STATISTICS, OPTIMIZATION AND INFORMATION COMPUTING
Stat., Optim. Inf. Comput., Vol. **x**, Month 202x, pp 0–14.
Published online in International Academic Press (www.IAPress.org)

A Method to Classify Shape Data using Multinomial Logistic Regression Model

Meisam Moghimbeygi

Department of Mathematics, Faculty of Mathematics and Computer Science, Kharazmi University, Tehran, Iran

Abstract We introduced a multinomial logistic regression model to classify the labeled configurations. In this modeling, we use a power-divergence test to find an estimator for belonging probability in each category. The estimator is introduced based on different distances. Since the estimator is biased, we modified the belonging probability by multinomial logistic regression. We evaluate the performance of the proposed technique in the comprehensive simulation study. Also, we classified the five real data sets using our multinomial logistic model.

Keywords Power-Divergence, Multinomial Logistic, Shape Analysis, Classification, Logistic Regression.

AMS 2010 subject classifications 62H30, 62Hxx.

DOI: 10.19139/soic-2310-5070-221

1. Introduction

Shape analysis is one of the areas of multivariate statistics in which the main focus is on the geometric structures of objects. "The shape of an object is all the geometric information left after filtering that object's location, scale, and rotation effects." This formal definition was first given by Kendall [26] about shape. In order to analyze the shape, it is first necessary to record the configurations. The two main approaches in recording configurations are the use of landmarks and curves. Also, there are various methods to register configurations. The two well-known coordinate systems are introduced by Kendall [25] and Bookstein[5]. While the former is mainly used for theoretical analysis, the latter is better for exploratory analysis. In this article, the landmark base approach and Kendall shape space are followed. A valuable resource to get acquainted with shape analysis is [12] and the `shape` package in R software. There are several ways to summarize data; one is the classification of data based on one or more standard features. Therefore, different methods for categorizing shapes have been introduced to classify the shape data. For example, [9] studied longitudinal shape data clustering in different unlabeled classes using a new mixed model. Their model created an average shape trajectory for each cluster and used these curves for clustering. Due to shape space belonging to the manifold space, another clustering approach is using manifold space This approach to clustering the shapes is similar to using nonlinear image algorithms such as Isomap projection [32]. Another example of this approach is the image of shapes on linear space and the use of PCs to classify shapes using logistic regression [28].

Binary decisions are one of the common challenges in many practical problems. In data analysis, there are two different methods for this work: hypothesis testing and binary classification. However, choosing one of these is often unclear and confusing. One solution to this indeterminacy is to create a relationship between these two methods. Studies have been done in this regard. For example, a general kernel mixture model on manifold was

*Correspondence to: Meisam Moghimbeygi (Email: Email: m.moghimbeygi@khu.ac.ir). Department of Mathematics, Kharazmi University No.43 South Mofateh St Tehran, 15719-14911 Iran.

ISSN 2310-5070 (online) ISSN 2311-004X (print)
Copyright © 202x International Academic Press

introduced to establish the relationship between two predictor and response variables in [2]. Also, this study presented a hypothesis test on the hypersphere using the classification model. Another study, [24], introduced a classification method using LRT statistics. As an application, we can refer to the [13] in the radar target classification using means of multiple testing. A binary classification was done in [15] as a two sample testing problem. This research classified the instances by calculating the distance between the test and training instances. Then the two-sample test was performed under the null hypothesis.

Logistic regression describes the relationship between explanatory variables and a binary response variable. This modeling is a convenient tool for classifying data with binary responses. A generalization of this technique can also be defined in terms of multiple responses in the form of multivariate logistic regression. Due to the limitations of various application problems, researchers have generalized the logistic regression model to overcome them. For example, [27] extended nominal logistic regression models for binary responses to ordinal responses by involving the cumulative logits. They also introduced a nonparametric local linear smoothing technique for testing the goodness-of-fit of ordinal logistic regression models with continuous and categorical covariates. The local likelihood logit regression for an application to female labor supply is used by [14]. To estimate parameters, they employed a kernel function and least square methods. As another application of this method, [23] utilized this approach to classify diabetes mellitus patients. Another method to classify the categorical data is presented by [19]. They presented a kernel smoothing method for multinomial regression. The estimator of the regression functions is constructed by minimizing a localized power-divergence measure. The estimators include the bandwidth and a single parameter originating in the power-divergence measure as a smoothing parameter. A valuable study in the classification of shape data was done by [30]. They introduced the generalized partially linear models with covariates on the Riemannian manifold. Their model combined two parts, a nonparametric model of shape and a function of sizes with error distribution models other than a normal distribution.

Most existing classification methods are in the cast of an optimization problem and do not address the issue of statistical significance. Then, this paper aims to present a multinomial logistic regression (MLR) based on a power-divergence as a test statistics measure. In this model, the configurations appear as covariates where the power-divergence measure calculates the difference between configurations based on different distances in shape space. The remainder of the paper is organized as follows. In Section 2, we briefly review statistical shape analysis. Then, the MLR for shape data using the power-divergence measure is presented in Section 3. The model's performance in the simulation research and five real data is described in Section 4. The paper ends with some concluding remarks.

2. A Brief Background on the Statistical Shape Analysis

In this section, we will explain the preliminary concepts of shape statistics. The content presented in this section is taken from [12]. In shape statistics, the set of finite points that determine the geometry of an object is called the landmarks, and the set of all the landmarks considered for an object is called the configuration of that object. In addition, to specify the configuration is used a $k \times m$ matrix X where k is the number of landmarks and m is the dimension of configuration. According to Kendall's definition of shape, the shape is all the remaining geometric information of an object after filtering out the effects of location, scale, and rotation [26]. Therefore, to get the shape of a configuration, it is necessary to remove the effects of location, scale, and rotation. Centering matrices can be used to remove the location effect. One of these centralizers is the Helmert submatrix H. To see the Helmert submatrix see definition 2.5 from [12]. Hence, matrix HX is called a centered configuration or Helmertized. In the following, we will explain some concepts used in this article.

Definition 2.1. Let X_1 and X_2 as two $k \times m$ configurations matrices. Given $Z_j = HX_j / \parallel HX_j \parallel$, $j = 1, 2$ which is called pre-shape, the Riemannian distance $\rho(X_1, X_2)$ is the closest great circle distance between Z_1 and Z_2 on the pre-shape sphere and calculated by

$$\rho(X_1, X_2) = \arccos(\sum_{i=1}^{m} \lambda_i),$$

where λ_i's are the eigenvalues which are obtained from the following singular value decomposition

$$Z_1^T Z_2 = V\text{diag}(\lambda_1, ..., \lambda_m)U^T. \tag{1}$$

Definition 2.2. The partial Procrustes distance $d_p(X_1, X_2)$ is obtained by matching the pre-shapes Z_1 and Z_2 as closely as possible over rotations and is given by

$$d_p = \sqrt{2}\left(1 - \sum_{i=1}^{m} \lambda_i\right)^{1/2}.$$

Definition 2.3. The full Procrustes distance $d_F(X_1, X_2)$ is obtained by matching the pre-shapes Z_1 and Z_2 as closely as possible over rotations and scales and is given by

$$d_F = \left(1 - (\sum_{i=1}^{m} \lambda_i)^2\right)^{1/2}.$$

Definition 2.4. The Riemannian distance in size-and-shape space is given by:

$$d_S(X_1, X_2) = \sqrt{S_1^2 + S_2^2 - 2S_1S_2\cos\rho(X_1, X_2)},$$

where S_1, S_2 are the centroid sizes of X_1 and X_2.

