Variation pattern classification of functional data

Shuhao JIAO1*, Ron D. FROSTIG2, and Hernando OMBAO1

1Statistics Program, King Abdullah University of Science and Technology, Saudi Arabia
2Department of Neurobiology and Behavior, University of California, Irvine, California, U.S.A.

Keywords and phrases: Dimension reduction; discriminative feature function; functional data analysis; nearest centroid classifier; second-moment structure.

Abstract: A new classification method for functional data is proposed in this article. This work is motivated by the need to identify features that discriminate between neurological conditions on which local field potentials (LFPs) were recorded. Regardless of the condition, these LFPs have zero mean, and thus the first moments of these random processes do not have discriminating power. We propose the variation pattern classification (VPC) method which employs the second-moment structure as the discriminating feature and uses the Hilbert–Schmidt norm to measure the discrepancy between the second-moment structure of different groups. The proposed VPC method is demonstrated to be sensitive to the discrepancy, potentially leading to a higher rate of classification. One important innovation lies in the dimension reduction where the VPC method adaptively determines the basis functions (discriminative feature functions) that account for the major discrepancy. In addition, the selected discriminative feature functions provide insights into the discrepancy between different groups because they reveal the features of variation pattern that differentiate groups. Consistency properties are established and, furthermore, simulation studies and the analysis of rat brain LFP trajectories empirically demonstrate the advantages and effectiveness of the proposed method.

Résumé: Le présent article propose une nouvelle méthode de classification de données fonctionnelles. Il est motivé par la nécessité d’identifier les caractéristiques capables de différencier entre les conditions neurologiques sous lesquelles les potentiels de champ local (LFPs) ont été enregistrés. Indépendamment de la condition neurologique, ces potentiels possèdent une moyenne nulle, ce qui prive les premiers moments de ces processus aléatoires de tout pouvoir discriminant. Les auteurs proposent une méthode de classification des profils de variation (VPC) qui utilise la structure du second moment comme caractéristique discriminante et applique la norme de Hilbert-Schmidt pour mesurer l’écart entre les structures du second moment au sein des différents groupes. Il s’avère que cette méthode VPC est sensible aux écarts et peut, par conséquent, donner lieu à un taux de classification plus élevé. Une innovation importante de cette approche réside dans la réduction de dimension car les fonctions de base (fonctions discriminantes) qui expliquent l’écart majeur sont déterminées de manière adaptative. En outre, les fonctions discriminantes sélectionnées fournissent des aperçus sur l’écart entre les différents groupes parce qu’elles révèlent les caractéristiques des profils de variation qui différencient les groupes. En plus d’établir quelques propriétés de convergence de la méthode proposée, les auteurs illustrent ses avantages et son efficacité en ayant recours à des simulations et à l’analyse des trajectoires LFP dans le cerveau des rats.

© 2022 Statistical Society of Canada / Société statistique du Canada
1. INTRODUCTION

1.1. Motivation from a Stroke Experiment

It is important for clinicians to be able to rapidly detect stroke onset in order to minimize the debilitating downstream effects of a stroke. Early detection gives the patients the best possible prognosis for quick recovery and minimal neurological damage. By contrast, late detection of a stroke is associated with poor prognosis including longer recovery times, and nonrecovery from profound effects on motor function, speech, and memory. In this article, we examine local field potentials (LFPs) that are recorded from rats during an experiment.

LFPs, regardless of clinical condition, are random fluctuations around zero. Thus, more interesting and discriminating features are captured by the second- or higher-order moments. The motivation for this work comes from the need to identify discriminating features of rat LFPs in a simulated stroke experiment. The goals of this article are the following:

1) To develop a statistical method for the discrimination between prestroke and poststroke onset brain signals. In this setting, there are training data available, which have known group labels (pre vs. poststroke onset). The goal here is to identify features that best separate the two classes of signals.

2) To develop a statistical method for classification of signals with unknown group labels (normal vs. abnormal or stroke in particular). We envision developing a method that can track brain signals online for the purpose of providing some warning for clinicians when the brain signals start to exhibit abnormal features.

1.2. Our Contributions and Existing Work

The main contribution of this work is a functional classifier based on the (lagged) second-moment operator under the setting where different groups of functional data have similar mean functions. The proposed method does not rely on any distributional assumption, and thus the classification procedure has broad potential applicability. It is noted that classification accuracy is influenced by two factors, the discrepancy between groups and background noise in the curve to be classified. It is often the case that, as more basis functions are used for discriminating different groups, the discrepancy typically becomes more pronounced. However, variability or uncertainty also increases. Therefore, it is not necessarily advantageous to incorporate many basis functions in discriminant analysis—if the employed basis functions have low power to discriminate between different classes. As an advantage, the proposed VPC method adaptively selects the basis functions that account for the major discrepancy between groups, and thus attenuates the nuisance effect from basis functions with low discriminating power.

In the past two decades, a variety of classification methods for functional data have been proposed. James & Hastie (2001), Preda, Saporta & Lévéder (2007), Shin (2008), and Delaigle & Hall (2012) studied functional linear discriminant analysis. Biau, Bunea & Wegkamp (2005) and Fromont & Tuleau (2006) studied k-nearest neighbour classification. Müller & Stadtmüller (2005) studied generalized functional linear models, which were used for classification in Leng & Müller (2006), and Li & Ghosal (2018) used multinomial logistic models for multiclass functional data classification. Glendinning & Herbert (2003), Chiu & Li (2007), and Song et al. (2008) studied functional principal component based classifiers. Delaigle & Hall (2013) and Dai, Müller & Yao (2017) studied a functional Bayesian quadratic classifier. Tian & James (2013) proposed an interpretable dimension reduction technique for functional data classification. Some work carefully studied the selection of basis functions in the classification problem. Saito & Coifman (1995, 1996) proposed a method to select the best discriminative functions from wavelet packets to extract local information for classification problems. Wang, Ray & Mallick (2007) and Fryzlewicz & Ombao (2009) employed wavelet methods. Huang, Ombao & Stoffer (2004) and
Ho et al. (2008) proposed discrimination procedures for nonstationary time series. The novelty of these works is that they select bases from the SLEX library that can best illuminate the difference between two or more classes of processes. Böhm et al. (2010) proposed a similar procedure for multivariate nonstationary processes. Some of these research studies incorporate the difference in second-moment structure in the classification, such as Huang, Ombao & Stoffer (2004), Chiou & Li (2007), Fryzlewicz & Ombao (2009), Delaigle & Hall (2013), and Dai, Müller & Yao (2017). However, in these methods, the discrepancy between covariance operator/matrix is accounted for by fixed basis functions or group-wise functional principal components, which do not necessarily capture the difference.

