Statistics
in Medicine WILEY

Pooling random forest and functional data analysis for biomedical signals supervised classification: Theory and application to electrocardiogram data

Fabrizio Maturo | **Rosanna Verde**

Department of Mathematics and Physics, University of Campania Luigi Vanvitelli, Caserta, Italy

Correspondence
Fabrizio Maturo, Department of Mathematics and Physics, University of Campania Luigi Vanvitelli, Viale Abramo Lincoln, 5, 81100 Caserta, CE, Italy.
Email: fabrizio.maturo@unicampania.it

Abstract

Scientific progress has contributed to creating many devices to gather vast amounts of biomedical data over time. The goal of these devices is generally to monitor people's health conditions, diagnose, and prevent patients' diseases, for example, to discover cardiovascular disorders or predict epileptic seizures. A common way of investigating these data is classification, but these instruments generate signals often characterized by high dimensionality. Learning from these data is definitely a challenging task due to many issues, for example, the trade-off between complexity and accuracy and the course of dimensionality. This study proposes a supervised classification method based on the joint use of functional data analysis, classification trees, and random forest to deal with massive biomedical data recorded over time. For this purpose, this research suggests different original tools to extract features and train functional classifiers, interpret the classification rules, assess leaves' quality and composition, avoid the classical drawbacks due to the COD, and improve the accuracy of the functional classifiers. Focusing on ECG data as a possible example, the final purpose of this study is to offer an original approach to identify and classify patients at risk using different types of biomedical signals. The results confirm that this line of research is exciting; indeed, the interpretative tools show evidence to be very useful for understanding classification rules. Furthermore, the performance of the proposed functional classifier, in terms of accuracy, is excellent because the latter breaks the previous classification record regarding a well-known ECG dataset.

KEYWORDS
functional between groups variability, functional between leaves variability, functional classification trees, functional data analysis, functional random forest

1 | INTRODUCTION

In recent decades, instruments to collect massive volumes of biomedical data have evolved due to technological innovations. Therefore, today, many medical devices can record different types of signals to monitor people's health conditions,

---

diagnose, and prevent diseases. For example, electrocardiograms (ECG) are useful to track cardiac activity and diagnose a heart attack, and electroencephalograms (EEG) are helpful for brain monitoring to predict epileptic seizures. Supervised and unsupervised classification, feature extraction, and dimensionality reduction techniques are among the most used strategies for dealing with these types of data; indeed, biomedical data are often characterized by high dimensionality, and thus, appropriate methodological approaches are required. Definitely, learning from high-dimensional biomedical data is always challenging due to the course of dimensionality (COD) that may drive to data sparsity, troubles in selecting a unique statistical model, multicollinearity, and distance concentration. In addition, other potential concerns may arise: the sampling units might be observed in a restricted set of time points that may be irregularly spaced and diverse for each statistical unit; computational time-consuming and algorithm convergence may be complicated due to probable local minimum; the research for a trade-off between accuracy and complexity; and finally, the interpretation of classification rules may become very hard in high-dimensional spaces. For these reasons, in recent decades, research on classification of massive biomedical data and dimensionality reduction techniques have assumed a fundamental role in many areas, such as medicine, biostatistics, statistics, engineering, and computer science.

The starting idea of this article is that biomedical signals observed over time can be analyzed as functions in the time domain, and consequently, be treated as single objects. The latter procedure seems like the most obvious approach to deal with this type of data and, indeed, it is basically the crucial idea of functional data analysis (FDA).[1,2] The purpose of exploiting FDA to deal with biomedical data is widely shared in the literature.[3-6] Effectively, FDA is one of the most popular approaches to dealing with high-dimensional data and fixing some drawbacks mentioned above, for example, COD. FDA supports reducing the dimensionality of the data and using supplementary critical sources of pattern and variation[1] without the demand of restrictive hypotheses.[7] For these reasons, recently, we are witnessing an uninterrupted growth of methodological research on FDA that attempts to replicate, in a functional key, a large part of classical statistics.[1,2,8-10] In addition, there is a constant development of novel applications and suggestions to answer particular dilemmas in singular contexts employing functional instruments.[11-16]