Definition 2.5. Partial Ordinary Procrustes Analysis (OPA) involves Euclidean transformation translation and rotation match two configurations by minimization of the expression

$$\| X_2 - X_1\Gamma - 1_k\gamma^T \|^2 . \tag{2}$$

Using the definition 2.5, the partial OPA fit of X_1 onto X_2 is:

$$X_1^p = X_1\hat{\Gamma} + 1_k\hat{\gamma}^T,$$

where $\hat{\gamma}$ and $\hat{\Gamma}$ are optimal location and rotation parameters obtained by minimizing the equation (2). The minimization is carried out over rotations. Now, the optimization procedure can be done easily. Then the partial OPA solution to the minimization of (2) is given by

$$\hat{\gamma} = 0, \quad \hat{\Gamma} = UV^T,$$

where U and V are special orthogonal matrices ($SO(m)$) obtained from equation (1). More details about the optimization procedures can be seen in section 7 from [12].

3. Shape Classification

Let X_i be a configuration matrix and a covariate with a binary response variable $Y_i(\in \{0, 1\})$ for $i = 1, 2, ..., n$. Then we will present an MLR model to classify the configurations based on their categories. To this end, we first present a transform of configurations based on the power-divergence measure. Then, an MLR model on size-and-shape space will be introduced.

3.1. The criterion

For $j = 1, \ldots, r$, suppose that B_j's are random variables from a beta distribution with parameters $a_j > 0$ and $b_j > 0$. Next, suppose that Y_j are indicator random variables with the property that given $B_j = p_j \in (0, 1)$, Then

$Y_1, ..., Y_r$ are conditionally independent with Bernoulli density where

$$P(Y_j = 1|B_j = p_j) = p_j.$$

For n independent trials, each leading to one of the k categories with a fixed success probability, the multinomial distribution gives the probability of any particular combination of the number of successes for the different categories. Now, the joint conditional distribution of $Y_1|B_1 = p_1, \ldots, Y_r|B_r = p_r$ is the multinomial distribution $MN(p_1, \ldots, p_r; r)$.

In order to evaluate the estimator, there are some different methods. One of them that can be used for multinomial data is the goodness-of-fit test. The power-divergence measure is a measure to evaluate the estimator, which is defined by [6] as

$$I_\lambda(\boldsymbol{p} : \boldsymbol{q}) = \frac{1}{\lambda(1+\lambda)} \sum_{j=1}^{r} p_j \left\{ \left(\frac{p_j}{q_j}\right)^{\lambda} - 1 \right\},$$

for $\lambda \in \mathbb{R}$, $\boldsymbol{p} = (p_1, ..., p_r)$ and $\boldsymbol{q} = (q_1, ..., q_r)$ are probability distribution parameters. By minimizing the power divergence measure with the Lagrange condition $\sum_{j=1}^{r} p_j = 1$. Then, the Lagrange function defined as

$$\mathcal{L} = I_\lambda(\boldsymbol{p} : \boldsymbol{q}) + \ell(\sum_{j=1}^{r} p_j - 1)$$

where ℓ is Lagrange multiplier The estimator of $p_j(x)$ as

$$\hat{p}_j = \frac{q_j}{\sum_{j=1}^{r} q_j}.$$

To apply the model for shape data, we consider r categories of configuration data X_{ij}, for $j = 1, ..., r$ and $i = 1, ..., n_j$. Also, we use the two independent samples of Goodall's test to define our criteria. From section 9 of [12], given X_{ij} as a $k \times m$ random variable from $N_{km}(\mu_{ij}, \sigma^2 I_{km})$, the full Procrustes distance of X_{ij} from any fixed configuration X is $d_F^2(X_{ij}, X)$ and approximately distributed as $\tau_0^2 \chi_v^2$ distribution where $v = km + m(m+1)/2 - 1$, $\tau_0 = \sigma/\delta_0$, δ_0 is the centroid size of X and σ is small. For r categories of configurations,

$$B_{ij}(X) = \frac{d_F^2(X_{ij}, X)}{\sum_{j=1}^{r} d_F^2(X_{ij}, X)},$$

is a random variable from a beta distribution with parameters $v/2$ and $(r-1)v/2$. So the mean of this classifier is $1/r$. Using $B_{ij}(X)$ as an estimation for q_j, then a classifier can be defined as

$$\hat{p}_j(X) = \frac{\sum_{i=1}^{n_j} d_F^2(X_{ij}, X)}{\sum_{j=1}^{r} \sum_{i=1}^{n_j} d_F^2(X_{ij}, X)}. \tag{3}$$

Again, the joint conditional distribution of $Y_{i1}|B_{i1} = \hat{p}_1, \ldots, Y_{ir}|B_{ir} = \hat{p}_r$ is the multinomial distribution $MN(\hat{p}_1, \ldots, \hat{p}_r; r)$. Since the small value of distance $d_F^2(X_{ij}, X)$ means two configurations X_{ij} and X are in the same categories, then $\hat{p}_j(X)$ is the failure probability estimation of X belonging to category j. Note that since our purpose in this paper is to classify the data, there is no difference in using success and failure probability. From sec 4.2.1 of [12], for shapes that are close together, we have

$$d_p = d_F + O(d_F^3), \quad \rho = d_F + O(d_F^3).$$

For the concentrated configurations, the distances are the same. According to the special features of each of the introduced distances, they can be used instead of each other. Since size also plays an important role in classifying shape data in practical examples, the use of d_S can be a more suitable criterion for use in the introduced model. Therefore, d_S can be used instead of d_F in the equation (3).

The equation (3) is derived based on the assumptions made about the configuration distribution, and as a result, it is expected that this estimation may be biased. In other words, the estimator (3) follows a beta distribution with a mean of $1/r$, and the probability values are determined relative to this mean. These probabilities are then classified according to their proximity to the mean. Moreover, finding a cut point based on the probability $\hat{p}_j$ for data with more than two categories presents additional challenges.

The underlying assumptions of the approach include the multinomial distribution of the response variable and the linear relationship between covariates and the log-odds in the multinomial logistic regression (MLR) model. For the power-divergence measure, the assumptions rely on the parametric structure of the shape data, such as the beta distribution used in Goodall's test for the full Procrustes distance. While these assumptions provide a basis for the estimator, deviations may introduce biases in practical applications. For instance, the probabilistic classification of shape configurations based on distance measures could lead to inaccuracies if the distance metric does not align with the intrinsic characteristics of the data.

To address these potential biases, we employ the MLR model to refine the probabilities estimated by the power-divergence measure. The MLR framework mitigates biases by incorporating covariate effects and accounting for the multidimensional structure of the data. Furthermore, the use of MLR facilitates handling challenges such as cut-point determination in multi-category classification problems.

Despite its robustness, the method has limitations. The performance of the classifier heavily depends on the chosen distance measure, such as size-and-shape, Riemannian, or full Procrustes distances. In addition, the computational cost of iterative optimization in the MLR can pose challenges for large datasets. However, these limitations are offset by the model's flexibility and its ability to adapt to diverse datasets through informed selection of distance metrics.

The following subsection provides a detailed description of the MLR model and the process of parameter estimation, elaborating on how it integrates with the power-divergence-based probabilities to enhance classification accuracy.

3.2. MLR Model

One method of examining the relationship between two variables is using regression modeling. In regression modeling, a model is created to predict the dependent variable based on the independent variable. Usually, in the created model, both independent and dependent variables are quantitative. However, we may want to measure the relationship between an independent variable (with continuous values) and a variable dependent with qualitative values. The standard linear regression method will not work in this case, so MLR should be performed intermittently. MLR is a classification method for classifying two or more discrete results.