Compared with the existing methods, the proposed VPC method has the following advantages:
1. The VPC method is entirely data driven and nonparametric, making it applicable for a broad range of data and robust to potential model misspecification.
2. The VPC method selects, in a data adaptive manner, the sequence of orthonormal basis functions that accounts for most of the discrepancy of the (lagged) second-moment operators. Thus, the selected basis functions improve the classification accuracy and reveal the features with the ability to differentiate groups.
3. The VPC method takes account of the intra-curve information, which potentially provides important information discriminating groups.
4. The proposed framework can be applied to both independent and dependent functions.

The rest of the article is organized as follows. Section 2 presents the classification procedure and also displays the theoretical result that the classification performs asymptotically perfectly under some regularity conditions. In Section 3, we study the finite-sample properties of the classifier by simulations. In Section 4, the VPC method is implemented to classify the LFP epochs. Conclusions are made in Section 5. Technical proofs and additional real data analysis on phoneme data can be found in the Supplementary Material. The relevant code can be found on GitHub at https://github.com/DrJiaoSH/VPC.

2. MODEL, CONSISTENCY, AND ALGORITHM
2.1. General Setting and Preliminaries
Let \{X_k(t) : k \in \mathbb{N}, t \in [0, 1]\} be a set of random functions such that the realizations are elements of the Hilbert space \(L^2[0, 1]\), where the inner product is defined as \((x, y) = \int_0^1 x(t)y(t) dt\), and the norm is defined as \(\|x\|^2 = (x, x) = \int_0^1 x(t)^2 dt\). Assume \(E\|X\|^2 < \infty\), and define the mean function by \(\mu(t) = E\{X(t)\}\), and the covariance operator \(C(\cdot) : L^2[0, 1] \to L^2[0, 1]\) by \(C(\cdot) = \mathbb{E}\{X(\cdot) - \mu(\cdot)(X - \mu)\}\). The Hilbert–Schmidt norm of an operator \(\Phi\) is defined as: \(\|\Phi\|^2 = \sum_n \|\Phi_{n,j}\|^2\), where \(\Phi_{n,j} = \langle \Phi(e_j), e_j \rangle\) and \(\{e_j : j \in \mathbb{N}_+\}\) is a sequence of orthonormal basis functions. This norm does not depend on the choice of \(\{e_j : j \in \mathbb{N}_+\}\).

Consider a sequence of functions in \(L^2[0, 1]\) for each group \(\Pi_g\) (\(g = 0, 1\)), \(X_1^{(g)}(t), \ldots, X_{n_g}^{(g)}(t), g = 0, 1\), and \(n_0 + n_1 = n\), where \(g\) is the group index and the functions in either group are weakly stationary. Here we assume the mean function of each group is zero, say, \(\mu_g(t) = E\{X_k^{(g)}(t)\} = 0\) for \(g = 0, 1\). The zero-mean assumption is reasonable for many brain signals (e.g., electroencephalograms, local field potentials, and magnetoencephalograms), of which recordings always fluctuate around the zero line. We define the (auto)covariance operator at lag \(h\) as \(C_g^{(h)}(\cdot) = E\{X_k^{(g)}(X_{k+h}^{(g)}, \cdot)\}\) and \(C_g^{(h)}(\cdot) = E\{X_k^{(g)}(X_{k+h}^{(g)}, \cdot)\}\), where \(h \in \mathbb{N}\). In practice, \(\{C_g^{(h)}(\cdot) : h \in \mathbb{N}, g = 0, 1\}\) are estimated by \(\hat{C}_g^{(h)}(\cdot) = (n_g - h)^{-1} \sum_{k=1}^{n_g-h} X_k^{(g)} X_{k+h}^{(g)}\) and \(\hat{C}_g^{(h)}(\cdot) = (n_g - h)^{-1} \sum_{k=1}^{n_g-h} X_k^{(g)} X_{k+h}^{(g)}\).

Here we consider functions with zero means. If the mean functions are the same across groups but are not zero, discriminating (auto-)covariance operators is equivalent to discriminating
(lagged) second-moment operators due to the fact that $C_g^{(h)}(\cdot) = E\{X_k^{(g)}(X_{k+h}^{(g)} \cdot)\} - \mu_g(\mu_g \cdot)$. Thus the zero-mean assumption does not lead to any loss of generality.

### 2.2. Discriminative Feature Functions

A widely accepted metric to evaluate the difference between the two trace-class operators $C_0^{(0)}$ and $C_1^{(0)}$ is the Hilbert–Schmidt metric $\| \cdot \|_S$, defined as

$$\left\| C_0^{(0)} - C_1^{(0)} \right\|_S^2 = \sum_{i,j=1}^{\infty} \left\langle \left( C_0^{(0)} - C_1^{(0)} \right)(v_i), v_j \right\rangle^2.$$ 

Note that the evaluation of the metric does not rely on the selection of the orthonormal basis $\{v_j : j \geq 1\}$. However, it is important to select the basis functions that capture the discrepancy due to the decision rule that is now briefly described. Assume that $\{X_k^{(g)}(t) : k \in \mathbb{N}\}$ are the observed functions for group $g$. Suppose $Y(t) \in L^2[0,1]$ is a new function whose group membership is to be determined. Notationally, let $Y(\cdot)(t) = \langle Y, \cdot \rangle Y(t)$ be the one-dimensional compact operator driven by $Y(t)$. Our lag-0 centroid classifier assigns $Y$ to group $\Pi_g$ if

$$D_d \left( Y, C_g^{(0)} \right) < D_d \left( Y, C_{1-g}^{(0)} \right),$$

where $D_d$ is the Hilbert–Schmidt metric distance determined by $d$ orthonormal basis functions $\{v_j(t) : j = 1, \ldots, d\}$, say, $D_d \left( Y, C_g^{(0)} \right) = \sum_{i,j=1}^{d} \left\langle \left( Y(Y, v_j) - C_g^{(0)}(v_j) \right), v_j \right\rangle^2$.