Because FDA is widely appreciated as a valuable tool for analyzing biomedical data, research on curves' supervised classification is lively.[7,17-19] However, functional data supervised classification using tree-based techniques is still little known and underdeveloped. Few prior investigations dealt with such a problem and proposed very diverse approaches. Yu et al[20] introduced spline trees for functional data, with an application to time-of-day patterns for consumers' international calls. Balakrishnan[21] recommended combining functional classification trees (FCTs) and clustering. Nerini et al[22] concentrated on the problem of regression trees to predict probability density functions via FDA. Fan et al[23] proposed kernel-induced random forests to predict temporal gene expression data. Gregorutti[24] focused on variables' importance measures in tree-based methods using FDA. Moller et al[25] suggested extending random forest to FDA by using different mean values calculated at varying time windows as possible features. El Haouij et al[26] recommended using functional random forest (FRF) via wavelet basis with an application to driver's stress level classification. Finally, Belli and Vantini[27] focused on constrained convex optimization to select multiple weighted integral characteristics from the input functions and define binary splits of trees trained utilizing functional inputs. These above-mentioned approaches are very different from each other and represent the first studies on the possible combined use of tree-based techniques and FDA. The common idea of these studies is certainly to extend classical classification trees to a datum that can be represented through a curve. The reason is that tree-based techniques are precise classification tools and powerful instruments that help the interpretation of classification rules; therefore, in a world of "streaming data" where high-dimensionality has become a very common issue to deal with, pooling FDA and decision trees would lead to accurate and interpretable classifiers, dimensionality reduction, and additional features to extract information from the original data.

With the same starting idea but with a different strategy, this research intends to pool tree-based techniques and FDA for supervised classification of curves representing high-dimensional biomedical data recorded over time. Remarkably, this investigation proposes FCTs and FRF. Using an application on ECG data, this study shows the power of the proposed functional classifiers in terms of accuracy and their usefulness in terms of interpretability of the classification rules. Based on the technique utilized to describe the curves, various solutions are presented to decrease the dimensionality of the data and represent phenomena. In this research, the b-spline representation and the functional principal components' decomposition (FPCD) are proposed as possible basis transformations to obtain features from curves and train FCTs and FRF classifiers. The first original part of this article is the introduction of new tools to support interpreting the functional classification rules in the functional framework, that is, the so-called empirical splitting curve (ESC) and theoretical splitting curve (TSC). The second innovative proposal concerns the presentation of different functional measures to evaluate the variability between groups (within the same terminal node) or between terminal nodes (given the group) and their interpretation (with particular attention to the case of a binary outcome and therefore in terms of true positives, false

positives, true negatives, and false negatives). The result of the joint use of these tools linked to the construction of a functional classification tree provides a powerful tool to interpret the classification rule and understand, in the time domain, how the discriminatory rule operates and how reliable it is.

This study shows compelling results from various viewpoints. First, it's worth noticing that the FRF classifier, in terms of accuracy, breaks the previous world record on a well-known ECG dataset. Next, the concept of employing the FPCD to derive new features for training a functional classifier is a charming proposal because it enables refining the classification rules from noise and redundant information of signals in the time domain. The last aspect drives to a more steady classifier when implemented on a test set because it reduces overfitting by controlling the number of FPCs used to represent the original data. Then, the representation of the FCT through our interpretative instruments is very natural and effective in explaining the purpose of each split in a functional context. Consequently, aiming to reduce the variance of a single FCT and decorrelate FCTs in the view of a collection of FCTs, the purpose of extending random forest to FDA demonstrates to be very appealing and worthy of additional studies. In summary, this line of investigation turns out to be very promising in a world that is increasingly dominated by considerable volumes of data that we continuously gather from various devices in multiple domains of application, for example, the medical one.

This research is structured as follows. Section 2 provides an introduction to FDA, FPCD, and the most used proximity measures among functional data. Section 3 illustrates FCTs and many innovative tools to help their interpretation and assess the quality and composition of each leaf. Section 4 displays the FRF procedure. Section 5 shows a detailed application of the proposed method to ECG data with two classes to predict. Moreover, the results of another application on ECG data, with a different number of classes, is presented. A comparison between the proposed approach and some well-known supervised classification methods for functional data is shown. Finally, Section 6 gives the conclusions and suggests possible prospective extensions of this study.

