A Novel Divide and Merge Approach for Improved Classification of Functional Data

Wei Zhao
wei.zhao-6@postgrad.manchester.ac.uk
Xiao-Jun Zeng*
x.zeng@manchester.ac.uk
Chengdong Shi
chengdong.shi@postgrad.manchester.ac.uk
Ching-Hsun Tseng
ching-hsun.tseng@postgrad.manchester.ac.uk
Yue Chang
yue.chang@manchester.ac.uk

Department of Computer Science dept.
University of Manchester
Manchester, UK

Abstract

To address the challenges posed by the complexities unsolved by traditional classification methods when dealing with high-dimensional functional data, we propose a divide and merge method based on B-spline based Functional Data Analysis (FDA). The proposed method not only optimizes the classification performance for functional data, but also achieves dimension reduction and simplify the model by transitioning the learning problem from an infinite-dimensional vector space to a corresponding finite-dimensional parameter space. Unlike traditional knot placement that focus on a specific part or subset of dataset, through FDA techniques, this method first divides the input data according to categories, then determines the knot vector for each category, and dynamically merges knots to adapt to the inherent structure and trends of all kinds of labeled data. Finally, we calculate the common B-spline basis function for entire dataset by the merged knot vector, then encode the raw data into a smaller parameter space through B-spline approximation, and the classification procedure can be done through existing machine learning method enhancing the accuracy and efficiency of classification on functional data. Through four comparative experiments on real-world datasets, we demonstrate how this method significantly improves the classification accuracy over the existing functional classification methods while maintaining a moderate model complexity. Besides, our work reveals the impact of knot configuration on model interpretability and on dimensional reduction.

Keywords—Functional Data Analysis, B-spline, Function Data Classification

1 Introduction

Functional data classification is one of the key technologies for analyzing complex data structures with inherent continuity and plays a crucial role in various fields. Although tra-

[1] *Corresponding author: Xiao-jun Zeng

ditional classification methods such as logistic regression, decision tree, and support vector machines (SVM) have achieved some success in processing such data, they often fail to fully capture the continuity and inherent characteristics of the data. In recent years, with the increase in functional data applications [20], researchers have begun to explore more refined data approximation methods based on FDA [15] to improve the accuracy and efficiency of functional data classification. By using FDA, the raw data points are transformed into curves rather than discrete vectors, emphasizing the function nature of the data, and it allows a more comprehensive understanding and representation of the data. As FDA considers data in their continuous form, it shows the underlying patterns and trends within the data, enhancing the analysis and subsequent classification processes.

The conversion from raw data points into the curves is an essential step in FDA, highlighting its fundamental principle of viewing data from a functional perspective. I.J. Schoenberg developed Splines interpolation and approximation [18] in 1946. It is a mathematical method to find a smooth approximation function, via using a series of knot vectors to build the smooth curve. There are several types of splines, the linear splines, composed of line segments, the quadratic and cubic splines, consists of 2nd or 3rd order piecewise polynomials, and B-spline. B-spline interpolation and approximation [18] shows great flexibility, allowing the use of a set of knots to construct the approximator to accurately represent the curve's shape, with each knot influencing only its corresponding interval so that each knot can be adjusted independently. B-spline curve fitting is the most widely used technique for smoothing and fitting functional data. FDA focus on handling continuous, curve-like data, and B-spline fitting has been a fundamental method in performing various analytical tasks on such data.

In the B-spline curve fitting and approximation field, knot selection is a core issue that has a decisive impact on model performance. The placement and number of knots determine the shape and flexibility of the B-spline, affecting the ability to accurately represent the raw data. There are many knots placement methods. One straightforward and simple approach is the uniform placement proposed by Piegl and Tiller [13], in which the knots are evenly distributed over the whole data range, but the performance is unsatisfactory where data curve or function variations are not evenly distributed. Liang *et al*. [11] suggested the iterative knot insertion (IKI) method that begins with a small group of knots and then the knots are dynamically inserted in segments where the given threshold was exceeded by the approximation error. This iterative insertion continues until the error in all segments no longer exceeds the threshold, optimizing the fitness of the spline approximator to the data. Yeh *et al*. [21] places the knots by distributing the "feature contents" evenly to ensure that the feature value covered between every two knots is the same. Also, a machine learning approach using SVM for knot placement is described by Laube *et al*. [9]. The performance of their approach depends heavily on the training datasets, which limits the performance of the algorithm. Although there are various methods to represent a function as we mentioned above, in the context of FDA, the challenge is not just to represent a single function. Instead, we need to identify a set of common basis functions that can represent all the functional data in a dateset. This problem and approach is fundamentally different from the traditional methods.