Clearly $C_0^{(0)} - C_1^{(0)}$ is compact, and suppose the $C_0^{(0)} - C_1^{(0)}$ admits the compact spectral decomposition $(C_0^{(0)} - C_1^{(0)})(\cdot) = \sum_{i,j=1}^{\infty} \lambda_{0,i,j} \langle v_{0,i,j}, \cdot \rangle v_{0,i,j}$, and accordingly $(C_0^{(0)} - C_1^{(0)})(\cdot)^2 = \sum_{i,j=1}^{\infty} \lambda_{0,i,j}^2 \langle v_{0,i,j}, \cdot \rangle v_{0,i,j}$. By the nature of the Hilbert–Schmidt norm, $\|C_0^{(0)} - C_1^{(0)}\|_S^2 = \sum_{i=1}^{\infty} \lambda_{0,i,j}^2$. Therefore, most discrepancy is captured by the eigenfunctions associated with the large eigenvalues of the positive definite operator $(C_0^{(0)} - C_1^{(0)})(\cdot)^2$, and we propose to use the eigenfunctions associated with the nontrivial eigenvalues of $(C_0^{(0)} - C_1^{(0)})(\cdot)^2$, which are termed the discriminative feature functions. In the cases where the discrepancy cannot be explained by group-wise functional principal components, the classifier based on group-wise functional principal components cannot yield satisfactory results. In comparison, the discriminative feature functions are extracted from the difference operator $(C_0^{(0)} - C_1^{(0)})(\cdot)^2$ and are always able to capture the difference.

Similarly, as for the (auto)covariance operators, the eigenfunctions of $(C_0^{(h)} + C_1^{(-h)} - (C_0^{(h)} + C_1^{(-h)}))^2$ associated with its nontrivial eigenvalues account for most of the discrepancy $C_0^{(h)} + C_0^{(-h)} - (C_0^{(h)} + C_1^{(-h)})$. To find the most important discriminative basis functions, select the first $d_h$ eigenfunctions of the positive definite operator $R_h = (C_0^{(h)} + C_1^{(-h)} - (C_0^{(h)} + C_1^{(-h)}))^2$, associated with the first $d_h$ largest eigenvalues of $R_h$. The estimator of $R_h$ is $\hat{R}_h = (\hat{C}_0^{(h)} + \hat{C}_1^{(-h)} - (\hat{C}_0^{(h)} + \hat{C}_1^{(-h)})^2$. In the following, $\{\nu_{h,j}(t) : j = 1, \ldots, d_h\}$ denote the selected discriminative feature functions for the lag-$h$ (auto)covariance operator.

### 2.3. Classification Procedure

Assume that $\{Y_k(t) \in L^2[0,1] : k = 1, \ldots, p+1\}$ are consecutively collected from the same group, whose group memberships are to be predicted or determined jointly. Lag $p$ signifies the maximal incorporated lag. In some cases, it might be unrealistic to assume that the consecutive
samples \( \{ Y_{k}(t) : k = 1, \ldots, p + 1 \} \) come from the same class. However, in our application, the LFP signals remained abnormal after the stroke onset for a sufficiently long time, so that it is possible to collect consecutive functions in the poststroke onset state. Let \( \kappa^{(h)}_{g} = C^{(h)}_{g} + C^{(-h)}_{g} \), \( \hat{k}^{(h)}_{g} = \hat{C}^{(h)}_{g} + \hat{C}^{(-h)}_{g} \) and

\[
\hat{k}_{y,h}(\cdot) = \frac{1}{p + 1 - h} \sum_{k=1}^{p+1-h} Y_{k}(Y_{k+h},) + \frac{1}{p + 1 - h} \sum_{k=1}^{p+1-h} Y_{k+h}(Y_{k})
\]

In addition, we define \( \{ \hat{v}_{h,j} : j \geq 1 \} \) to be the empirical discriminative feature functions obtained from \( \hat{R}_{h} = (\hat{C}^{(h)} - \hat{C}^{(-h)} - (\hat{C}^{(h)} + \hat{C}^{(-h)})^2 \). The (auto)covariance operators at different lags may have varying levels of discriminating power, so we consider the weighted classifier. The procedure is summarized in Algorithm 1.

**Algorithm 1.** Classification algorithm

**Step 1.** Fix \( d_{h} \), obtain the eigenfunction of \( \hat{R}_{h} \), say, \( \{ \hat{v}_{h,j} : j = 1, \ldots, d_{h} \} \).

**Step 2.** Compute the scores \( \hat{S}^{h}_{g,i,j} = \langle \hat{k}^{(h)}_{g}, \hat{v}_{h,j} \rangle \) for \( i, j = 1, \ldots, d_{h} \).

**Step 3.** Compute

\[
\hat{D}_{g} = \sum_{h=0}^{p} W(h) \sum_{i,j=1}^{d_{h}} (\hat{S}^{h}_{g,i,j} - \langle \hat{k}_{y,h}(\hat{v}_{h,i}), \hat{v}_{h,j} \rangle)^2.
\]

If \( \hat{D}_{0} - \hat{D}_{1} < 0 \), classify \( Y_{1}, \ldots, Y_{p+1} \) to \( \Pi_{0} \), otherwise, classify them to \( \Pi_{1} \).

The weight \( W(h) \) depends on two factors: (1) the amplitude \( \| \hat{C}^{(h)} \|_{S} + \| \hat{C}^{(h)} \|_{S} \); and (2) the classification rate based on lag \( h \) only, \( P(h) \). The explicit expression of \( P(h) \) cannot be obtained here because no assumptions on distribution are made. However, an empirical version of \( P(h) \) can always be obtained by classifying the training set based on lag \( h \) only through some cross-validation procedure. In our application, \( W(h) = (\| \hat{C}^{(h)} \|_{S} + \| \hat{C}^{(h)} \|_{S})^{-1} \exp(\alpha P(h)) \), where \( \alpha \) is a tuning parameter and can be fixed by a cross-validation (CV) procedure. Here we briefly describe a feasible procedure (Monte-Carlo CV) to select \( \alpha \), the maximal lag \( p \), and \( P(h) \). A sequence of consecutive functions is randomly selected from the samples as the testing set. We use the rest of the samples to train the classifier based on a single lag \( h \) and obtain the empirical classification rate of the two groups. For a sufficient range of \( h \), repeat the procedure multiple times to obtain the average rate (over repetitions and groups) and employ it as \( P(h) \). Then apply the same procedure to obtain the empirical classification rates of the entire trained classifier that incorporates all \( p \) lags, where a set of prespecified candidate tuning parameters \( p \) and \( \alpha \) are tested. The resampling and classification steps are repeated multiple times to calculate the average classification rates under different pairs of \( p \) and \( \alpha \), and the tuning parameters corresponding to the largest average classification rate are selected.