2 | FUNCTIONAL DATA REPRESENTATION

The fundamental idea of FDA is to manage data functions as single objects. However, in practical applications, functional data are usually observed as series of point data. Thus, the function denoted by $z = f(x)$ reduces to a record of discrete measurements that are indicated by the T pairs $(x_j; z_j)$ where $x \in \Re$ and z_j are the values of the function calculated at the points $x_j, j = 1, 2, \ldots, T$.[1] Generalizing the reference framework, a functional variable X is a random variable assuming values in a functional space Ξ, and a functional data set is a sample $x_1, \ldots, x_N$, also denoted $x_1(t), \ldots, x_N(t)$, drawn from a functional object.[28]

Focusing our attention to the case of a Hilbert space with a metric $d(\cdot, \cdot)$ associated with a norm so that

$$d(x_1(t), x_2(t)) = \|x_1(t) - x_2(t)\|, \tag{1}$$

and where the norm $\| \cdot \|$ is associated with an inner product $\langle \cdot, \cdot \rangle$ so that

$$\|x(t)\| = \langle x(t), x(t) \rangle^{1/2}, \tag{2}$$

we can obtain as a specific case the space $\mathcal{L}_2[a, b]$ of real square-integrable functions defined on $[a, b]$ by

$$\langle x_1(t), x_2(t) \rangle = \int_a^b x_1(t) x_2(t) \mathrm{d}t. \tag{3}$$

If $x(t) \in \mathcal{L}_2$, a basis function system is a set of known functions $\varphi_j(t)$ that are linearly independent of each other and which span $\mathcal{L}_2$.[1] The first step in FDA is to convert the observed values $x_{i1}, x_{i2}, \ldots, x_{iT}$ for each unit $i = 1, 2, \ldots, N$ to a functional form. The most popular method to estimate the functional datum is the basis approximation. Depending on the characteristics of the curves, various basis systems can be adopted. A simple approach is that functions can be obtained employing a finite representation in a fixed basis system[1] as follows:

$$x_i(t) \approx \sum_{s=1}^{S} c_{is} \varphi_s(t), \tag{4}$$

where c_i is the vector of coefficients defining the linear combination of the ith curve, $\varphi_s(t)$ is the sth basis function with $s \in S$ (S being a finite set of fixed basis).

A different general approach consists of employing a data-driven basis rather than a fixed basis system. The most common procedure is the functional principal component decomposition (FPCD). The latter leads to a dimensionality reduction whilst maintaining the maximum volume of information from the initial data.[1,8,29] In this situation, the functional data can be approximated as follows:

$$x_i(t) \approx \sum_{k=1}^{K} v_{ik}\xi_k(t), \tag{5}$$

where K is the total number of selected FPCs, v_{ik} is the score of the generic FPC ξ_k for the generic function $x_i(t)$ ($i = 1, 2, \dots, N$). Therefore, we can achieve an approximation of the sample curves, whose explained variance is given by the sum of the eigenvalues $\sum_{k=1}^{K} \lambda_k$. Particularly, when dealing with high-dimensional data, the latter dimensionality reduction technique is necessary for explaining the main features of the data by a reduced set of uncorrelated FPCs.

This approach is clearly an extension of the classical PCA. Hereafter, we quickly recall the basics of FPCD. If we assume that the observed curves are centered so that the sample mean is equal to 0, the ith FPCs scores are given by

$$v_{ik} = \int_T x_i(t)\xi_k(t) \, dt \quad i = 1, \dots, N, \tag{6}$$

where the weight function ξ_k is obtained by maximizing the variance, solving:

$$Max_f Var \left[\int_T x(t)\xi_k(t) \, dt \right], \tag{7}$$

s.t.