For the classification problem on functional data, there are some popular approaches developed and published. The Nearest Centroid Classifier proposed by Delaigle and Hall [4], classifies based on the centroid function, obtained by calculating the pointwise average of all function in each label. But it is weak when dealing with functional data who show high intra-class variability or overlapping categories, as it assumes the data distribution between different labels is uniform. Another application of distance-based classification methods, is

the k-Nearest Neighbors (k-NN) [1], which can be used for functional data by defining an appropriate distance between functions, such as L2 norm. But it is difficult to determine the right value of K or the uniformed K for all input data may not exist. The concept of data depth was first introduced by Cuevas *et al*. [2] into functional data classification. The idea of data depth offers a significant advantage for creating robust estimators of a "location parameter" within high-dimensional or functional frameworks. *E.g*., once the depth is defined, the calculation of other parameters of robust estimators such as the median and trimmed become clear and direct. From the classification perspective, data depth is employed to classify new data points based on the comparative depth relative to the existing training data. And some other depth-based classification methods are proposed in Sguera *et al*. [19], Hlubinka *et al*. [5], and Kwon *et al*. [8]. A further method about functional data classification is SVM. The development of a functional SVM aims to identify the decision boundary $\beta(t)$ that optimally separates various categories of functional observations. This is achieved by maximizing the margin between classes. Rossi and Villab [17] were first applied SVM for classifying functional data based on linear decision boundaries. Additionally, to overcome the nonlinear separation, the functional-adjusted kernels were also taken into consideration. B.Li and Yu [10] introduced SVM to a low-dimensional subspace formed by the projection of Linear Discriminant Analysis (LDA) for classifying functional data. This approach initially uses LDA to reduce the dimensionality of the data and enhance the separability between classes, and then employs SVM for classification within this simplified subspace to maximize the classification margin. This combined method, integrating the dimensionality reduction capabilities of LDA with the classification strengths of SVM, is particularly suitable for handling high-dimensional datasets, effectively improving classification accuracy and computational efficiency. But LDA is sensitive to outliers because it relies on computing covariances within and between classes. Outliers can distort these statistics, thereby affecting the robustness of the overall model. Furthermore, Huang *et al*. [6], proposes a new similarity measurement method for functional data based on functional Mahalanobis distance and regenerative kernel theory. The method is applied to functional kernel principal component analysis (PCA) and combined with classification algorithms like support vector machines and random forests to classify functional data. Through empirical analysis, the proposed method outperforms traditional regenerative kernel methods based on Euclidean distance in classification tasks. Additionally, the similarity measurement can be extended to other machine learning algorithms based on regenerative kernel theory for further analysis of functional data.

However, in the classification problems we deal with data from different categories, and we require a set of knots to record the area of essential information from every label, and transform data from all labels into different curves. It is hard to find that kind of knots vector from the traditional method such as uniform placement or IKI we talked above, since they focus on one curve. If the knot placement for one curve is applied to transform all the data, the information from other labels will be lost, as it is not represented in the knot vector.

In this work, we propose a novel approach to build the knot vector. We first calculate the knot vectors for every label which refers to the divide part, and merge the knot vectors which capture the corresponding label's crucial information into a new set of knots to ensure that the information from different label can be contained. And the merged knot enables us to calculate the common B-spline basis functions and fit the entire dataset. After that, we represent the raw data by curves based on FDA with a small number of parameters-the coefficients of B-spline basis function. By doing this, we achieve the B-spline approximation of the entire dataset and the dimension reduction by encoding the discrete raw data points

from high dimensional vector space into a parameter space consists of a smaller number of coefficients from the B-spline fitting process. Then the classification process can be done within the parameter space with conventional classification method to avoid the issue of high dimensionality and complexity when directly feeding raw functional data to meet the demand of modern science and technology in industry.