**Remark 1.** One major challenge in classification is the unbalanced variance of \( X(t)X(s) \) across groups. Pronounced unequal levels of variance lead to unbalanced classifications, necessitating additional steps (e.g., scaling the functions by their norm and preclassifying by the amplitude). Scaling forces the variability of different groups to be at the same level and helps yield balanced classifications, but may lead to the loss of classification power if the amplitude is informative of discrimination. Thus, we propose that, if the variability of one group is significantly higher than that of the other one, we first implement preclassification based on the amplitude. Specifically,
we first set a threshold \( \tau \), which can be selected by CV, then given a curve \( Y(t) \) to be classified, if \( \|Y\| > \tau \), we classify the curve into the group with higher variation level, and the remaining functions are scaled to the unit norm and classified by the proposed method.

### 2.4. Classification of Multiple Groups

If we have \( G \) groups \( \Pi_1, \ldots, \Pi_G \) of functions, where \( G > 2 \), the VPC procedure can be easily extended to this case. In multiclass classification, we propose to do pairwise classification iteratively for different pairs of groups. More specifically, we first discriminate the first two groups \( \Pi_1 \) and \( \Pi_2 \). If curves with unknown group labels are classified into \( \Pi_1 \), then we do additional pairwise comparisons between \( \Pi_1 \) and \( \Pi_3 \). The procedure is finished when all necessary pairwise discriminations are conducted. Formally, the predicted index is defined as \( \hat{g} = \arg \min_{g=1,\ldots,G} \hat{D}_g \).

### 2.5. Consistency

Now we discuss the consistency property of the classifier. By Mercer’s theorem, assume that \( R_h(\cdot) = \sum_{j=1}^{\infty} \lambda_h \langle v_{h,j}, \cdot \rangle v_{h,j} \), where \( \lambda_{h1} > \lambda_{h2} > \cdots \). Notationally, let \( \sigma_{ij}^h = E \{ \gamma_{ij}^h(v_{h,i}, v_{h,j}) \} \) and \( \gamma_{ij}^h = \langle \kappa_{y,h}(v_{h,i}), v_{h,j} \rangle \). The following theorem gives an upper bound on the misclassification rate of the classifier with known discriminative feature functions.

**Theorem 1.** Assuming that there exists at least one \( h \) such that \( \| R_h \|_S > 0 \) and \( E \| X_k^{(y)} \|_4^4 < \infty \) for \( g = 0, 1 \), the misclassification rate satisfies

\[
P \left( \Pi_{1-g} \mid \Pi_g, \{ v_{h,j} : h \geq 0, j \geq 1 \} \right) \leq \frac{4 \sum_{h=0}^{P} \left( W(h) \sum_{j=1}^{d_h} \sigma_{ij}^h \right)}{\sum_{h=0}^{P} \left( W(h) \sum_{j=1}^{d_h} \lambda_{hj} \right)} \wedge 1.
\]

In this upper bound, the numerator is the variance, and the denominator is the discrepancy evaluated by the discriminative feature functions. The result below follows.

**Corollary 1.** Assume that the conditions in Theorem 1 hold. If

\[
\sum_{h=0}^{P} \left( W(h) \sum_{j=1}^{d_h} \lambda_{hj} \right) / \sum_{h=0}^{P} \left( W(h) \sum_{j=1}^{d_h} \sigma_{ij}^h \right) \rightarrow \infty,
\]

then the classification tends to be perfect as \( d_h, p \rightarrow \infty \).

This corollary states that if the discrepancy is large enough compared with the variance, perfect classification can be achieved if a sufficient number of lags and discriminative feature functions are incorporated.

The problem is that the discriminative feature functions \( \{ v_{h,j} : h \geq 0, j \geq 1 \} \) and \( \{ \kappa_{y}^{(h)} : h \geq 0, g = 0, 1 \} \) are unknown and should be estimated from the samples, and it is impossible to incorporate all lags. If a huge amount of discriminative basis functions and lags were incorporated, the estimation error could reduce the classification power. However, under some regularity conditions, perfect classification can still be achieved as the sample size increases to infinity. To establish the consistency of the estimated classifier, several assumptions are introduced below.

**A1** For each \( 0 \leq h \leq p_n \), \( R_h(\cdot) = \sum_{j=1}^{\infty} \lambda_{hj} \langle v_{h,j}, \cdot \rangle v_{h,j} \), where \( \lambda_{h1} > \lambda_{h2} > \cdots \) and \( \sum_j \lambda_{hj} < \infty \), and there exists \( \alpha_h > 2 \), such that \( \theta_{hj} = \lambda_{hj} - \lambda_{h,j+1} \geq \text{const.} j^{-\alpha_h} \). Here, “const.” signifies a constant not related to \( h \).
(A2) \( p_n \) and \( d_{h,n} \) are selected such that \( p_n/n \to 0 \) and \( d_{h,n} = n^{1/\tau_h} \), and that
\[
\frac{p_n}{\tau_h} \sum_{h=0}^{p_n} W(h) n^{(2d_{h,n}+1)/\tau_h-1} \to 0,
\]
\[
\frac{p_n}{\tau_h} \sum_{h=0}^{p_n} W(h) n^{1/\tau_h-1/2} \to 0.
\]

(A3) The process \( \{X_k^{(g)}(t) : k \geq 1\} \) is weakly dependent such that
\[
E \left| \hat{C}_g^{(h)}(t) - C_g^{(h)}(t) \right|^2 \leq O((n_g - h)^{-q})
\]
for any \( h, g, \) and \( q = 1, 2, n_0/n_1 \to \eta \) where \( \eta \neq 0, \infty. \)

Assumption (A1) guarantees the identifiability of the discriminative feature functions of each lag \( h \) and \( \alpha_h > 2 \) comes from \( \sum \lambda_{hj} < \infty. \) Assumption (A2) assures the selection of \( p_n \) and \( d_{h,n} \) not leading to an overly large estimation error. In Assumption (A3), Hörmann & Kokoszka (2010) have shown
\[
E \left| \hat{C}_g^{(h)}(t) - C_g^{(h)}(t) \right|^2 = O(n^{-1})
\]
for \( L^4 - m \)-approximable process, and similar arguments can be employed to extend this result to \( h \neq 0 \) cases. Note that
\[
E \left| \hat{C}_g^{(h)}(t) - C_g^{(h)}(t) \right|^2 \leq O((n_g - h)^{-q})
\]
always holds for \( m \)-dependent processes.

