accuracy was shown to be a slightly different value.

5. Application of a Real Dataset

We applied LDA and QDA to classify a healthy person and a diabetic patient from the Pima Aboriginals diabetes dataset from the Applied Physics Laboratory of John Hopkins University. This dataset was obtained from the University of California, Irvine website (https://archive.ics.uci.edu/ml/datasets/). The binary dependent variables were 0 or 1, with 0 indicating a healthy person and 1 indicating a diabetic patient. The independent variables were defined by the number of pregnancies (x_1), years of age (x_2), diabetes pedigree function (x_3), triceps skin fold thickness (x_4), 2-h serum insulin (x_5), plasma glucose concentration (x_6), diastolic blood pressure (x_7), and body mass index (kg/m2) (x_8). These data consisted of 768 records: 500 healthy person records and 268 diabetic patient records. Table 6 presents the descriptive statistics related to the diabetic disease dataset.

Pearson's correlation analysis was used to determine whether there was a relationship between the eight continuous independent variables. The formula for computing the correlation between two variables was:

$$r = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\left[\sum_{i=1}^{n}(x_i - \overline{x})^2\right]\left[\sum_{i=1}^{n}(y_i - \overline{y})^2\right]}}. \quad (28)$$

TABLE 2: The average percentage of accuracy of moments (MM), maximum likelihood (ML), minimum value ellipsoid (MVE), and t-distribution (t) methods under two independent variables.

Sample sizes (n)	Correlation Coefficient	Linear discriminant analysis				Quadratic discriminant analysis			
		MM	ML	MVE	t	MM	ML	MVE	t
200	0.1	98.15	98.15	97.61	97.92	98.18	98.18	98.00	98.12
	0.5	98.12	98.12	97.70	97.90	98.27	98.27	98.00	98.21
	0.9	98.15	98.15	97.72	97.93	98.22	98.22	98.05	98.21
300	0.1	98.47	98.47	98.10	98.31	98.55	98.55	98.36	98.48
	0.5	98.45	98.45	98.07	98.29	98.50	98.50	98.39	98.49
	0.9	98.37	98.38	98.05	98.21	98.45	98.46	98.43	98.43
400	0.1	98.58	98.58	98.28	98.49	98.66	98.66	98.64	98.67
	0.5	98.67	98.67	98.28	98.51	98.69	98.69	98.62	98.66
	0.9	98.60	98.60	98.30	98.46	98.66	98.66	98.65	98.65
500	0.1	98.75	98.75	98.40	98.61	98.78	98.78	98.74	98.74
	0.5	98.72	98.72	98.37	98.58	98.76	98.76	98.73	98.73
	0.9	98.73	98.73	98.39	98.60	98.81	98.81	98.73	98.71

TABLE 3: The average percentage of accuracy of moments (MM), maximum likelihood (ML), minimum value ellipsoid (MVE), and t-distribution (t) methods under four independent variables.

Sample sizes (n)	Correlation Coefficient	Linear discriminant analysis				Quadratic discriminant analysis			
		MM	ML	MVE	t	MM	ML	MVE	t
200	0.1	97.59	97.59	97.32	97.29	97.49	97.49	96.75	97.27
	0.5	97.60	97.60	97.28	97.40	97.49	97.60	97.28	97.36
	0.9	97.65	97.65	97.33	97.39	97.50	97.50	96.51	97.32
300	0.1	97.95	97.95	97.75	97.74	97.83	97.83	97.38	97.69
	0.5	97.97	97.97	97.65	97.73	97.81	97.81	97.42	97.66
	0.9	97.94	97.94	97.71	97.74	97.81	97.82	97.41	97.61
400	0.1	98.10	98.10	98.91	97.90	97.95	97.96	97.77	97.81
	0.5	98.10	98.10	98.96	97.93	97.93	97.94	97.72	97.84
	0.9	98.21	98.21	98.00	98.05	98.05	98.06	97.83	97.91
500	0.1	98.32	98.32	98.09	98.11	98.16	98.17	97.99	98.05
	0.5	98.31	98.31	98.09	98.13	98.16	98.16	97.93	98.03
	0.9	98.32	98.32	98.16	98.13	98.19	98.19	97.96	98.05

TABLE 4: The average percentage of accuracy of moments (MM), maximum likelihood (ML), minimum value ellipsoid (MVE), and t-distribution (t) methods under six independent variables.