The main contribution of this work as follows:

a) We proposed the divide-merge method to find a right set of knots and then construct a set of common basis functions to represent all the functions in a raw dataset, which is the first step in solving the representation problems with functional data.

b) Through the common basis functions we got, the traditional high-dimensional learning problems are converted from an infinite-dimensional vector space to a corresponding finite-dimensional parameter space, which can be solved by traditional machine learning methods, thus reducing dimensionality, enhancing simplicity and improving efficiency.

c) Through several experiments, it is shown that the approach led to simpler classifiers and improved classification accuracy, thereby enhancing the model's interpretability and performance, confirming the feasibility and effectiveness of the method.

The remainder of the paper organized as follow. Section II briefly reviews the FDA. Our method is presented in Section III, and Section IV introduce the experiments and results compare with prior work. The conclusions is displayed in Section V, followed by the acknowledgment in Section VI.

2 Functional Data Analysis

FDA is a branch which concentrates on scenarios where for each subject there are multiple observations being continuously measured over time or space. FDA methods show valuable insights in numerous fields including Economics, Public Health, Meteorology, Paleopathology, Graphology and Criminology [16].

Given a n-dimensional vector $X_i = [x_{i,1}, x_{i,2}, .., x_{i,n}]^T$ $(\forall i = 1, 2, ..N)$, it can be described as a continues function $x_i(t)$. We aim to fit the discretized function using the model $X_i = x_i(t) + \varepsilon$. The simple method for represent such a functional data is using basis functions in continuous function spaces. There are several different types of basis functions including B-spline basis, Fourier basis, wavelet basis. The function $x_i(t)$ can be represented by basis functions:

$$x_i(t) = \sum_{k=1}^{K} c_{i,k} \varphi_k(t) \tag{1}$$

$\varphi_k(t)$ is the basis function evaluated at time t, K is the total number of basis functions. $c_{i,k}$ $(k = 1, 2, ..., K)$ are the real number coefficients.

In this work, B-spline basis is used for the functional data representation. The definition of B-spline basis functions can be defined recursively. For a given series of knots $\{t_i\}$, B-spline basis function $\{B_{i,n}(t)\}$, where n is the order of B-spline functions, can be calculated by:

When n=1 (zero-order B-spline basis function):

$$B_{i,1}(t) = \begin{cases} 1 & \text{if } t_i \le t \le t_{i+1}, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

When n>1(higher-order B-spline basis function):

$$B_{i,n}(t) = \frac{t - t_i}{t_{i+n-1} - t_i} B_{i,n-1}(t) + \frac{t_{i+n} - t}{t_{i+n} - t_{i+1}} B_{i+1,n-1}(t) \tag{3}$$

For a given coefficients $\{P_i\}$, the B-spline curve function can be described as:

$$S(t) = \sum_{i=1}^{k} P_i B_{i,n}(t) \tag{4}$$

Here, the coefficients $\{P_i\}$ form as vectors. They crucially determine the shape and the aesthetics of the B-spline curve. And $S(t)$ corresponds to the B-spline curve, a parametric curve that describes the combined influence of the coefficients and basis functions over the parameter t. The coefficients P decide the shape of $S(t)$, and the shape refers to the form or outline created by the curve or surface as it spans through and around P.

In B-spline curve fitting, the knot placement determines the setting of basis. We prefer to put more knot when the function or curve fluctuates more since more information could be in there, and less on flat part. As the result of that, an optimized knot placement leads to a good or accurate performance of the functional data representation and as a consequence the better classification result. In the next section, we will develop our new method about knot placement.

3 Proposed Method

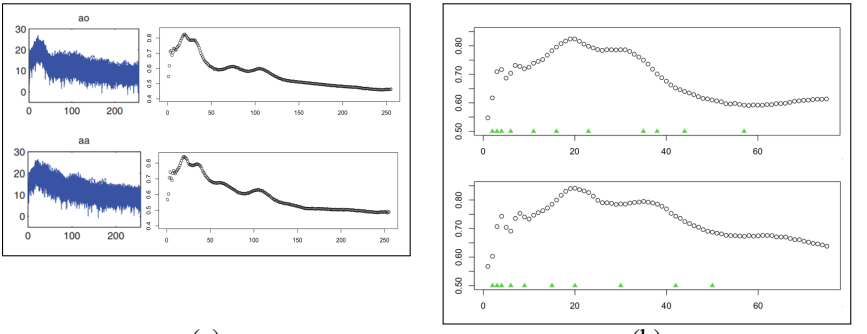
To develop our approach, the first problem we need to solve is to propose a new method to represent all the input functions or curves for a given FDA problem. More exactly, we want to solve the problem how to find a common set of B-spline basis functions which can be used to accurately represent all input functions. As the B-spline basis functions are totally determined by their knots, this problem in turn is to find a common set of knots whose corresponding B-spline basis functions can be used to accurately approximate all input functions or curves. Although several knot selection methods were available as discussed in the introduction, such a method is designed to approximate one given function and therefore it does not fit to our purpose here. Therefore we need to develop a new knot placement method for a given set of functions.