Theorem 2. Assuming that (A1)–(A3) and the assumptions in Theorem 1 hold, if
\[
\frac{p_n}{\tau_h} \sum_{h=0}^{p_n} \left( W(h) \sum_{j=1}^{d_{h,n}} \lambda_{hj} \right) \left/ \sum_{h=0}^{p_n} \left( W(h) \sum_{i,j=1}^{d_{h,n}} \sigma_{ij}^p \right) \right. \to \infty,
\]
then \( P(\Pi_{1-g} \mid \Pi_g, \{ \hat{v}_{h,j}, : h \geq 0, j \geq 1 \}) \to 0, \) as \( n \to \infty. \)

Theorem 2 illustrates that under some regularity conditions, the classification based on the estimated discriminative feature functions still approaches perfection.

3. SIMULATION

3.1. Classification of Functional Moving Average Processes

To study the finite-sample performance of the proposed method, we employed our proposed method to classify two FMA(3) processes, which follow the recursive equation
\[
X_k^{(g)}(t) = \epsilon_k(t) + \sum_{k=1}^{3} \Psi_{g,\epsilon}(\epsilon_{k-j}) \Psi_{g,\epsilon}(t), \quad (\epsilon_k(t) : k \geq 1) \text{ are zero-mean i.i.d innovation functions across } k.
\]
All the functions are simulated with the first 21 Fourier basis functions, which are denoted here by a functional vector \( F(t) = (F_1(t), \ldots, F_{21}(t))' \). Two templates \( \Psi_g(\cdot), g = 0, 1 \) are created for the coefficient operators, specifically, \( \psi_{g,\epsilon}(\cdot) = a_{\epsilon} \Psi_{g,\epsilon}(\cdot) \). Here we set \( a_{\epsilon} = 0.4 \). The template coefficient operators admit the basis representation \( \Psi_g(\cdot) = F^T K_g \Psi(\cdot) \), where \( K_g \) is a 21 × 21 matrix and each element is sampled from the normal distribution \( K_{g,ij} \sim \mathcal{N}(0, \sigma_g^2) \), where
\[
\sigma_0 = (1, 1_5 \otimes (0.8, 0.8, 1, 1)), \quad \sigma_1 = (1, 1_5 \otimes (1, 1, 0.8, 0.8)),
\]
and \( 1_5 = (1, 1, 1, 1, 1) \). We simulate 50, 100, or 600 functions for each group as the training set, and simulate another 100 functions for the testing set to obtain the empirical \( P(h) \) and select \( \alpha \) by the CV procedure described in Section 2.3. The dimension \( d_{h,n} \) is selected as the smallest dimension satisfying \( \sum_{j=1}^{d_{h,n}} \lambda_{hj} / \sum_{j \geq 1} \lambda_{hj} \geq 90\% \).

The candidate values for maximal lag \( p \) under consideration are 0, 1, 2, 3, 4. Under each setting, 100 sets of new functions (\( p+1 \) consecutive simulated functions in each set) are classified and the classification rates are computed for comparison. Under each simulation setup, the simulation runs are repeated 200 times and the average classification rates and the
The standard errors of the empirical classification rates are displayed in Table 1. Here, we set \( W(h) = \left( \| \widehat{C}_0^{(h)} \|_S + \| \widehat{C}_1^{(h)} \|_S \right)^{-1} \exp(10P(h)) \).

The optimal classification performance is typically reached at \( p = 3 \). This tendency is because the discrepancy between the two FMA processes concentrates on the first 3 lags, and two curves become independent as the lag is greater than 3.

### 3.2. Comparison with Other Methods

For comparison, the three functional methods under consideration are as follows: (1) the proposed method (VPC); (2) the projection method (denoted by PJ, Chou & Li, 2007); (3) the functional quadratic classifier (denoted by FQC, Delaigle & Hall, 2013, and Dai, Müller & Yao, 2017). Since the competitor methods do not incorporate the dependence across curves, we only consider lag 0 in the comparison. We simulated two classes \((n_1 = n_2 = 200)\) of independent functions with 24 B-spline basis functions \( \{ B_1(t), \ldots, B_{24}(t) \} \) over the unit interval \([0, 1]\). The functions in each group admit the following basis representation

\[
X_k(g) \equiv \sum_{j=1}^{24} \xi_{k,j}^{(g)} B_j(t), \quad g = 0, 1.
\]

The scores of the two groups follow two different normal distributions respectively \( \{ \xi_{k,j}^{(g)} : j = 1, \ldots, 24 \} \sim \mathcal{N}(0, \sigma^b) \), where

\[
\sqrt{\sigma^b_0} = I_3 \otimes \text{diag}(b, b, b, b, b, a, b, b), \quad \sqrt{\sigma^b_1} = I_3 \otimes \text{diag}(b, b, b, b, b, a, b, b),
\]

\( I_3 \) is the 3 \times 3 identity matrix and \( a \) and \( b \) are two constants satisfying \( a^2 + 7b^2 = 100 \). Clearly, a larger value of \( a \) leads to a more substantial discrepancy between the two groups. We simulated 200 curves for each group and then split these curves into 100 curves for training and 100 for testing. The classification procedure was repeated 200 times. The average classification rates

| \( p \) | \( n \) | \( \Pi_0 \) | \( \Pi_1 \) | \( \Pi_0 \) | \( \Pi_1 \) | \( \Pi_0 \) | \( \Pi_1 \) |
|---|---|---|---|---|---|---|---|
| 0 | 50 | 0.826 | 0.807 | 0.889 | 0.872 | 0.950 | 0.946 |
|  |  | (0.071) | (0.068) | (0.051) | (0.058) | (0.028) | (0.032) |
| 1 | 100 | 0.843 | 0.827 | 0.907 | 0.892 | 0.962 | 0.962 |
|  |  | (0.062) | (0.057) | (0.042) | (0.047) | (0.022) | (0.023) |
| 2 | 600 | 0.847 | 0.834 | 0.909 | 0.899 | 0.964 | 0.964 |
|  |  | (0.062) | (0.055) | (0.041) | (0.045) | (0.020) | (0.023) |
| 3 | 200 | 0.850 | 0.839 | 0.912 | 0.904 | 0.966 | 0.966 |
|  |  | (0.062) | (0.057) | (0.040) | (0.043) | (0.018) | (0.021) |
| 4 | 50 | 0.848 | 0.838 | 0.912 | 0.903 | 0.966 | 0.966 |
|  |  | (0.062) | (0.056) | (0.040) | (0.043) | (0.018) | (0.022) |
Table 2: Average classification rates and the corresponding standard errors (in parentheses).