Suppose X is a predictor variable with a nominal response variable Y. In the standard linear regression, the expected value of the response variable Y is a linear combination of covariates. A simple linear model based on $p(x)$ can be defined as

$$\mathbb{E}(Y|X = x) = \alpha + \beta p(x),$$

where α is the constant parameter as intercept and $p(x)$ is some deterministic function. When the outcome variable Y is dichotomous, a choice is to use logistic regression. Other options to analyze these models can be found in [7]. Using the logistic distribution, we can define the conditional mean of Y given x ($\pi(x) = \mathbb{E}(Y|x)$) as

$$\pi(x) = \frac{e^{\alpha+\beta p(x)}}{1 + e^{\alpha+\beta p(x)}},$$

where this is also known as success probability. A simple transformation of $\pi(x)$ that plays a key role in this paper is logit transformation that is defined in terms of $\pi(x)$ as

$$\ln\left[\frac{\pi(x)}{1-\pi(x)}\right] = \alpha + \beta p(x). \tag{4}$$

Logistic functions lead to very simple structures (linear logits) for the analysis of the impact of the covariates [1].

For r categories with p covariate, the MLR model, is defined as the extension of (4) as

$$\begin{aligned} g_m(\boldsymbol{x}) &= \ln\left[\frac{\pi_m(\boldsymbol{x})}{\pi_k(\boldsymbol{x})}\right] \\ &= \beta_{m0} + \beta_{m1}p_1(\boldsymbol{x}) + \cdots + \beta_{mr}p_r(\boldsymbol{x}), \end{aligned} \tag{5}$$

where the $k \in \{1, \ldots, r\}$ is the category's index taken as a "pivot". A general expression for the conditional probability in the r category model is

$$\pi_j(\mathbf{x}) = \Pr(Y = j \mid \mathbf{x}) = \frac{e^{g_j(\mathbf{x})}}{\sum_{k=1}^{r-1} e^{g_k(\mathbf{x})}}, \quad j = 1, \ldots, r-1.$$

Let configuration matrices $X_{ij},\ i = 1, \ldots, n, j = 1, \ldots, r$ as explanatory variables in r different classes correspondence to Y_{ij} as nominal response variables. A way to parameterize a relatively general nonlinear function is to use feed-forward neural networks. For a multinomial log-linear model with r classes, a neural network with r outputs and the negative conditional log-likelihood is

$$E = -\sum_{i=1}^{n}\sum_{j=1}^{r} t_{ij} \log \pi_{ij}, \quad \pi_{ij} = \frac{e^{y_{ij}}}{\sum_{j=1}^{r} e^{y_{ij}}},$$

where t_{ij} is the target and y_i the output for the i-th example pattern. Since exactly one of the responses will be one (for the class which occurred) and the others all zero, the pretentious name has often known as 'softmax' fitting. The procedure of fitting the MLR with the neural network method in R software is available in [31] in detail. Also, the package `nnet` is considered in this paper for fitting the MLR to classify the shape data.

4. Simulation Study and Real Data Analysis

In this section, we will classify the simulated and real data examples. Then in the first subsection, we simulated data and classified them based on the MLR model. The second subsection is related to the classification of five real data sets.

4.1. Simulation Study

In order to evaluate the performance of our model, we use the method described in [29] to simulate configurations. Then, using the simulated data, we classify them based on our methods. The model is

$$X_{ij} = \beta_{ij}\mu_j\Gamma_{ij} + 1_k\gamma_{ij}^T + \varepsilon_{ij}, \ \ i = 1, ..., n, \ \ j = 1, ..., m, \tag{6}$$

where β_{ij}, Γ_{ij} and γ_{ij} are scale, rotation and location parameters. Also, μ_j is j-th reference configuration and ε_{ij} is a $m \times k$ random error from the multivariate normal distribution. We used the four reference configurations , star, heart, circle, and ellipse, with ten landmarks in two dimensions. The schematic representation of references is shown in Figure 1.

In the model, the parameter β is simulated from the uniform distribution on $(1, 2)$. Also, residuals generated from centered normal distribution from standard deviation $\sigma = 0.1, 0.2, 0.3, 0.4, 0.5$. Since the location and rotation parameters are less important in the classification procedures, we use the identity matrix for rotation and fix $\gamma_{ij} = 0$. The simulation process were done 1000 times and estimate the class of each obsevation with leave-one-out approach. The result can be seen in Tables 1-3 for different values of σ and number of configurations 10, 50 and 100 for each categories based on diferent distances. Some well-known statistical criteria such as pseudo-R-squares, Akaike information criterion(AIC), Bayesian information criterion(BIC), residual deviance (RD), and percentage of overall correct prediction (OCP) were reported. By increasing σ, the values of pseudo-R-squares and OCP decreased, and BIC and RD increased. This pattern there exists in all four used distances. Also, by increasing

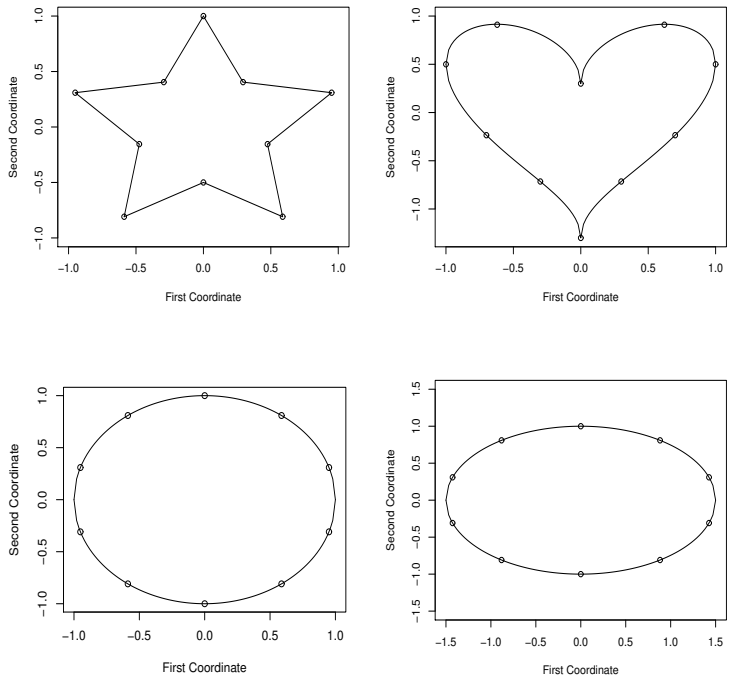

Figure 1. The schema of references configurations: Star, Heart, Circle, Ellipse

the number of configurations, the slope in the plot increased. The amount of standard deviation also increased by increasing σ. The schematic representation of mentioned result can be seen in Figure 2, 3 and Figures SM-1-SM-6 in Supplementary materials. In evaluating the accuracy of predictions, it can be seen that the three distances, Riemannian, partial and full Procrustes, have almost the same performance, and the size-and-shape distance has a lower accuracy compared to the other three distances (see Figure 4).