Sample sizes (n)	Correlation Coefficient	Linear discriminant analysis				Quadratic discriminant analysis			
		MM	ML	MVE	t	MM	ML	MVE	t
200	0.1	97.27	97.27	97.01	97.00	96.98	96.98	95.76	96.74
	0.5	97.26	97.26	97.03	96.95	97.07	97.08	96.00	96.80
	0.9	97.30	97.30	97.04	97.04	97.17	97.18	96.06	96.84
300	0.1	97.59	97.59	97.47	97.46	97.28	97.29	96.71	97.10
	0.5	97.61	97.61	97.39	97.39	97.32	97.32	96.63	97.09
	0.9	97.88	97.88	97.37	97.37	97.23	97.23	96.60	97.03
400	0.1	97.88	97.88	97.65	97.65	97.46	97.46	96.99	97.28
	0.5	97.84	97.84	97.60	97.60	97.45	97.46	97.00	97.26
	0.9	97.84	97.85	97.66	97.66	97.51	97.51	97.04	97.30
500	0.1	98.02	98.02	97.83	97.83	97.67	97.67	97.30	97.49
	0.5	98.02	98.02	97.83	97.83	97.66	97.67	97.30	97.46
	0.9	98.03	98.03	97.82	97.82	97.69	97.69	97.28	97.50

The correlation coefficients of the independent variables are presented in Table 7 and Figure 1. The hypothesis testing for the significance of the relationship is defined as:

$H_0: \rho = 0$, $H_1: \rho \neq 0$, and the t-statistics were used for computing the significance of Pearson's correlation by $t = r\sqrt{n-2/1-r^2}$ with a degree of freedom (df) n-2.

Finally, when the p-value of the t-statistics was less than 0.05, that indicated a significant relationship between two variables shown in Table 7.

Based on our findings, a significant positive relationship at a moderate level was found in most cases such as in x_1, x_6 and x_3 to x_8, and a significant relationship at a strong level

TABLE 5: The average percentage of accuracy of moments (MM), maximum likelihood (ML), minimum value ellipsoid (MVE), and t-distribution (t) methods under eight independent variables.