$$|\xi||^2 = \int \xi_k(t)^2 dt = 1, \tag{8}$$

and

$$\int \xi_k(t)\xi_l(t) \, dt = 0 \quad for \quad l \neq k. \tag{9}$$

2.1 | Most used proximity measures amongst functional data

Proximity measures amongst statistical units have a decisive role in FDA. Evidently, according to various chosen distances, contrasting results can be obtained. Hence, the selection of a proximity measure depends on the nature of data and aim of the analysis. In the context of the FDA, several metrics and semi-metrics have been proposed over time; nevertheless, focusing on the case of the $\mathcal{L}_2$-space, the most frequently used distance between curves are the following.[1,2,29-30] The L_2-distance is the most utilized and can be computed as follows:

$$\|x_1(t) - x_2(t)\|_2 = \left\{ \frac{1}{\int_a^b w(t)dt} \int_a^b |x_1(t) - x_2(t)|^2 w(t)dt \right\}^{1/2}, \tag{10}$$

where $w(t)$ is a strictly positive weight function, and the observed points on every curve are equally spaced. Usually, the semi-metric of the r-order derivatives of two curves, for example, $x_1(t)$ and $x_2(t)$, could be considered because it produces interesting information depending on the purpose of the study. It can be calculated as follows:

$$d_2^{(r)}(x_1(t), x_2(t)) = \left(\frac{1}{T} \int_T \left(x_1^{(r)}(t) - x_2^{(r)}(t) \right)^2 dt \right)^{\frac{1}{2}}, \tag{11}$$

where $x_1^{(r)}(t)$ and $x_2^{(r)}(t)$ are the r-derivatives of $x_1(t)$ and $x_2(t)$, respectively. Finally, the semi-metric of the FPCs is especially attractive when researchers require dimensionality reduction and wish to understand similarity among curves according

to distinct parts of the domain. An additional advantage is that such a measure reduces noise while keeping the most important piece of information. The semi-metric of the FPCs is provided by:

$$d_2\left(x_1(t), x_2(t)\right) \approx \left(\sum_{k=1}^{K} (v_{1k} - v_{2k})^2 \|\xi_k\|\right)^{\frac{1}{2}},\tag{12}$$

where $v_{i,k}$ is the score of the linear combination, and ξ_k is the kth functional principal components.

3 | FUNCTIONAL CLASSIFICATION TREES (FCTS)

The functional classification framework intends to predict an outcome Y by means of a variable X taking values in a separable metric space (Ξ, d). Theoretically, Y could be both categorical or numerical, leading to classification or regression problems, respectively. Nevertheless, this research concentrates on the scalar-on-function classification problem and thus, we consider the case of a categorical outcome Y. Hence, the procedure is designed for functional data of the form $\{y_i, x_i(t)\}$, with a predictor curve $x_i(t), t \in T$, and y_i being the response value observed at sample $i = 1, \ldots, N$. Let the feasible values of Y be 0 or 1, classification of a novel observation x from X is carried out by a mapping $f : F \rightarrow \{0, 1\}$, called a "*classifier*," which maps x into its predicted label. The latter problem can easily be extended to the case of Y with multiple modalities.

A decision tree (DT) classifier is one of the most successful supervised learning techniques to predict values of responses by learning decision rules from features.[31-33] The starting idea is that DTs can be extended to the FDA framework by employing the scores of a fixed basis system like those in Equation (4) or (5) as original features to train the functional classifier. In the case of a data-driven basis system, for example, Equation (5), the features' matrix is given by:

$$\mathbf{V} = \begin{pmatrix} v_{11} & \ldots & v_{1K} \\ \vdots & \ddots & \vdots \\ v_{N1} & \ldots & v_{NK} \end{pmatrix},\tag{13}$$

where v_{ik} is the score of the ith curve $(i = 1, \ldots, N)$ relative to the kth functional principal component ξ_k $(k = 1, \ldots, K)$.

In the instance of a fixed basis system, for example, Equation (4) using B-splines, the features' matrix is given by:

$$\mathbf{C} = \begin{pmatrix} c_{11} & \ldots & c_{1S} \\ \vdots & \ddots & \vdots \\ c_{N1} & \ldots & c_{NS} \end{pmatrix},\tag{14}$$

where the generic element c_{is} is the coefficient of the ith curve $(i = 1, \ldots, N)$ relative to the sth $(s = 1, \ldots, S)$ basis function $\varphi_s(t)$ involved in the linear combination. The most challenging element of these strategies is to look for a functional interpretation of the classification rules provided by FCTs. For this reason, the selection of one of the two approaches depends on diverse reasoning, for example, desired interpretability and dimensionality reduction needs.