Toward this purpose, we proposed the following divide and merge approach, which identifies the main features of each class and leads the required common B-spline basis functions, and then solve the accurate input-function representation problem as the foundation of FDA classification. The proposed method is comprised of the following steps:

3.1 Divide

In this step, we divide the data according to labels. We assume that all data from same label share some common features, therefore, the mean series composed of the mean values of every observation point with the same label were treated as that label's representative. From a function point of view, it shows the overall shape of one kind of input data.

Figure 1 (a) shows two normalized mean series from one of our experiments, the left part is the observations, and the right is the mean series. Data 'aa' and 'ao' are the two most difficult groups to classify in the Phoneme dataset refers to section 4.1. As can be



(a) (b)

Figure 1: (a) The mean series of the two labels; (b) Knot placement for two mean series.

seen in Figure 1 (a), their differences of features are concentrated in the region where the x-coordinate is between 40 and 80, with other parts being quite similar.

3.2 Knot placement

To serve to our purpose, the Fast Automatic Knot Placement method [21] is applied and extended to be used in this step. Following this method, we use the central difference to calculate the derivatives of the mean series for every label, then the feature functions of different categories which measures the amount of information can be built by the approximation value of each point on the derivative. After calculating the integral of every feature function, the corresponding knot vector can be determined by evenly distribution of the integral.

We calculate the knot placement of the two mean series in Figure 1 (a). The knot vector from 0 to 80 which is the most different part between data 'aa' and 'ao' is displayed in Figure 1 (b). The triangle shows the knot in the mean series.

3.3 Knot merge

As we got the knot vectors for different labels from the Fast Automatic Knot Placement method mentioned above, which distributes the feature evenly to guarantee that the feature value spanned between adjacent knots is the same. Based on FDA, the concentrated distribution of knots indicates that the data variation in that region is more significant or information-rich. Therefore, these areas may contain more vital information about category features, aiding in distinguishing between different labels. *E.g.*, in Figure 1 (b), when x-axis reaches 30-50, more knots were placed on the first mean series since its volatility increase.

Our final knot vector for defining the common B-spline basis functions should capture all the critical information. So, we need to merge the knots which are relatively dense in different categories into one vector as our final common knot vector. The merge algorithm is as follows: first, calculate the average distance between adjacent knots in the mean series of each category. Let K_i be the knot of vector for mean series of category i, $d_{ij} = K_i[j+1] - K_i[j]$ represents the distance between adjacent knots in category i, and $\bar{d}_i$ be the average distance between adjacent knots for category i:

$$\bar{d}_i = \frac{1}{n_i - 1} \sum_{j=1}^{n_i - 1} d_{i,j} \tag{5}$$

Then, use half of the average of these averages as a threshold. The threshold $Threshold$ is defined as half of the average of these average distances across all categories:

$$Threshold = \frac{1}{2} \cdot \frac{1}{m} \sum_{i=1}^{m} \bar{d}_i \tag{6}$$

The purpose of defining this threshold is to determine that, for all categories, if the distance between adjacent knots in a category is less than this threshold, it indicates that the category contains a substantial amount of information in this region, and thus this area should be integrated into the final common knot vector. Finally, for each category i, if the distance between two adjacent knots satisfies:

$$d_{i,j} < Threshold \tag{7}$$

Then the knots $K_i[j]$ and $K_i[j+1]$ are considered important and should be merged into the final knot vector. And the final common knot vector should be:

$$K_{final} = sort\left(\{K_i[j] \mid d_{i,j} < Threshold\}\right) \tag{8}$$

3.4 Smoothing

Given the knot vectors merged from above, as we mentioned in Equation 4, when the common B-spline basis functions are used, the only variable $\{P_i\}$ decide the shape of the B-spline curve. Then the coefficients $\{P_{ik}\}$of B-spline basis function, is set to represent the k-th input function in parameter space.