4.2. Application

In this subsection, five real data sets are examined. We also intend to compare the method introduced in this paper with the results of [28] for real data. The five real data sets used in this paper where are taken from the package `shapes` from "R" software, which is as follows:

- *Schizophrenia (Schi)*: 13 landmarks selected on near midsagittal 2D slices from magnetic resonance (MR) brain scans of 13 schizophrenia patients and 14 control people. The data are in two dimensions [4].
- *Sand*: The data were collected from sea and river sand grains. The sea particles are from the Baltic Sea, and the river particles are from the Caucasian River Selenchuk. There is 24 sea sand and 25 river sand grain profiles in two dimensions with 50 landmarks of the sands curve outline [22].
- *Macaque (Maca)*: A case study into sex differences in the crania of a species of Macaca fascicularis (a type of monkey). The data collected randomly from 9 male and 9 female skulls were registered in 7 landmarks and 3D. see [11] for more details on this data
- *Apes*: These data were collected from 167 skulls of great apes. The skull configuration of 30 female (gorf) and 29 male (gorm) adult gorillas, 26 female (panf) and 28 male (panm) adult chimpanzees, and 24 female (pongof) and 30 male (pongom) adult orangutans were considered. The data are described in detail by [17], and [18].

Table 1. Mean and standard deviation (SD) of Pseudo-R-Square measures (McFadden, CoxSnell, and Nagelkerke) and Model Fitting Criteria (AIC, BIC, RD and OCP) based on 1000 replications of the model (6) for different values of σ and configurations.

Number of configurations is 10								
Size-and-shape distance								
		Pseudo-R-Square			Model Fitting Criteria			
σ	Est.	McFadden	CoxSnell	Nagelkerke	AIC	BIC	RD	OCP
0.1	Mean	0.9955	0.9366	0.9991	24.4814	51.5035	0.4814	0.9368
	SD	0.0097	0.0018	0.0019	1.0533	1.0533	1.0533	0.0305
0.2	Mean	0.9918	0.9359	0.9984	24.8862	51.9082	0.8862	0.9334
	SD	0.0161	0.0030	0.0032	1.7422	1.7422	1.7422	0.0318
0.3	Mean	0.9885	0.9352	0.9977	25.2408	52.2628	1.2408	0.9317
	SD	0.0213	0.0041	0.0044	2.3050	2.3050	2.3050	0.0341
0.4	Mean	0.9884	0.9352	0.9977	25.2517	52.2738	1.2517	0.9296
	SD	0.0200	0.0038	0.0041	2.1638	2.1638	2.1638	0.0339
0.5	Mean	0.9865	0.9349	0.9973	25.4628	52.4849	1.4628	0.9217
	SD	0.0216	0.0041	0.0044	2.3341	2.3341	2.3341	0.0360
Riemannian distance								
0.1	Mean	1	0.9374	1	24.0001	51.0222	0.0001	1
	SD	0	0.0000	0	0.0000	0.0000	0.0000	0
0.2	Mean	1	0.9374	1	24.0006	51.0227	0.0006	0.9845
	SD	0.0001	0.0000	0	0.0067	0.0067	0.0067	0.0215
0.3	Mean	0.9994	0.9373	0.9999	24.0660	51.0881	0.0660	0.9565
	SD	0.0047	0.0009	0.0009	0.5064	0.5064	0.5064	0.0274
0.4	Mean	0.9974	0.9369	0.9995	24.2855	51.3076	0.2855	0.9471
	SD	0.0094	0.0018	0.0019	1.0179	1.0179	1.0179	0.0318
0.5	Mean	0.9945	0.9364	0.9989	24.5890	51.6111	0.5890	0.9387
	SD	0.0146	0.0028	0.0030	1.5755	1.5755	1.5755	0.0355
Partial distance								
0.1	Mean	1	0.9374	1	24.0001	51.0222	0.0001	1
	SD	0	0.0000	0	0.0000	0.0000	0.0000	0
0.2	Mean	1	0.9374	1	24.0003	51.0224	0.0003	0.9856
	SD	0	0.0000	0	0.0030	0.0030	0.0030	0.0208
0.3	Mean	0.9994	0.9373	0.9999	24.0671	51.0892	0.0671	0.9560
	SD	0.0046	0.0009	0.0009	0.5018	0.5018	0.5018	0.0275
0.4	Mean	0.9975	0.9369	0.9995	24.2651	51.2872	0.2651	0.9478
	SD	0.0089	0.0017	0.0018	0.9640	0.9640	0.9640	0.0317
0.5	Mean	0.9955	0.9365	0.9991	24.4897	51.5118	0.4897	0.9390
	SD	0.0132	0.0025	0.0027	1.4230	1.4230	1.4230	0.0338
Full distance								
0.1	Mean	1	0.9374	1	24.0001	51.0222	0.0001	1
	SD	0	0.0000	0	0.0000	0.0000	0.0000	0
0.2	Mean	1	0.9374	1	24.0005	51.0225	0.0005	0.9891
	SD	0.0001	0.0000	0	0.0073	0.0073	0.0073	0.0185
0.3	Mean	0.9995	0.9373	0.9999	24.0522	51.0743	0.0522	0.9567
	SD	0.0039	0.0007	0.0008	0.4188	0.4188	0.4188	0.0290
0.4	Mean	0.9984	0.9371	0.9997	24.1770	51.1990	0.1770	0.9466
	SD	0.0070	0.0013	0.0014	0.7518	0.7518	0.7518	0.0296
0.5	Mean	0.9972	0.9369	0.9995	24.2972	51.3192	0.2972	0.9436
	SD	0.0093	0.0017	0.0019	1.0082	1.0082	1.0082	0.0324

- *Steroid (Ster)*: The dataset consists of 42 to 61 atoms for each of 31 steroid molecules. This dataset is an array of dimensions $61 \times 3 \times 31$. The remaining coordinates are all zero if a molecule has less than 61 atoms. The steroids have different binding affinities to the corticosteroid-binding globulin receptor, so each molecule has an activity class of either high, intermediate, or low binding affinity. The aim is to classify steroids based on their activity class using the MLR model. This dataset has been analyzed before by [10], and [8].

The MLR model was used to classify five real datasets. The results are shown in Table 4. Note that in Table 4, we leave out each observed and then predict its categories to calculate OCP. As seen for schizophrenia and macaques

Table 2. Mean and standard deviation (SD) of Pseudo-R-Square measures (McFadden, CoxSnell, and Nagelkerke) and Model Fitting Criteria (AIC, BIC, RD and OCP) based on 1000 replications of the model (6) for different values of σ and configurations.