| Sample sizes (n)	Correlation Coefficient	Linear discriminant analysis				Quadratic discriminant analysis			
		MM	ML	MVE	t	MM	ML	MVE	t
200	0.1	97.11	97.11	96.72	96.82	96.85	96.85	95.12	96.41
	0.5	97.04	97.04	96.72	96.74	96.81	96.81	95.14	96.45
	0.9	97.13	97.14	96.78	96.87	96.90	96.90	95.18	96.47
300	0.1	97.38	97.38	97.10	97.07	96.92	96.92	95.99	96.65
	0.5	97.28	97.28	97.13	97.05	96.91	96.92	95.91	96.60
	0.9	97.32	97.32	97.12	97.06	96.89	96.90	95.88	96.61
400	0.1	97.62	97.62	97.17	97.34	97.17	97.17	96.47	96.91
	0.5	97.54	97.54	97.38	97.30	97.10	97.10	96.37	96.83
	0.9	97.57	97.57	97.40	97.35	97.16	97.16	96.50	96.90
500	0.1	97.74	97.74	97.56	97.50	97.27	97.27	96.79	97.30
	0.5	97.74	97.74	97.55	97.49	97.31	97.31	96.78	97.07
	0.9	97.77	97.77	97.58	97.60	97.32	97.32	96.77	97.08
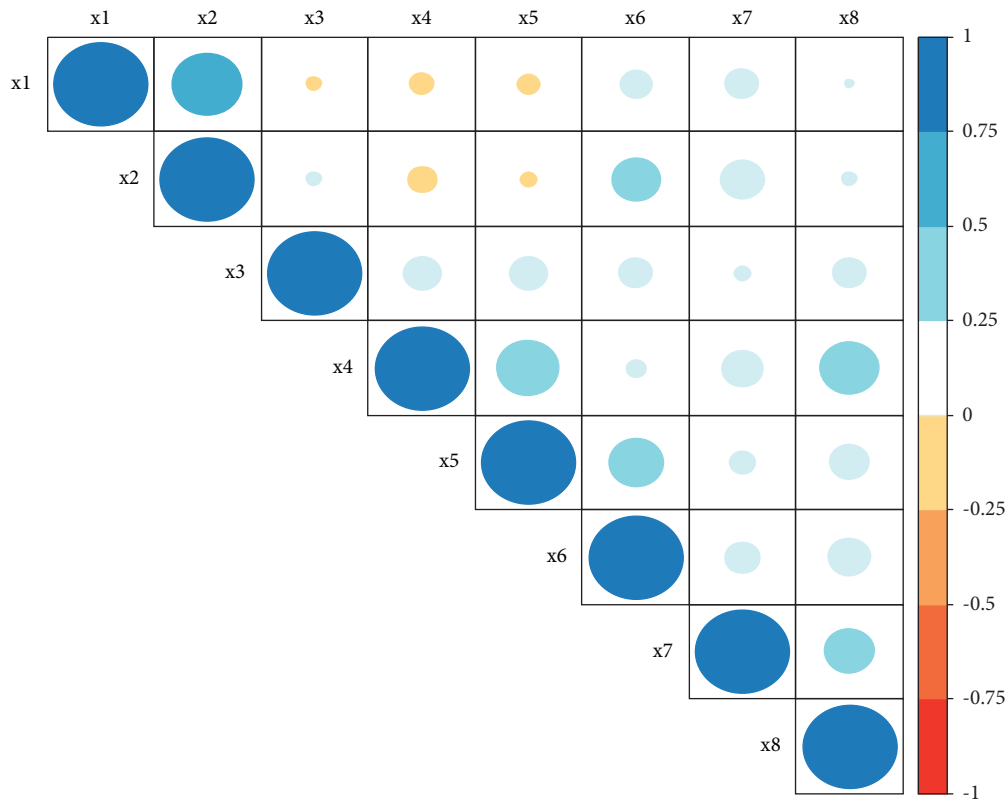
FIGURE 1: The graph of the correlation matrix between eight independent variables.

was shown in some cases such as x_1, x_2 and x_4, x_5. Most of the independent variables had a significant relationship except for x_2, x_3, x_2, x_4, and x_2, x_5.

The Pearson correlation matrix from Table 7 is rewritten in Figure 1, and is easier to understand using different shades. The light shade signifies a moderate correlation, and the dark shade denotes a strong correlation. Most of the independent variables have the light shade, which means there were either correlations among the independent variables or multicollinearity problems.

The MM, ML, MVE, and t-distribution methods via LDA and QDA were approximated to compute the accuracy percentages in Table 8. The sets of two, four, six, or eight independent variables were similar to those of the simulation data. The two, four, and six independent variables were selected when the correlations were at the significant 0.05 level.