Follow this idea, we need to calculate the $\{P_{ik}\}$ of the B-spline functions.A set of n common B-spline basis functions $B_i(x)$ were used to fit the raw data. Assume K input curves, for each discrete observation $(x_{jk}, y_{jk})(j = 1..m,\ k = 1..K)$, where x_{jk} represents the independent variable, and y_{jk} the dependent variable. y_{jk} can be expressed as follows:

$$y_{jk} = \sum_{i=1}^{n} P_{ik} B_i\left(x_{jk}\right) + \varepsilon_{jk} \tag{9}$$

The least squares method is utilized to calculate P_{ik}, aiming to find a set of vectors P_{ik} to minimize ε_{jk}, which represents the error between the predicted and the actual observed value, and turns out to be an optimization problem:

$$min_P \sum_{j=1}^{m} \left(y_{jk} - \sum_{i=1}^{n} P_{ik} B_i\left(x_{jk}\right) \right)^2 \tag{10}$$

Converted to matrix view, the above optimization problem should be:

$$min_P \left\| Y - BP \right\|^2 \tag{11}$$

We can get P by solving the normal equation:

$$B^T BP = B^T Y \tag{12}$$

P is a matrix with dimension n*1, where n represents the number of B-spline basis functions applied in the curve fitting. This indicates n parameters were used to represent the m observations.

With the coefficient vector $\{P_{ik}\}$, the curve fitted by B-spline can be calculated by Equation 4, and one random "ao" fitting result is shown in Figure 2.

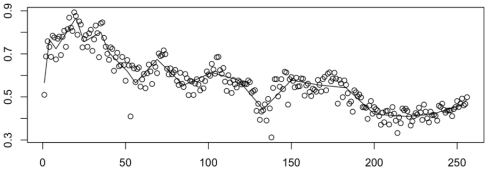


Figure 2: Raw data point and the corresponding continues curve

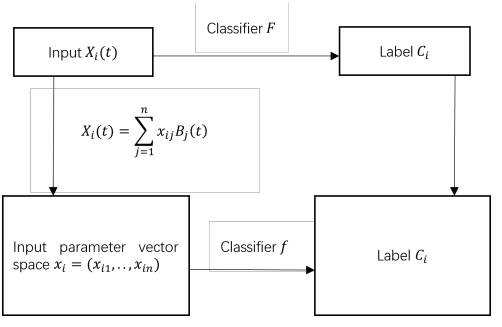


Figure 3: The process of Classification section

3.5 Classification

After the coefficient vector calculated from the last section, we use the coefficient vector $\{P_{ik}\}$ to represent the k-th input function, and the classification can be done in the parameter vector space.

The classification procedure is illustrated in Figure 3. That is, given input function $X_i(t)$, classifier F to identifies the corresponding label C_i can be expressed via f as follows:

$$C_i = F\left[X_i(t)\right] = F\left[\sum_{j=1}^{n} x_{ij}(t) B_j(t)\right] = f\left(x_{i1}, \ldots, x_{in}\right) \tag{13}$$

In general, the challenging task of defining classifier F in the infinite-dimensional input spaces is simplified by converting it into a manageable learning problem for the model f, which is a model in the finite-dimensional spaces of input parameters. Since the model f's learning problem exists in a traditional vector space, the classification issue can be addressed using the existing machine learning method, such as SVM or k-NN. And the classification results are displayed in next section.

4 Experimental Result

The performance of our approach was tested against 4 prior works. For every experiment, the performance depends on the accuracy classification rate. And also, we wrote down the number of B-spline bases we use in converting the raw input data into curves, which means the dimension reduced from the original to the number of bases in our experiment.