Number of configurations is 50								
		Size-and-shape distance						
		Pseudo-R-Square			Model Fitting Criteria			
σ	Est.	McFadden	CoxSnell	Nagelkerke	AIC	BIC	RD	OCP
0.1	Mean	0.9797	0.9339	0.9961	35.1869	87.9599	11.1869	0.9814
	SD	0.0101	0.0019	0.0020	5.5660	5.5660	5.5660	0.0087
0.2	Mean	0.9473	0.9276	0.9894	53.0531	105.8262	29.0531	0.9597
	SD	0.0174	0.0035	0.0037	9.6111	9.6111	9.6111	0.0129
0.3	Mean	0.9143	0.9206	0.9819	71.2985	124.0715	47.2985	0.9402
	SD	0.0223	0.0049	0.0053	12.3241	12.3241	12.3241	0.0167
0.4	Mean	0.8777	0.9120	0.9728	91.4998	144.2729	67.4998	0.9194
	SD	0.0259	0.0063	0.0068	14.3173	14.3173	14.3173	0.0186
0.5	Mean	0.8404	0.9024	0.9626	112.0835	164.8566	88.0835	0.8973
	SD	0.0283	0.0076	0.0081	15.5947	15.5947	15.5947	0.0205
Riemannian distance								
0.1	Mean	1	0.9375	1	24.0001	76.7732	0.0001	1
	SD	0	0.0000	0	0.0000	0.0000	0.0000	0.0003
0.2	Mean	0.9984	0.9372	0.9997	24.8908	77.6639	0.8908	0.9911
	SD	0.0040	0.0007	0.0008	2.2340	2.2340	2.2340	0.0059
0.3	Mean	0.9682	0.9317	0.9938	41.5332	94.3063	17.5332	0.9697
	SD	0.0179	0.0034	0.0036	9.8531	9.8531	9.8531	0.0118
0.4	Mean	0.9142	0.9205	0.9819	71.3461	124.1191	47.3461	0.9394
	SD	0.0246	0.0054	0.0058	13.5588	13.5588	13.5588	0.0160
0.5	Mean	0.8593	0.9074	0.9679	101.6222	154.3953	77.6222	0.9079
	SD	0.0275	0.0070	0.0075	15.1553	15.1553	15.1553	0.0196
Partial distance								
0.1	Mean	1	0.9375	1	24.0001	76.7732	0.0001	1
	SD	0	0.0000	0	0.0000	0.0000	0.0000	0.0003
0.2	Mean	0.9984	0.9372	0.9997	24.8782	77.6512	0.8782	0.9913
	SD	0.0040	0.0007	0.0008	2.2314	2.2314	2.2314	0.0060
0.3	Mean	0.9686	0.9317	0.9938	41.3490	94.1221	17.3490	0.9698
	SD	0.0178	0.0034	0.0036	9.8446	9.8446	9.8446	0.0119
0.4	Mean	0.9150	0.9207	0.9821	70.8797	123.6528	46.8797	0.9398
	SD	0.0244	0.0054	0.0057	13.4715	13.4715	13.4715	0.0160
0.5	Mean	0.8612	0.9079	0.9684	100.5944	153.3674	76.5944	0.9091
	SD	0.0272	0.0069	0.0074	15.0122	15.0122	15.0122	0.0194
Full distance								
0.1	Mean	1	0.9375	1	24.0001	76.7732	0.0001	1
	SD	0	0.0000	0	0.0000	0.0000	0.0000	0.0003
0.2	Mean	0.9985	0.9372	0.9997	24.8416	77.6147	0.8416	0.9932
	SD	0.0040	0.0007	0.0008	2.2179	2.2179	2.2179	0.0066
0.3	Mean	0.9690	0.9318	0.9939	41.1296	93.9027	17.1296	0.9702
	SD	0.0174	0.0033	0.0036	9.6086	9.6086	9.6086	0.0115
0.4	Mean	0.9178	0.9213	0.9827	69.3634	122.1365	45.3634	0.9416
	SD	0.0239	0.0052	0.0056	13.2021	13.2021	13.2021	0.0157
0.5	Mean	0.8677	0.9096	0.9702	97.0104	149.7835	73.0104	0.9131
	SD	0.0264	0.0066	0.0070	14.5644	14.5644	14.5644	0.0191

datasets, the performance of three distances, Riemannian, partial, and full, is the same and better compared to size-and-shape distance. The size-and-shape distance performed the classification better for the sand and apes dataset. This is because of the importance of size in classifying these types of datasets. The steroids datasets were classified better by using the full Procrustes distance.

The RMS criteria is a measure used in shape analysis to calculate the variation. This criterion is defined as

$$RMS = \frac{d_s}{\sqrt{k}},$$

Table 3. Mean and standard deviation (SD) of Pseudo-R-Square measures (McFadden, CoxSnell, and Nagelkerke) and Model Fitting Criteria (AIC, BIC, RD and OCP) based on 1000 replications of the model (6) for different values of σ and configurations.

Number of configurations is 100								
		Size-and-shape distance						
		Pseudo-R-Square			Model Fitting Criteria			
σ	Est.	McFadden	CoxSnell	Nagelkerke	AIC	BIC	RD	OCP
0.1	Mean	0.98005	0.93393	0.99619	46.07474	109.93817	22.07474	0.98545
	SD	0.00768	0.00141	0.00151	8.49516	8.49516	8.49516	0.00583
0.2	Mean	0.93958	0.92606	0.98779	90.83986	154.70330	66.83986	0.96105
	SD	0.01246	0.00256	0.00273	13.78325	13.78325	13.78325	0.00950
0.3	Mean	0.90222	0.91795	0.97915	132.1748	196.0382	108.1748	0.93893
	SD	0.01637	0.00372	0.00397	18.1092	18.1092	18.1092	0.01247
0.4	Mean	0.85593	0.90669	0.96713	183.37619	247.23963	159.37619	0.91334
	SD	0.01874	0.00485	0.00518	20.72775	20.72775	20.72775	0.01383
0.5	Mean	0.81320	0.89493	0.95460	230.64461	294.50804	206.64461	0.88749
	SD	0.01986	0.00578	0.00617	21.97053	21.97053	21.97053	0.01539
Riemmanian distance								
0.1	Mean	1	0.9375	1	24.00014	87.86358	0.00014	0.99998
	SD	0	0.0000	0	0.00003	0.00003	0.00003	0.00022
0.2	Mean	0.99627	0.93684	0.9993	28.12136	91.98479	4.12136	0.99348
	SD	0.00479	0.00085	0.0009	5.29585	5.29585	5.29585	0.00404
0.3	Mean	0.95513	0.92917	0.99111	73.63328	137.49671	49.63328	0.97005
	SD	0.01387	0.00273	0.00291	15.34267	15.34267	15.34267	0.00847
0.4	Mean	0.89036	0.91519	0.97621	145.2865	209.1499	121.2865	0.93385
	SD	0.01796	0.00422	0.00450	19.8660	19.8660	19.8660	0.01171
0.5	Mean	0.82990	0.89969	0.95967	212.1686	276.0321	188.1686	0.89872
	SD	0.01965	0.00547	0.00583	21.7405	21.7405	21.7405	0.01410
Partial distance								
0.1	Mean	1	0.9375	1	24.00014	87.86358	0.00014	0.99999
	SD	0	0.0000	0	0.00003	0.00003	0.00003	0.00018
0.2	Mean	0.99633	0.93685	0.99931	28.05795	91.92138	4.05795	0.99336
	SD	0.00479	0.00085	0.00090	5.30399	5.30399	5.30399	0.00405
0.3	Mean	0.95530	0.92920	0.99115	73.44581	137.30924	49.44581	0.97004
	SD	0.01389	0.00273	0.00291	15.36419	15.36419	15.36419	0.00842
0.4	Mean	0.89087	0.91531	0.97633	144.72514	208.58857	120.72514	0.93414
	SD	0.01792	0.00421	0.00449	19.82849	19.82849	19.82849	0.01174
0.5	Mean	0.83100	0.89999	0.96000	210.95948	274.82292	186.95948	0.89934
	SD	0.01958	0.00543	0.00579	21.65837	21.65837	21.65837	0.01403
Full distance								
0.1	Mean	1	0.9375	1	24.00014	87.86358	0.00014	0.99941
	SD	0	0.0000	0	0.00003	0.00003	0.00003	0.00106
0.2	Mean	0.99639	0.93686	0.99932	27.99447	91.85790	3.99447	0.99328
	SD	0.00483	0.00085	0.00091	5.34043	5.34043	5.34043	0.00403
0.3	Mean	0.95572	0.92928	0.99124	72.98874	136.85217	48.98874	0.97036
	SD	0.01381	0.00271	0.00289	15.27843	15.27843	15.27843	0.00849
0.4	Mean	0.89244	0.91568	0.97673	142.98737	206.85080	118.98737	0.93515
	SD	0.01778	0.00415	0.00443	19.67309	19.67309	19.67309	0.01162
0.5	Mean	0.83468	0.90101	0.96108	206.88699	270.75043	182.88699	0.90154
	SD	0.01933	0.00530	0.00566	21.38621	21.38621	21.38621	0.01396

where k is the number of landmarks. The small value of RMS means the category is homogeneous. In other words, the classification was done effectively if the RMS within each category before and after classification was close. The RMS measure helps us to compare our method in classification with others. We compare our method with the methods described in [28] and [29] based on the gorilla skulls data. These results are shown in Table 5. As seen in this table, the RMS quantities before and after clasification with our model are too close together, indicating that the model presented in this article outperforms the method presented in [28] and [29].