From Table 8, it is apparent that the MM and ML methods showed equal accuracy percentages across all cases and the highest accuracy percentages in most cases.
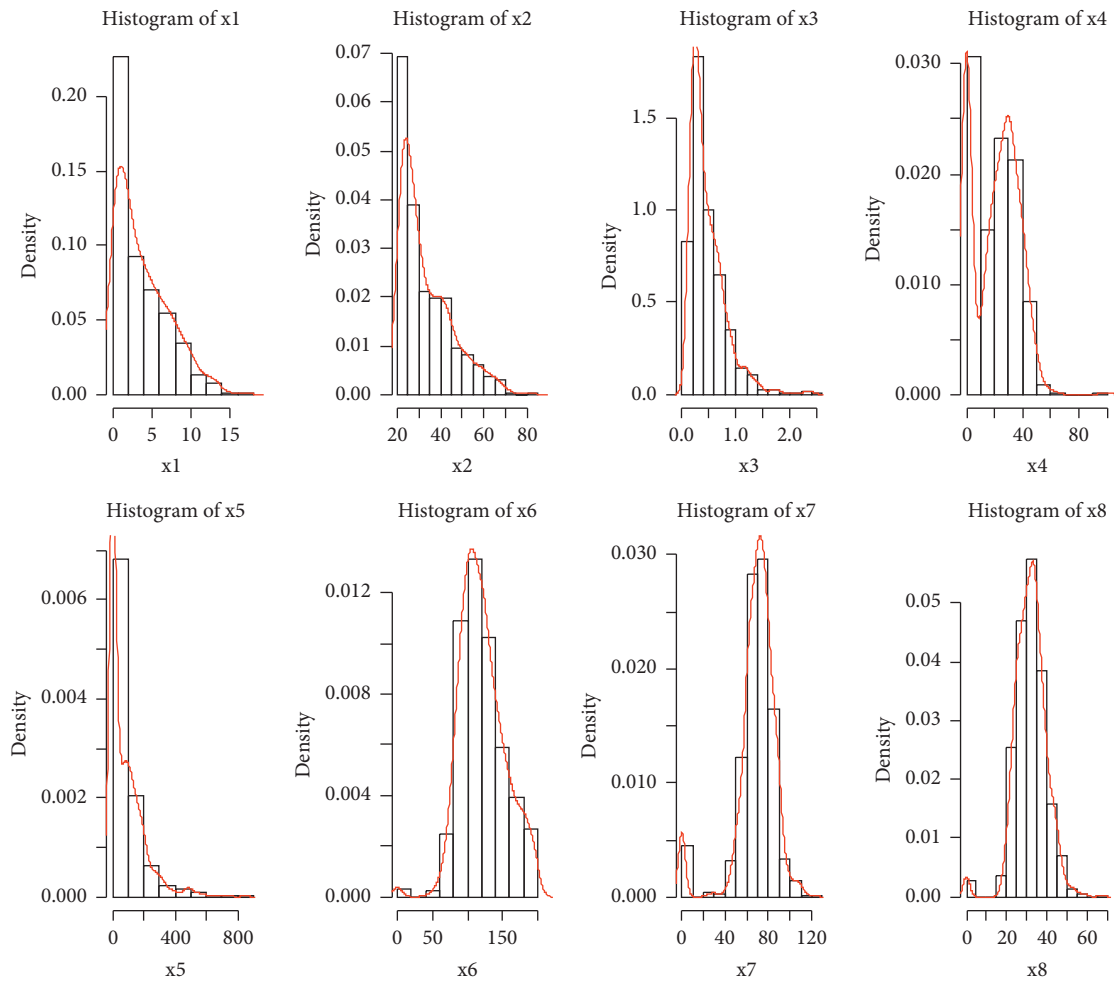
FIGURE 2: The histogram of eight independent variables.

TABLE 6: Descriptive statistics of the diabetic patient dataset.

Variables	Mean	Median	Std. Dev.	Min	Max
x_1	3.84	3.00	3.36	0.00	17.00
x_2	33.24	29.00	11.76	21.00	81.00
x_3	0.47	0.37	0.33	0.07	2.42
x_4	20.54	23.00	15.95	0.00	99.00
x_5	79.80	30.5	115.24	0.00	846.00
x_6	120.90	117.00	31.97	0.00	199.00
x_7	69.10	72.00	19.35	0.00	122.00
x_8	31.99	32.00	7.88	0.00	67.10

TABLE 7: Pearson correlation matrix for the relationship between eight independent variables.