	Divide & Merge + SVM	RKVS+QDA [14]	RF [14]	k-NN [14]
Phoneme	**0.930±0.004**	0.927±0.005	0.924±0.006	0.911±0.004
Duo	**0.819±0.030**	0.815±0.015	0.810±0.015	0.803±0.015
Tecator	0.989±0.019	0.978±0.015	**0.990±0.012**	0.980±0.019
Fish	**0.894±0.056**	0.824±0.037	0.806±0.037	0.776±0.033

Table 1: Classification results

4.1 Phoneme Dataset

This data set is a part of TIMIT database. The dataset consists of a total of 4509 phonemes, categorized into five types: “ao”, “sh”, “iy”, “dcl” and “aa” (Figure 1 shows ’aa’ and ’ao’). Each data is a log-periodogram of length 256 corresponding to phonemes in 32ms duration.

The smoothing of one random normal sample “ao” is shown in Figure 2.

Table 1 includes the first experiment with Phoneme data set is a normal classification between 5 different phonemes. We compared with prior works same as their setting for experiments.

Delaigle and Hall [9] also used this data set, but restricted it to the use of the first 50 discretization points and to the binary case using the two most difficult phonemes (“aa”, “ao”) , we duplicate the experiment with same parameter and Table 1 shows the duo classification result.

In the two experiment of Phoneme data set, our method performed better with two different parameter settings. We used 23 and 10 B-spline bases and accomplished the dimension reduction for each input vector in both the tests from 256 to 23 and 50 to 10, respectively.

4.2 Tecator Dataset

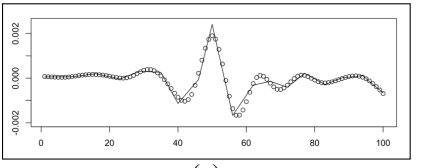
This data set has 215 spectra, each with a length of 100. Each data is the near infrared absorbance spectrum of a meat sample. It is classified by the fat content of the spectra.In order to amplify their difference, we use their second derivatives. After smoothing, one high fat content sample’s curve and raw data points were shown in the Figure 4 (a).

The result of the classification problem of Tecator data set is displayed in Table 1. Our method performs almost the same against two best prior works under two different experiment contexts. Also, we use 18 B-spline bases in this experiment and reduce the dimension from 100 to 18 within every input vector.

4.3 Fish Dataset

D.Lee [12] collects and uses the 7 classes of fish outline data with contour matching. We obtained this data from the UCR Time Series Classification Archive [3]. This dataset contains 350 observations and each has 463 raw data points.

We random picked one observation, and Figure 4 (b) is the smoothing result. Green lane is used to make the curve clear. And the classification rate was displayed in Table 1. Our method performed the best in this data set, almost 90 percents of classification rate, followed by RKVS+QDA suffered under 18 percents of mislabeling, and 33 bases were used in this experiment to reduce the dimension of input vector from 463 to 33.



(a) (b)

Figure 4: (a) The observation and fitted curve for one sample with high fat content; (b) The smoothed result for one random picked sample.

Divide & Merge + SVM (10 nbasis)	Distspace [7]
0.728±0.035	Around 0.70

Table 2: Classification result for MRI Image

4.4 Medical Image Dataset

This data set was also downloaded from the UCR Time Series Classification Archive [3], and it includes 9 different classes, together contain 547 observations and the length is 99. Table 2 shows the result of the classification problem of Medical Image data set.

Our method achieves a better result against Distspace in the Medical Image data set. In our experiment, the length of each input vector is only 10, instead of 99 data points in traditional machine learning methods.

5 Conclusions

We proposed a divide and merge method for functional data classification based on FDA. At the beginning, we divided the dataset according to the label and identify the knot vector for each dataset with the same label, and then apply the knot merge to build the knot vector for whole dataset with all labels in classification problems. Then the merged knot vector accomplished the fitting of B-spline for the raw data, and least square method was utilized to calculate the coefficient vector P in B-spline curve fitting. This function representation step encodes the raw data into a parameter space, which is smaller than the original input space. Subsequently, conventional machine learning techniques were used to address the classification problems within this parameter space. We tested our method on 4 real data sets. The fitted curves were displayed and illustrate we use the coefficient vector P calculated by merged knot vector to capture the function nature of these datasets. And the accuracy classification results show advantages in multiclass datasets (Phoneme, Fish, Medical Image).

6 Acknowledgment

The first author is sponsored by the University of Manchester and China Scholarship Council joint PhD studentship.

References

[1] Amparo Baíllo, Antonio Cuevas, and Ricardo Fraiman. Classification methods for functional data. 2010.