Compared to the clustering-based approach of [29], which does not leverage probabilistic classification frameworks, our method provides a more precise grouping by incorporating probabilistic estimates derived from the power-divergence measure. Similarly, the approach in [28] uses principal components but does not integrate

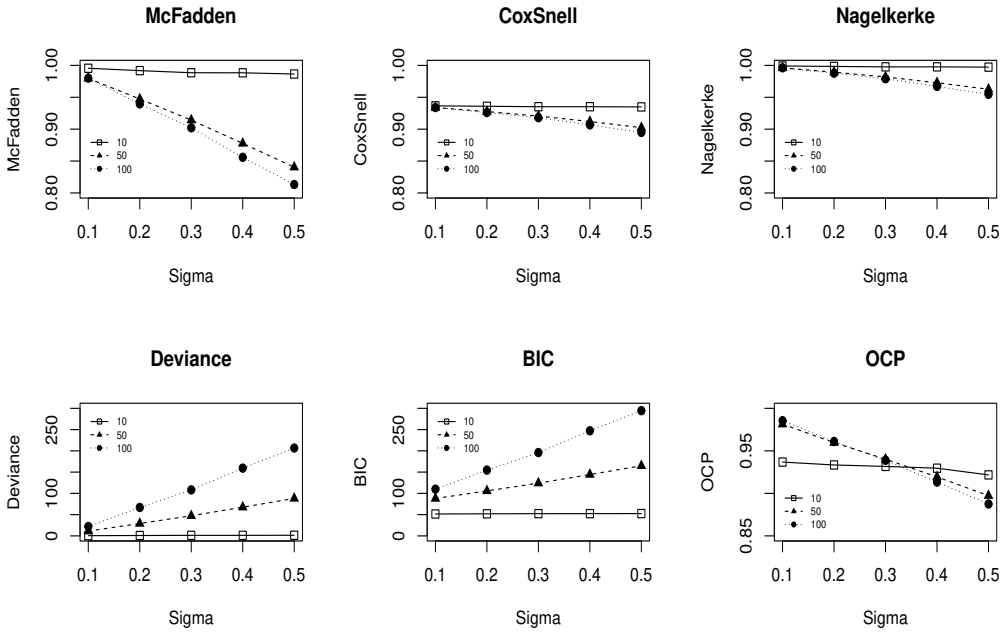

Figure 2. Mean of Pseudo-R-Square measures (McFadden, CoxSnell, and Nagelkerke) and Model Fitting Criteria (BIC, RD, and OCP) based on 1000 replications of the model (6) for different values of σ and number of configurations. The plots are based on Tables 1-3 and size-and-shape distance.

Table 4. McFadden, CoxSnell and Nagelkerke pseudo R-square and the summary of fitting the logistic regression such as AIC, BIC, residual deviance (RD) and percentage of overal correct prediction (OCP) for data sets schizophrenia, macaques, sand apes and steroids.

		Pseudo R-Square			Model Summary			
Data Set	Method	McFadden	CoxSnell	Nagelkerke	AIC	BIC	RD	OCP
Schi	Size-and-shape	0.56	0.54	0.72	21.12	26.45	17.12	85.71
	Riemannian	0.70	0.62	0.83	15.80	21.13	11.80	92.86
	Partial	0.70	0.62	0.83	15.79	21.12	11.79	92.86
	Full	0.70	0.62	0.83	15.75	21.08	11.75	92.86
Maca	Size-and-shape	0.86	0.70	0.93	7.46	11.02	3.46	88.89
	Riemannian	0.99	0.74	0.99	4.37	7.93	0.37	100.00
	Partial	0.98	0.74	0.99	4.61	8.17	0.61	100.00
	Full	0.95	0.73	0.98	5.24	8.80	1.24	100.00
Sand	Size-and-shape	0.60	0.57	0.76	30.90	38.47	26.90	87.76
	Riemannian	0.29	0.34	0.45	51.90	59.47	47.90	69.39
	Partial	0.30	0.34	0.45	51.84	59.40	47.84	69.39
	Full	0.30	0.34	0.45	51.64	59.21	47.64	69.39
Apes	Size-and-shape	0.94	0.97	0.99	96.59	128.64	36.59	91.62
	Riemannian	0.88	0.96	0.98	129.74	161.80	69.74	83.83
	Partial	0.88	0.96	0.99	129.47	161.53	69.47	83.83
	Full	0.88	0.96	0.99	128.66	160.72	68.66	83.83
Ster	Size-and-shape	0.98	0.88	0.99	13.21	21.22	1.21	96.77
	Riemannian	0.98	0.88	1.00	13.05	21.06	1.05	87.10
	Partial	0.96	0.88	0.99	14.33	22.34	2.33	96.77
	Full	1.00	0.89	1.00	12.20	20.21	0.20	100.00

logistic regression or the flexibility of multiple distance measures as employed in this paper. These distinctions underscore the advantages of our two-stage classification framework.

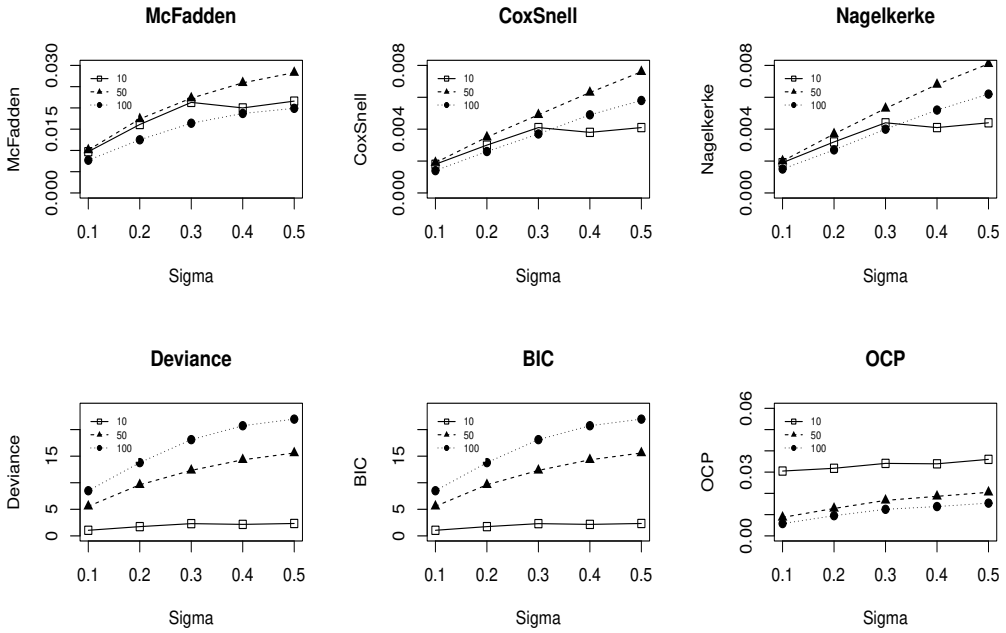

Figure 3. The standard deviation of Pseudo-R-Square measures (McFadden, CoxSnell, and Nagelkerke) and Model Fitting Criteria (BIC, RD, and OCP) based on 1000 replications of the model (6) for different values of σ and number of configurations. The plots are based on Tables 1-3 and size-and-shape distance.