| $a^2$ | Methods | VPC | PJ | FQC |
|-------|---------|-----|----|-----|
|       | Group   | $\Pi_0$ | $\Pi_1$ | $\Pi_0$ | $\Pi_1$ | $\Pi_0$ | $\Pi_1$ |
| 20    | ACR     | 0.55 (0.08) | 0.55 (0.08) | 0.53 (0.11) | 0.52 (0.11) | 0.51 (0.07) | 0.51 (0.07) |
| 40    | ACR     | 0.78 (0.09) | 0.79 (0.09) | 0.68 (0.10) | 0.70 (0.10) | 0.65 (0.08) | 0.64 (0.09) |
| 60    | ACR     | 0.89 (0.08) | 0.88 (0.08) | 0.85 (0.06) | 0.84 (0.006) | 0.81 (0.08) | 0.80 (0.08) |
| 80    | ACR     | 0.94 (0.07) | 0.95 (0.07) | 0.96 (0.03) | 0.96 (0.03) | 0.93 (0.05) | 0.93 (0.05) |

(ACRs) for each group are presented in Table 2, where the standard errors of the empirical classification rates are shown in parentheses. The simulation results demonstrate that the projection method and the functional quadratic classifier are inferior to the VPC method even though they also incorporate the covariance operator in the classification procedure, especially when the discrepancy is not substantial. One explanation for this suboptimal performance is that the methods use group-wise principal components that are not guaranteed to capture the discrepancy.

3.3. Discriminative Feature Functions

To show the necessity of selecting the basis function with the most discriminative power, and how the discriminative feature functions discriminate different groups, we simulate two groups of independent functions with both common and uncommon components. Two hundred functions are simulated with 21 Fourier bases for training, and another 100 functions are simulated in the same way for testing. All the simulated functions have the following basis representation $X_k^{(g)}(t) = \sum_{j=1}^{21} \psi_{kj}^{(g)} F_j(t)$, where $F_j(t)$ is the $j$th Fourier basis function. The score vectors of the two groups follow the following multivariate normal distribution, $\{ \psi_{kj}^{(g)} : j = 1, \ldots, 21 \} \sim \mathcal{N}(0, \Sigma_g)$, where $\Sigma_g$ is a diagonal matrix with diagonal elements $\delta_k$.

Two settings of $\{ \delta_k : g = 0, 1 \}$ were considered here, namely,

- Setting 1: $\delta_0 = (1, 1, 1, 0, 0, 1, 1, \ldots, 1, 1)$, $\delta_1 = (1, 0, 0, 1, 1, 1, \ldots, 1, 1)$,
- Setting 2: $\delta_0 = (1, 1, 0, 1, 0, 1, 0, \ldots, 1)$, $\delta_1 = (1, 0, 1, 0, 1, 0, 1, \ldots, 1)$.
We applied the VPC to classify the 100 functions in the testing group with finite dimensions \(d = 1, \ldots, 9\) and without dimension reduction \(d = \infty\), and repeated this procedure 200 times. Figure 1 displays the average classification rate of the two groups in different settings and the corresponding selected discriminative feature functions.

The two groups are differentiated into four dimensions in the first setting, and are differentiated into six dimensions in the second setting. Figure 1 shows that the classification rate first increases and then decreases as \(d\) increases. The optimal performance is achieved as \(d = 4\) and \(d = 6\) for settings 1 and 2, respectively. Another interesting point is that the discriminative feature functions are similar to those Fourier basis functions that differentiate the two groups, which provide insights on the discrepancy.

4. ANALYSIS OF RAT LOCAL FIELD POTENTIALS

4.1. Description of the LFP Data and Preprocessing

The VPC method was used to study the impact of a shock (simulated stroke by occlusion) on the functional architecture of a rat brain using the local field potential data. Here, LFPs were recorded from multi-tetrodes, implanted on the rat cortex, continuously over a period of 6 h (see Figure 2). In this article, we analyzed a 10-min window (5 min before the occlusion and 5 min after the occlusion).

From these micro-tetrodes, LFPs were recorded at the rate of 1000 observations per second (thus there are \(T = 1000\) time points per 1-s epoch). A total of \(r = 600\) epochs were recorded.
Midway through this period (at epoch \( r = 300 \)), a stroke was mechanically induced on the rat by clamping the medial cerebral artery. We extracted the component in different frequency bands from each epoch and used the filtered trajectories in six different frequency bands \((1) 1–4 \text{ Hz}, (2) 5–9 \text{ Hz}, (3) 10–14 \text{ Hz}, (4) 15–19 \text{ Hz}, (5) 20–24 \text{ Hz}, \) and \((6) 25–30 \text{ Hz}) for classification. The filtered trajectories are obtained by functional smoothing with Fourier bases. For example, to extract the \( k \)-Hz component, one can smooth the trajectories with the following two Fourier basis functions

\[
\left\{ \begin{array}{ll}
\sqrt{2} \cos(2\pi kt), & t \in [0, 1], \\
\sqrt{2} \sin(2\pi kt), & t \in [0, 1].
\end{array} \right.
\]

The goals of the analysis are (1) to identify features that best differentiate the prestroke from the poststroke onset signals using training data (where group labels are known) and (2) to classify a future brain signal with an unknown label.

For this particular dataset, the discrepancy is expressed primarily at lag 0 and thus we did not consider other lags. Data visualization reveals several outliers with extreme amplitude in the poststroke onset group, which can be easily identified as poststroke onset epochs, so we removed them from the samples. All the remaining functions (epochs) are scaled to unit norm to achieve a balanced classification.