[2] Antonio Cuevas, Manuel Febrero, and Ricardo Fraiman. On the use of the bootstrap for estimating functions with functional data. *Computational statistics & data analysis*, 51(2):1063–1074, 2006.

[3] Hoang Anh Dau, Anthony Bagnall, Kaveh Kamgar, Chin-Chia Michael Yeh, Yan Zhu, Shaghayegh Gharghabi, Chotirat Ann Ratanamahatana, and Eamonn Keogh. The ucr time series archive. *IEEE/CAA Journal of Automatica Sinica*, 6(6):1293–1305, 2019.

[4] Aurore Delaigle and Peter Hall. Achieving near perfect classification for functional data. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 74(2): 267–286, 2012.

[5] Daniel Hlubinka, Irène Gijbels, Marek Omelka, and Stanislav Nagy. Integrated data depth for smooth functions and its application in supervised classification. *Computational Statistics*, 30:1011–1031, 2015.

[6] Xinyu Huang, Ziyang Pan, et al. A functional data classification model utilizing functional mahalanobis distance and regenerative kernel methods. *Journal of Electronics and Information Science*, 8(6):104–110, 2023.

[7] Mia Hubert, Peter Rousseeuw, and Pieter Segaert. Multivariate and functional classification using depth and distance. *Advances in Data Analysis and Classification*, 11: 445–466, 2017.

[8] Amy M Kwon, Ming Ouyang, and Andrew Cheng. Resampling-based classification using depth for functional curves. *Communications in Statistics-Simulation and Computation*, 45(9):3329–3338, 2016.

[9] Pascal Laube, Matthias O Franz, and Georg Umlauf. Learnt knot placement in b-spline curve approximation using support vector machines. *Computer Aided Geometric Design*, 62:104–116, 2018.

[10] Bin Li and Qingzhao Yu. Classification of functional data: A segmentation approach. *Computational Statistics & Data Analysis*, 52(10):4790–4800, 2008.

[11] Fusheng Liang, Ji Zhao, Shijun Ji, Cheng Fan, and Bing Zhang. A novel knot selection method for the error-bounded b-spline curve fitting of sampling points in the measuring process. *Measurement Science and Technology*, 28(6):065015, 2017.

[12] Changyi Park, Ja-Yong Koo, Sujong Kim, Insuk Sohn, and Jae Won Lee. Classification of gene functions using support vector machine for time-course gene expression data. *Computational Statistics & Data Analysis*, 52(5):2578–2587, 2008.

[13] Les Piegl and Wayne Tiller. *The NURBS book*. Springer Science & Business Media, 2012.

[14] Carlos Ramos-Carreño, José Luis Torrecilla, and Alberto Suárez. Classification of functional data: A comparative study. In *2022 21st IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 866–871. IEEE, 2022.

[15] J. Ramsay and B.W. Silverman. *Functional Data Analysis*. Springer Series in Statistics. Springer New York, 2006. ISBN 9780387227511. URL https://books.google.co.uk/books?id=REzuyz_V6OQC.

[16] J.O. Ramsay and B.W. Silverman. *Applied Functional Data Analysis: Methods and Case Studies*. Springer Series in Statistics. Springer New York, 2007. ISBN 9780387224657. URL https://books.google.co.uk/books?id=WE3SzeVEvDkC.

[17] Fabrice Rossi and Nathalie Villa. Support vector machine for functional data classification. *Neurocomputing*, 69(7-9):730–742, 2006.

[18] Isaac Jacob Schoenberg. Contributions to the problem of approximation of equidistant data by analytic functions. part b. on the problem of osculatory interpolation. a second class of analytic approximation formulae. *Quarterly of Applied Mathematics*, 4(2): 112–141, 1946.

[19] Carlo Sguera, Pedro Galeano, and Rosa Lillo. Spatial depth-based classification for functional data. *Test*, 23:725–750, 2014.

[20] Shahid Ullah and Caroline F Finch. Applications of functional data analysis: A systematic review. *BMC medical research methodology*, 13:1–12, 2013.

[21] Raine Yeh, Youssef SG Nashed, Tom Peterka, and Xavier Tricoche. Fast automatic knot placement method for accurate b-spline curve fitting. *Computer-aided design*, 128:102905, 2020.