Variables	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
x_1	1.00	0.54*	−0.03	−0.08	−0.07	0.13*	0.14-->	0.02
p-value		<0.05	1.00	0.21	0.33	<0.05	<0.05	1.00
x_2	—	1.00	0.03	−0.11*	−0.04	0.26	0.23*	0.03
p-value			1.00	0.02	1.00	<0.05	<0.05	1.00
x_3	—	—	1.000	0.18*	0.18*	0.13*	0.04	0.14*
p-value				<0.05	<0.05	<0.05	1.00	<0.05
x_4	—	—	—	1.000	0.43*	0.05	0.20*	0.39*
p-value					<0.05	0.79	<0.05	<0.05
x_5	—	—	—	—	1.000	0.33*	0.08	0.19*
p-value						<0.05	0.14	<0.05
x_6	—	—	—	—	—	1.000	0.15*	0.22*
p-value							<0.05	<0.05
x_7	—	—	—	—	—	—	1.000	0.28*
p-value								<0.05
x_8	—	—	—	—	—	—	—	1.000
p-value								

Note. *, The significant correlation at 0.05 level.

However, the accuracy percentage of the t-distribution method was good in two independent variables using QDA. Therefore, the MM and ML methods outperformed with LDA for four, six, and eight independent variables, which was the same as the simulation results. When the number of independent variables increased for two or four, the accuracy percentage was slightly different. However, the large independent variables demonstrated excellent performances in the diabetes dataset.

TABLE 8: The percentage of prediction accuracy on two, four, six, and eight independent variables.

Number of independent variables	Independent variables	Linear discriminant analysis				Quadratic discriminant analysis			
		MM	ML	MVE	t	MM	ML	MVE	t
2	x_1, x_2	66.53	66.53	65.10	66.53	65.88	65.88	65.75	67.44
	x_1, x_7	66.53	66.53	64.97	66.53	65.88	65.88	65.75	67.44
	x_3, x_4	66.53	66.53	65.75	66.53	65.88	65.88	65.75	67.44
	x_5, x_8	66.53	66.53	65.75	66.53	65.88	65.88	65.49	67.44
	x_6, x_7	66.53	66.53	65.49	66.53	65.88	65.88	65.62	67.44
	x_7, x_8	66.53	66.53	65.10	66.53	65.88	65.88	65.75	67.44
4	x_1, x_2, x_4, x_8	67.96	67.96	66.01	67.44	67.31	67.31	65.75	66.92
	x_2, x_5, x_6, x_8	67.96	67.96	67.05	67.44	67.31	67.31	66.40	66.92
	x_2, x_3, x_4, x_8	67.96	67.96	66.14	67.44	67.31	67.31	66.53	66.92
	x_3, x_5, x_7, x_8	67.96	67.96	65.57	67.44	67.31	67.31	66.01	66.92
6	$x_2, x_3, x_5, x_6, x_7, x_8$	76.51	76.51	74.47	76.04	76.43	76.43	70.44	72.39
	$x_1, x_2, x_3, x_4, x_6, x_7$	76.52	76.51	75.00	76.04	76.43	76.43	72.00	72.39
8	$x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8$	78.38	78.38	77.77	77.21	76.43	76.56	74.21	73.3

6. Discussion

The simulated results are listed in Tables 2–5. The number of independent variables and sample sizes influenced the average accuracy percentage. QDA outperformed on the small independent variables, but the large independent variables were more appropriate for LDA. However, the average accuracy percentages of the large independent variables were less than the percentages for the small independent variables. The correlation coefficient increase did not affect the classification because the average accuracy percentage showed slightly different values. If the sample size increased, the accuracy of all methods increased in all cases.

For the real data results in Table 8, the MM, ML, and t-distribution methods showed the highest accuracy percentages for two independent variables. We found that the independent variables of real data demonstrated skewed data (Figure 2). The Shapiro-Wilk test [22] was used to confirm that all independent variables showed non-normality. However, the MM and ML methods supported classification on large independent variables. The multivariate normal distribution generated the simulation dataset. Therefore, the real data results were different from the simulation study in some cases. Overall, it is clear that the MM and ML methods are most suitable for classifying diabetes.