4.2. Change-point Detection and Classification with Change points

Since brain signals are nonstationary within each group, we considered the structural breaks in the classification. Our approach is to approximate the nonstationary functional sequence as a concatenation of a piecewise-stationary functional sequence. Suppose there are \( L_0 \) and \( L_1 \) structural breaks in the covariance operator, dividing the entire sequence of the two groups \((g = 0, 1)\) into \( L_0 + 1 \) and \( L_1 + 1 \) stationary segments respectively. Hence, the covariance operator \( C^{(0)}_{g,k} (\cdot) = E \{ \langle X^{(g)}_k (\cdot) X^{(g)}_k (\cdot) \rangle \} \) is modelled as follows, \( C^{(0)}_{g,k} (\cdot) = \sum_{\ell=1}^{L_0+1} \tilde{C}^{(0)}_{g,k} (\cdot), \) where \( \tilde{C}^{(0)}_{g,k} (\cdot) \) is the covariance operator of the \( \ell \)th quasi-stationary segment of group \( g \), and \( \tilde{C}^{(0)}_{g,k} (\cdot) = 1 \) if \( k \) lies in the \( \ell \)th segment of group \( g \) and \( \tilde{C}^{(0)}_{g,k} (\cdot) = 0 \) otherwise.

Here we applied the break point detection method developed in Jiao, Frostig & Ombao (2022). Binary segmentation was applied to determine all pronounced break points. The break points of the 32 micro-tetrodes were searched simultaneously, and a set of common break points was found for the prestroke/poststroke onset epoch sequences of the 32 tetrodes. Let \( X^{(g)}_{k_i} (\Omega, t) \) be the \( \Omega \)-frequency component of the \( k \)th epoch obtained from the \( i \)th tetrode, and define

\[
Z^{(g)}_{k_i} (\Omega, t, s) = \sum_{i=1}^{32} X^{(g)}_{k_i} (\Omega, t) X^{(g)}_{k_i} (\Omega, s).
\]

DOI: 10.1002/cjs.11738

The Canadian Journal of Statistics / La revue canadienne de statistique
The break detection method was applied to the sequence $Z^{(g)}_k(\Omega, t, s)$. Under significance level $\alpha = 0.05$, several break points were detected in different frequency bands (see Figure 3). The epochs in the same local stationary segment are marked in the same colour. For each local stationary block, we concatenated the 32 tetrodes’ epochs by stacking the observations and obtained the estimated covariance function of the concatenated epochs, which serves as the classification feature. The covariance functions of the concatenated epochs are estimated for all segments. Given a new concatenated function $Y(t)$ with unknown group label, first find the covariance operator closest to $\mathcal{Y}(\cdot) = Y(Y, \cdot)$ within each group, denoted as $\hat{C}^{(0)}_{g, opt}(\cdot)$ and $\hat{C}^{(0)}_{1, opt}(\cdot)$. Then $Y(t)$ is assigned to group $g$ if the functional operator $\hat{C}^{(0)}_{g, opt}(\cdot)$ is closer to $\mathcal{Y}(\cdot)$. See Algorithm 2 for a summary of the classification procedure.

Algorithm 2. Classification algorithm with change points

Step 1. For each group $(g = 0, 1)$, apply break point detection method (e.g., Jiao, Frostig & Ombao (2022)) to find the structural break points in the covariance operator, and estimate the covariance operator for each local stationary subsequence.

Step 2. For each group $(g = 0, 1)$, find the estimated covariance operator $\hat{C}^{(0)}(\cdot)$ that is closest to $\mathcal{Y}(\cdot)$ with respect to the Hilbert–Schmidt norm. Fix $d$, and obtain the eigenfunctions $(\hat{\nu}_{opt,i} : j = 1, \ldots, d)$ of the functional operator $(\hat{C}^{(0)}_{g, opt} - \hat{C}^{(0)}_{1, opt})^2(\cdot)$ associated with its $d$ largest eigenvalues.

Step 3. Compute the scores $\hat{S}^{(g)}_{opt,i,j} = \langle \hat{C}^{(0)}_{g, opt}(\hat{\nu}_{opt,i}), \hat{\nu}_{opt,j} \rangle$ for $i, j = 1, \ldots, d$.

Step 4. Compute

$$\hat{D}_{g, opt} = \sum_{i,j=1}^d (\hat{S}^{(g)}_{opt,i,j} - \langle Y, \hat{\nu}_{opt,i} \rangle \langle Y, \hat{\nu}_{opt,j} \rangle)^2.$$ 

If $\hat{D}_{0, opt} - \hat{D}_{1, opt} < 0$, then classify $Y$ to $\Pi_0$, otherwise, classify $Y$ to $\Pi_1$.

4.3. Discriminative Feature Functions

To check the overall discriminative features between the two different brain states, the entire prestroke and poststroke onset epochs were employed to calculate the most discriminative feature.
functions for different frequency bands (1–4, 5–9, 10–14, 15–19, 20–24, and 25–30 Hz). Figure 4 displays the first discriminative feature functions that best explain the discrepancy between the covariance operators of the concatenated epochs for different frequency bands. Each function has 32 equal-length blocks corresponding to 32 tetrodes.

The discriminative functions are informative. First, they reveal the tetrodes that have low power of discrimination. The blocks with low amplitude indicate low power of discrimination of the corresponding tetrodes. For example, Figure 4 shows that the discriminative feature functions over blocks 9, 10, 13, 17, and 27 are comparatively flat, which indicates that the discrepancies between the two brain states across these tetrodes are comparatively trivial. Second, they reveal the period in the epochs that best differentiates the two states. An interesting finding is that for lower frequency bands (1–9 Hz), the discrepancy is mainly present in the middle part of the epochs, and for higher frequency bands (10–24 Hz), the discrepancy is more distributed across the epochs.
bands (10–30 Hz), the discrepancy is mainly present near the starting and ending points of the epochs.

4.4. Classification Performance

To check the discrimination power of different tetrodes, we applied the VPC method to each tetrode separately across different frequency bands, and applied a Monte-Carlo procedure. Specifically, in each step, 250 curves are randomly selected as the training set and the remaining curves are treated as the testing set. The Monte-Carlo procedure was repeated 200 times, and the average classification rates (ACRs) are displayed in Figure 5.

Each subfigure of Figure 5 presents the tetrode-wise average classification rates of one frequency band. An interesting phenomenon is that the $\delta$-band (1–4 Hz) has the highest power of discrimination. The final classification was conducted on the concatenated epoch trajectories. The same Monte-Carlo procedure was repeated. The resulting average classification rate of different frequency bands is shown in Table 3 together with the standard errors of the empirical classification rates. From the table, it is clear that the classification performance of VPC is nearly perfect over 1–9 Hz, and worsens as the frequency increases.