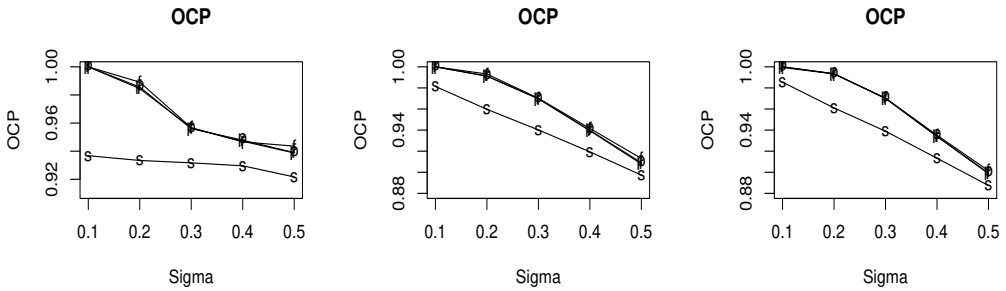

Figure 4. Left to right: OCP of classification with different distance measures size-and-shape (s), Riemannian (r), partial (p), and full (f) Procrustes for different values of σ and number of configurations 10, 50, and 100, respectively.

Additionally, our method's ability to incorporate diverse distance measures, such as Riemannian, partial, and full Procrustes distances, enhances its adaptability to various datasets. This flexibility is especially advantageous for datasets with complex structures or significant size variations, as demonstrated by the improved performance on the gorilla skulls data. The close alignment of RMS values before and after classification further reflects the robustness of our approach in maintaining within-category homogeneity.

Future comparisons could expand upon this analysis by including methods that utilize Riemannian manifold approaches or kernel smoothing techniques, which are often used in shape analysis. However, the results presented here strongly suggest that our method is effective and competitive, particularly for datasets where probabilistic classification and flexible metric selection are critical.

Table 5. The RMS measure for both groups of male and female gorillas skulls before and after classification (clas.) according to the method described in [29](a) and [28] (b) and our model (c) based on size-and-shape distance.

		Before clas.	After clas. (a)	After clas. (b)	After clas. (c)
RMS	Female	0.0437	0.0469	0.0432	0.0437
	Male	0.0499	0.0466	0.0479	0.0499
Mean of two RMS		0.0468	0.0468	0.0456	0.0468
Distance between means		0.0586	0.0591	0.0595	0.0586

Table 6. The overall correct prediction (OCP) for the schizophrenia, macaque, sand, ape, and steroid datasets was evaluated using machine learning methods, including MLR, SVM, k-NN, Naive Bayes, XGBoost, and LDA, with a 70-30 train-test split approach.

Data set	Method	MLR	SVM	k-NN	Naive Bayes	XGBoost	LDA
Schi	Size-and-shape	100	100	100	100	100	100
	Riemannian	100	100	87.5	100	100	100
	Partial	100	100	87.5	100	100	100
	Full	100	100	87.5	100	100	100
Maca	Size-and-shape	75	75	75	75	75	75
	Riemannian	100	100	100	100	100	100
	Partial	100	100	100	100	100	100
	Full	100	100	100	100	100	100
Sand	Size-and-shape	85.71	85.71	78.57	85.71	71.43	85.71
	Riemannian	57.14	78.57	71.43	78.57	78.57	71.43
	Partial	57.14	78.57	71.43	78.57	78.57	71.43
	Full	64.29	78.57	71.43	78.57	71.43	71.43
Apes	Size-and-shape	93.75	2.08	39.58	91.67	91.67	97.92
	Riemannian	81.25	2.08	14.58	81.25	77.08	85.42
	Partial	81.25	2.08	14.58	81.25	77.08	85.42
	Full	83.33	2.08	14.58	81.25	77.08	85.42
Ster	Size-and-shape	66.67	11.11	66.67	55.56	88.89	55.56
	Riemannian	55.56	11.11	66.67	55.56	77.78	55.56
	Partial	66.67	11.11	66.67	55.56	88.89	55.56
	Full	55.56	11.11	55.56	66.67	77.78	55.56

The method proposed in this paper consists of two stages for classifying shape data. In the first stage, the estimation (3) is introduced. This estimation, based on the power-divergence measure, calculates probabilities of data points belonging to specific categories by leveraging shape distances. In the second stage, these probabilities are used in combination with a logistic regression model to classify the shape data. To enhance classification efficiency and adaptability, various machine learning models are employed in the second stage. These include Multinomial Logistic Regression (MLR), Support Vector Machine (SVM), k-Nearest Neighbors (k-NN), Naive Bayes, Extreme Gradient Boosting (XGBoost), and Linear Discriminant Analysis (LDA).

The OCP values for classifying the real data using these methods are presented in Table 6, based on an 70-30 train-test split approach. As observed, methods other than MLR occasionally demonstrate better performance. However, the choice of method should depend on the dataset's structure and characteristics. Additionally, it is evident that for datasets with insufficient volume, this approach does not yield optimal results due to the limited information available for model training and validation.

From a computational perspective, the method's two stages involve challenges in terms of efficiency, particularly when applied to large datasets. The calculation of shape distances, especially computationally intensive metrics like full Procrustes distance, can become a bottleneck. Furthermore, the iterative parameter estimation required in MLR or alternative machine learning methods adds to the computational load.

To address these bottlenecks, we suggest leveraging parallel computing techniques to accelerate both the distance calculations and the model fitting process. Modern libraries in *R*, such as *nnet* and *Rcpp*, can provide optimized implementations for these tasks, significantly reducing execution time. Additionally, dimensionality reduction techniques like principal component analysis (PCA) can be employed to preprocess the data, thereby reducing computational complexity without sacrificing accuracy.

For future improvements, exploring GPU computing or approximate methods for distance calculations, such as approximate nearest neighbor techniques, could further enhance scalability. These strategies, combined with algorithmic refinements and distributed computing, could extend the applicability of the proposed method to much larger datasets while maintaining classification accuracy.

Conclusion

The paper focuses on shape data classification. This classification is done using the MLR model based on the estimator in shape space. The estimator part of the model is obtained by minimizing the power-divergence measure. Four different distances, size-and-shape, Riemannian, partial and full Procrustes, were used in the classification method. In this model, the estimation of the MLR parameters was done using the neural network method. The performance of the model was evaluated using simulated data. Also, as an application of our classification method, five real data sets were classified using our MLR model. In the simulation study, since most of the differences in the configurations were due to their shape structure differences, the size-and-shape distance was less accurate compared to the other three distances. Versus, in the study of the sand and apes real datasets, the size-and-shape distance performed better than the others due to the point of view of large size differences. The mentioned method was compared with the methods described in [28] and [29] based on size-and-shape distance on the gorilla's skulls dataset.

Two key factors distinguish shapes: size differences and shape differences. For shapes of the same size, metrics such as Riemannian distance are particularly effective. Conversely, for shapes that differ primarily in size, size-and-shape distance measures are more suitable. In the sand and apes datasets, for example, size differences are prominent across categories. River sands are generally larger than sea sands, and in the ape dataset, male gorillas are typically larger than other monkeys, although some species exhibit similar sizes. In such cases, size serves as a critical factor in classification. The result showed that the MLR method in this paper classified data better than the two methods mentioned in the literature.