The Pima Indian diabetic dataset is popular to analyze for classification and to improve medical diagnosis. Gupta et al. [23] compared Gaussian process, LDA, QDA, statistical gradient descent, ridge regression classifier, support vector machines, k-nearest neighbors, decision tree, naïve Bayes, logistic regression, random forest, and ELM for multiquadric, RBF, sigmoid activation functions. The results suggested that the logistic regression performs better than the other techniques. Chang et al. [24] proposed an e-diagnosis system based on a machine learning algorithm: Naïve Bayes classifier, random forest classifier, and J48 decision tree models to be trained and tested using the Pima Indians diabetes. It can be concluded that a Naïve Bayes model works well with a more fine-tuned selection of features for binary classification. Abedin et al. [25] studied a hierarchical ensemble model to combine two classifiers that had been trained, a decision tree and a logistic regression model, and feed the output of those models to a neural network. The proposed model achieved classification accuracy by using PIMA Indian diabetes database.

7. Conclusion

This paper proposes the classification of binary data by applying MM, ML, MVE, and t-distribution methods based on LDA and QDA for data multicollinearity. The solutions provide the advantages and disadvantages of these methods. The advantage of the QDA results when using small independent variables, MM and ML, outperformed in classifying all correlation coefficients, but these methods are slightly different from the other methods. However, the LDA results showed that MM and ML perform the same when provided with four to eight independent variables. These methods did not relate to the classification performance when considering the correlation coefficient. However, the large sample sizes showed good performances across all cases. Eight independent variables were selected for the medical data when using real data to classify diabetes patients. We selected two, four, six, and eight independent variables for large correlation. These results showed that the t-distribution was efficient on two independent variables via QDA. However, the MM and ML methods were effective at classifying four to eight variables based on LDA. Therefore, we concluded that MM and ML methods could classify in the presence of multicollinearity. Medical data classification has been extended to study a novel random vector functional link [26] and a novel random vector functional link with ε-insensitive Huber loss function [27]. A fuzzy-based Lagrangian twin parametric-margin support vector machine [28] reduced the effect of the outliers in medical data.

Data Availability

The data used to support the findings of this study are available from the author on request.

Conflicts of Interest

Acknowledgments

References

[1] D. C. McLachlan, *Discriminant Analysis and Statistical Pattern Recognition*, John Wiley & Sons, New York, US, 2004.

[2] S. Shinmura, *New Theory of Discriminant Analysis after R. Fisher Advanced Research by the Feature-Selection Method for Microarray Data*, Springer, Singapore, 2016.

[3] E. I. Altman, G. Marco, and F. Varetto, "Corporate distress diagnosis: c," *Journal of Banking & Finance*, vol. 18, no. 3, pp. 505–529, 1994.

[4] Y. Guo, T. Hastie, and R. Tibshirani, "Regularized linear discriminant analysis and its application in microarrays," *Biostatistics*, vol. 8, no. 1, pp. 86–100, 2007.

[5] M. A. Fernandez, C. Rueda, and B. Salvador, "Incorporating additional information to normal linear discriminant rules," *Journal of the American Statistical Association*, vol. 101, no. 474, pp. 569–577, 2006.

[6] R. A. Fisher, *The Precision of Discriminant Functions*, Cambridge University Pres, London, UK, 1940.

[7] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Annals of Eugenics (London)*, vol. 7, no. 2, pp. 179–188, 1936.

[8] S. Dudoit, J. Fridlyand, and T. P. Speed, "Comparison of discrimination methods for the classification of tumors using gene expression data," *Journal of the American Statistical Association*, vol. 97, no. 457, pp. 77–87, 2002.

[9] W. Wu, Y. Mallet, B. Walczak et al., "Comparison of regularized discriminant analysis linear discriminant analysis and quadratic discriminant analysis applied to NIR data," *Analytica Chimica Acta*, vol. 329, no. 3, pp. 257–265, 1996.