Figure 5: The average classification rate based on each individual tetrode. The first row corresponds to tetrodes 1–8, the second row corresponds to tetrodes 9–16, etc.
TABLE 3: Average classification rates and the standard errors (in parentheses) across different frequency bands.

| Frequency | 1–4 Hz       | 5–9 Hz       | 10–14 Hz      | 15–19 Hz      | 20–24 Hz      | 25–30 Hz      |
|-----------|--------------|--------------|---------------|--------------|--------------|--------------|
| State     | Π₀           | Π₁           | Π₀           | Π₁           | Π₀           | Π₁           | Π₀           | Π₁           | Π₀           | Π₁           | Π₀           | Π₁           |
| ACR       | 0.990        | 0.999        | 0.980        | 0.975        | 0.874        | 0.970        | 0.898        | 0.961        | 0.906        | 0.919        | 0.841        | 0.870        |
| SE        | (0.015)      | (0.002)      | (0.017)      | (0.020)      | (0.047)      | (0.026)      | (0.043)      | (0.027)      | (0.041)      | (0.043)      | (0.051)      | (0.045)      |

5. CONCLUSION

This article proposes a general functional approach (VPC) to discriminate between and classify different groups of functions with similar means. The VPC method takes advantage of the divergence of second-moment structures. The comparison between groups is constrained to the subspace spanned by the discriminative feature functions that account for most of the divergence. The VPC method has a built-in dimension reduction step that extracts the major information of discrepancy. Therefore, it overcomes the disadvantage of using group-wise functional principal components, which only explain variation for each individual group, but do not explain the difference between groups. In addition to improving the classification effectiveness, the discriminative feature functions also reveal the features that differentiate groups and thus provide insights for classification. We have shown that the proposed classifier works nearly perfectly under some regularity conditions.

In the presented application, our methodology provides a general picture of the discrepancy between prestroke and poststroke onset epochs. Specifically, the method reveals the tetrodes, frequencies, and features of epochs that effectively illuminate the discrepancy between different brain states. The VPC has potential medical applications because, if implemented in an online setting, it can help quickly detect the onset of stroke-related abnormal brain electrical rhythms and hence minimize the debilitating effects of a stroke. The method also has a wide range of applicability including speech signal classification, which is demonstrated in the Supplementary Material.

ACKNOWLEDGEMENT

We are grateful to the associate editor and two referees for their comments and suggestions that led to significant improvement of the article.

REFERENCES

Biau, G., Bunea, F., & Wegkamp, M. H. (2005). Functional classification in Hilbert spaces. *IEEE Transactions on Information Theory*, 51, 2163–2172.

Böhm, H., Ombao, H., von Sachs, R., & Sanes, J. (2010). Classification of multivariate non-stationary signals: The SLEX-shrinkage approach. *Journal of Statistical Planning and Inference*, 140, 3754–3763.

Chiou, J.-M. & Li, P.-L. (2007). Functional clustering and identifying substructures of longitudinal data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 69, 679–699.

DOI: 10.1002/cjs.11738
Dai, X., Müller, H. G., & Yao, F. (2017). Optimal Bayes classifiers for functional data and density ratios. *Biometrika*, 104, 545–560.

Delaigle, A. & Hall, P. (2012). Achieving near perfect classification for functional data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 74, 267–286.

Delaigle, A. & Hall, P. (2013). Classification using censored functional data. *Journal of the American Statistical Association*, 108, 1269–1283.

Fromont, M. & Tuleau, C. (2006). Functional classification with margin conditions. In *International Conference on Computational Learning Theory*, Springer, Berlin, Heidelberg, 94–108.

Fryzlewicz, P. & Ombao, H. (2009). Consistent classification of non-stationary time series using stochastic wavelet representations. *Journal of the American Statistical Association*, 104, 299–312.

Glendinning, R. H. & Herbert, R. A. (2003). Shape classification using smooth principal components. *Pattern Recognition Letters*, 24, 2021–2030.

Ho, M.h. R., Ombao, H., Edgar, J. C., Cañive, J. M., & Miller, G. A. (2008). Time–frequency discriminant analysis of MEG signals. *Neuroimage*, 40, 174–186.

Hörmann, S. & Kokoszka, P. (2010). Weakly dependent functional data. *The Annals of Statistics*, 38, 1845–1884.

Huang, H.-Y., Ombao, H., & Stoffer, D. S. (2004). Discrimination and classification of nonstationary time series using the SLEX model. *Journal of the American Statistical Association*, 99, 763–774.

James, G. M. & Hastie, T. J. (2001). Functional linear discriminant analysis for irregularly sampled curves. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63, 533–550.

Jiao, S., Frostig, R., & Ombao, H. (2022). Break point detection for functional covariance. *Scandinavian Journal of Statistics*, 1–36. https://doi.org/10.1111/sjos.12589

Leng, X. & Müller, H.-G. (2006). Classification using functional data analysis for temporal gene expression data. *Bioinformatics*, 22, 68–76.

Li, X. & Ghosal, S. (2018). Bayesian classification of multiclass functional data. *Electronic Journal of Statistics*, 12, 4669–4696.

Müller, H.-G. & Stadtmüller, U. (2005). Generalized functional linear models. *The Annals of Statistics*, 33, 774–805.

Preda, C., Saporta, G., & Lévéder, C. (2007). PLS classification of functional data. *Computational Statistics*, 22, 223–235.

Saito, N. & Coifman, R. (1996). On local feature extraction for signal classification. *Zeitschrift fur Angewandte Mathematik und Mechanik*, 76, 453–456.

Saito, N. & Coifman, R. R. (1995). Local discriminant bases and their applications. *Journal of Mathematical Imaging and Vision*, 5, 337–358.

Shin, H. (2008). An extension of Fisher’s discriminant analysis for stochastic processes. *Journal of Multivariate Analysis*, 99, 1191–1216.

Song, J., Deng, W., Lee, H., & Kwon, D. (2008). Optimal classification for time-course gene expression data using functional data analysis. *Computational Biology and Chemistry*, 32, 426–432.

Tian, T. S. & James, G. M. (2013). Interpretable dimension reduction for classifying functional data. *Computational Statistics & Data Analysis*, 57, 282–296.

Wang, X., Ray, S., & Mallick, B. K. (2007). Bayesian curve classification using wavelets. *Journal of the American Statistical Association*, 102, 962–973.

Wann, E. G. (2017). *Large-scale spatiotemporal neuronal activity dynamics predict cortical viability in a rodent model of ischemic stroke*. Ph.D. thesis, UC Irvine.

Received 5 October 2021
Accepted 20 June 2022