For highly similar shapes, most distance metrics produce nearly identical results, making it difficult to classify each shape accurately. One common challenge in shape classification is the issue of collinearity between shape and reflection. However, this challenge was not encountered in the analysis of the data presented in this article. Researchers such as [25] and [16] have proposed methods for calculating distances in these special cases. While the shapes package performs efficiently on small datasets, its speed becomes a limitation when applied to medium and large datasets. To address this, researchers like [20] have introduced optimizations to enhance the performance of certain functions within the package.

While the proposed method demonstrates strong performance, there are several exciting directions for future research. One potential extension is to handle more complex shapes, such as 3D objects or shapes with topological features, by incorporating techniques like persistent homology or topological data analysis. Additionally, incorporating other covariates, such as size, orientation, or texture, alongside shape data, could enhance classification accuracy, particularly in fields like biometrics and robotics.

Another important direction is to explore multiscale and multimodal analyses, combining shape data with other data types, such as sensor data, to improve classification performance in applications like medical imaging and autonomous systems. Improving the method's computational efficiency is also a key area for future work, with opportunities to optimize the approach through GPU acceleration, parallel processing, or approximate distance calculations to handle larger datasets and enable real-time applications.

Finally, further theoretical advancements could focus on alternative divergence measures or integrating the method with other statistical models to enhance its robustness and adaptability. These directions will help expand the method's applicability and improve its practical use across diverse fields.

REFERENCES

1. A. Agresti, *Analysis of Ordinal Categorical Data, 2nd edition*, J. Wiley & Sons, Hoboken, New Jersey, 2010.
2. A. Bhattacharya and D. Dunson, *Nonparametric Bayes classification and hypothesis testing on manifolds*, Journal of Multivariate Analysis, vol. 111, pp. 1–19, 2012.
3. F. Boas, *The horizontal plane of the skull and the general problem of the comparison of variable forms*, Science, vol. 21, pp. 862–863, 1905.
4. F. L. Bookstein, *Biometrics, biomathematics and the morphometric synthesis*, Bulletin of Mathematical Biology, vol. 58, pp. 313–365, 1996.
5. F. L. Bookstein, *Size and shape spaces for landmark data in two dimensions (with discussion)*, Statistical Science, vol. 1, pp. 181–242, 1986.
6. N. Cressie and T. R. Read, *Multinomial goodness-of-fit tests*, Journal of the Royal Statistical Society: Series B (Methodological), vol. 46, pp. 440–464, 1984.
7. D. R. Cox and E. J. Snell, *Analysis of binary data*, Chapman and Hall/CRC, Boca Raton, 1989.
8. I. Czogiel, I. L. Dryden, and C. J. Brignell, *Bayesian matching of unlabeled marked point sets using random fields, with an application to molecular alignment*, Annals of Applied Statistics, vol. 5, pp. 2603–2629, 2011.
9. V. Debavelaere, S. Durrleman, and S. Allassonnière, *Learning the clustering of longitudinal shape data sets into a mixture of independent or branching trajectories*, International Journal of Computer Vision, vol. 128, pp. 2794–2809, 2020.
10. I. L. Dryden, J. D. Hirst, and J. L. Melville, *Statistical analysis of unlabeled point sets: comparing molecules in chemoinformatics*, Biometrics, vol. 63, pp. 237–251, 2007.
11. I. L. Dryden and K. V. Mardia, *Multivariate shape analysis*, Sankhyā: The Indian Journal of Statistics, Series A, vol. 55, pp. 460–480, 1993.
12. I. L. Dryden and K. V. Mardia, *Statistical Shape Analysis: With Applications in R*, John Wiley and Sons, Chichester, 2016.
13. A. Farina and A. Visconti, *Classification of radar targets by means of multiple hypotheses testing*, In Radar-87; Proceedings of the International Conference, pp. 73–78, 1987.
14. M. Frölich, *Non-parametric regression for binary dependent variables*, The Econometrics Journal, vol. 9, pp. 511–540, 2006.
15. Z. He, C. Sheng, Y. Liu, and Q. Zou, *Instance-based classification through hypothesis testing*, IEEE Access, vol. 9, pp. 17485–17494, 2021.
16. H. Le and B. Bhavnagri, *On simplifying shapes by subjecting them to collinearity constraints*, Mathematical Proceedings of the Cambridge Philosophical Society, vol. 122, pp. 315–323, 1997.
17. P. O'Higgins, *A morphometric study of cranial shape in the Hominoidea*, Ph.D. diss., University of Leeds, 1989.
18. P. O'Higgins and I. L. Dryden, *Sexual dimorphism in hominoids: further studies of craniofacial shape differences in pan, gorilla, pongo*, Journal of Human Evolution, vol. 24, pp. 183–205, 1993.
19. H. Okumura and K. Naito, *Non-parametric kernel regression for multinomial data*, Journal of Multivariate Analysis, vol. 97, pp. 2009–2022, 2006.
20. G. Quintana-Ortí and A. Simó, *A kernel regression procedure in the 3D shape space with an application to online sales of children's wear*, Statistical science, vol. 34, pp. 236–252, 2019.
21. J. G. Staniswalis, *The kernel estimate of a regression function in likelihood-based models*, Journal of the American Statistical Association, vol. 84, pp. 276–283, 1989.
22. D. Stoyan and H. Stoyan, *Fractals, Random Shapes and Point Fields: Methods of Geometric Statistics*, John Wiley & Sons, Ltd, Chichester, 1994.
23. M. Suliyanto, Rifada, and E. Tjahjono, *Estimation of nonparametric binary logistic regression model with local likelihood logit estimation method (case study of diabetes mellitus patients at Surabaya Hajj General Hospital)*, In AIP Conference Proceedings, vol. 2264, no. 1, p. 030007, AIP Publishing LLC, 2020.
24. V. Y. F. Tan, S. Sanghavi, J. W. Fisher, and A. S. Willsky, *Learning graphical models for hypothesis testing and classification*, IEEE Transactions on Signal Processing, vol. 58, no. 11, pp. 5481–5495, 2010.
25. D. G. Kendall, *Shape manifolds, Procrustean metrics, and complex projective spaces*, Bulletin of the London Mathematical Society, vol. 16, pp. 81–121, 1984.
26. D. G. Kendall, *The diffusions of shape*, Advances in Applied Probability, vol. 9, pp. 428–430, 1977.
27. K. C. Lin and Y. J. Chen, *Assessing ordinal logistic regression models via nonparametric smoothing*, Communications in Statistics-Theory and Methods, vol. 37, pp. 917–930, 2008.
28. M. Moghimbeygi and A. Nodehi, *Multinomial principal component logistic regression on shape data*, Journal of Classification, 2022.
29. M. Nabil and M. Golalizadeh, *On clustering shape data*, Journal of Statistical Computation and Simulation, vol. 86, pp. 2995–3008, 2016.
30. A. Simó, M. Victoria Ibáñez, I. Epifanio, and V. Gimeno, *Generalized partially linear models on Riemannian manifolds*, Journal of the Royal Statistical Society: Series C (Applied Statistics), vol. 69, pp. 641–661, 2020.
31. W. N. Venables and B. D. Ripley, *Modern Applied Statistics with S, 4th ed.*, Springer, 2002.
32. D. Yankov and E. Keogh, *Manifold clustering of shapes*, In Proceedings of IEEE International Conference on Data Mining (ICDM), 2006.