[10] C. L. Morrais and K. M. Lima, "Principle component analysis with linear and quadratic discriminant analysis for identification of cancer samples based on mass spectrometry," *Journal of the Brazilian Chemical Society*, vol. 29, no. 3, pp. 472–481, 2018.

[11] E. Nikita and P. Nikitas, "Sex estimation: a comparison of techniques based on binary logistic, probit and cumulative probit regression, linear and quadratic discriminant analysis, neural networks, and naïve Bayes classification using ordinal variables," *International Journal of Legal Medicine*, vol. 134, no. 3, pp. 1213–1225, 2020.

[12] K. Sarkodie and A. Fergusson-Rees, "Flow regime identification in vertical upward gas-liquid flow using an optical sensor with linear and quadratic discriminant analysis," *Journal of Fluids Engineering*, vol. 143, no. 2, pp. 1–12, 2021.

[13] K. O. Bowman and L. R. Shenton, *Encyclopedia of Statistical Sciences*, John Wiley & Sons, New York, US, 1998.

[14] R. J. Rossi, *Mathematical Statistics: An Introduction to Likelihood Based Inference*, John Wiley & Sons, New York, US, 2018.

[15] P. J. Rousseeuw, "Least median of squares regression," *Journal of the American Statistical Association*, vol. 79, no. 388, pp. 871–880, 1984.

[16] P. J. Rousseeuw, *Multivariate Estimation with High Break-Down point*, pp. 283–297, Mathematical Statistics and Application, 1985.

[17] D. Peel and G. J. McLachlan, "Robust mixture modelling using the t distribution," *Statistics and Computing*, vol. 10, no. 4, pp. 339–348, 2000.

[18] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the *EM* Algorithm," *Journal of the Royal Statistical Society: Series B*, vol. 39, pp. 1–22, 1977.

[19] B. Ghojogh and M. Crowley, "Linear and Quadratic Discriminant Analysis: Tutorial," 2019, https://arxiv.org/abs/1906.02590.

[20] S. Van Aelst and P. Rousseeuw, "Minimum volume ellipsoid," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 1, no. 1, pp. 71–82, 2009.

[21] G. J. McLachlan, D. Peel, K. E. Basford, and P. Adams, "The EMMIX software for the fitting of mixtures of normal and t-components," *Journal of Statistical Software*, vol. 4, no. 2, pp. 1–14, 1999.

[22] S. S. Shapiro and M. B. Wilk, "An analysis of variance test for normality (complete samples)," *Biometrika*, vol. 52, no. 3/4, pp. 591–611, 1965.

[23] D. Gupta, A. Choudhury, U. Gupta, P. Singh, and M. Prasad, "Computational approach to clinical diagnosis of diabetes disease: a comparative study," *Multimedia Tools and Applications*, vol. 80, no. 20, pp. 30091–30116, 2021.

[24] V. Chang, J. Bailey, Q. A. Xu, and Z. Sun, "Pima Indians diabetes mellitus classification based on machine learning (ML) algorithms," *Neural Computing & Applications*, pp. 1–17, 2022.

[25] M. Abedind and T. Banirostam, "Classification of Pima Indian diabetes dataset using ensemble of decision tree, logistic regression and neural network," *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 9, no. 7, pp. 1–4, 2022.

[26] B. B. Harzarika and D. Gupta, "1-Norm random vector functional link networks for classification problems," *Complex & Intelligent Systems*, vol. 308, pp. 1–17, 2022.

[27] B. B. Hazarika and D. Gupta, "Random vector functional link with ε-insensitive Huber loss function for biomedical data classification," *Computer Methods and Programs in Biomedicine*, vol. 215, Article ID 106622, 2022.

[28] D. Gupta, P. Borah, U. M. Sharma, and M. Prasad, "Data-driven mechanism based on fuzzy Lagrangian twin parametric-margin support vector machine for biomedical data analysis," *Neural Computing & Applications*, vol. 34, no. 14, pp. 11335–11345, 2022.