A FCT consists of recursive binary separations of the feature space into rectangular regions (terminal nodes or leaves) made by collections of curves $x_i(t) \in X$. To build the FCT, an optimal binary separation is provided at every step of the algorithm, based on the optimization of a cost criterion (eg, decrease of the impurity of the node via the Gini or Shannon-Weiner index).[33,34]

The Gini index is a measure of heterogeneity for categorical variables, and thus the lower the value of the index, the more homogeneous the observations in a node (the more the node is "pure"). It can be computed as follows:

$$G = 1 - \sum_{i}^{F} f_{zi}^2,\tag{15}$$

where f_{zi} is the proportion of training observations in the zth node that are from the ith class, and F is the number of modalities of the outcome Y.

The Shannon-Wiener entropy index is also an index of heterogeneity for categorical variables, and thus the interpretation is similar to the Gini's index. It can be calculated as follows:

$$E = \sum_{i}^{F} f_{zi} \cdot \log f_{zi}, \tag{16}$$

where f_{zi} represents the proportion of training observations in the zth node that are from the ith class.[33,35]

The algorithm starts with the entire functional data set using the scores of the FPCD obtained using Equation (13) or the coefficients obtained using Equation (4), and proceeds until the leaves are obtained. Having obtained the best split in one node, the data are partitioned into two nodes; the rule is replicated to produce the most proper binary division on all resulting nodes. Typically, a vast FCT is produced at the beginning, which is then pruned according to an optimization criterion, for example, to look for an acceptable trade-off between complexity and accuracy. In summary, the FTC is an extension of the classical classification tree algorithm (see eg, References 35-37) to the functional context.

3.1 | The theoretical splitting curve (TSC) and empirical splitting curve (ESC)

Given that the scores of the linear combination are used as new features to predict the response Y, the interpretation of FCT is totally different if compared to the classical CTs. Indeed, the values of the splits should be interpreted according to the part of the domain that the single FPC $\xi_k(t)$ or B-Spline $\varphi_s(t)$, mostly describe and the scores' thresholds v_{0k} or c_{0s}, respectively. The subscript "0" (rather than i) in v_{0k} and c_{0s} symbolizes that the threshold recognized for the score relating to a specific FPC $\xi_k(t)$ or B-spline $\varphi_s(t)$ becomes a fixed value to divide the curves into two subsamples (son nodes). Effectively, v_{0k} and c_{0s} are fixed value that does not depend on the ith unit but are the same for all the units. For example, contemplating the first split rule on a hypothetical FCT based on FPCs, that is, the division rule of the root node, v_{0k_0} is the threshold value related to a generic FPC K_0. Thus, all the curves meeting the condition $v_{ik_0} < v_{0k_0}$ form a subgroup whereas all the residual functions, that is, those meeting the condition $v_{ik_0} \geq v_{0k_0}$, enter the other subset.

To make the split rule better readable in the FDA setting, we offer two notions: the TSC and ESC. The TSCs are given by each theoretical separation rules generated by every single split of the FCTs. In the case of a data-driven basis system, for example, Equation (5), the TSP can is given by:

$$TSC_z^{FPC}(t) = \sum_{k \in R} v_{0k} \xi_k(t), \tag{17}$$

where R is the set of FPCs $\xi_k(t)$ involved in the classification rule path until the splitting of the zth intermediate node $(z = 1, \ldots, Z)$. Therefore, the generic intermediate node that produces a cut is indicated with z, and the whole number of these middle nodes is identified with Z. Consequently, each $TSC_z^{FPC}(t)$ can be associated to each intermediate node and, of course, also to the root node $(TSC_1^{FPC}(t))$. The limit of $TSC_z^{FPC}(t)$ is that it does not help to understand the separating laws in terms of the time domain. Indeed, TSCs are generally very smooth with few functional variability.

Consequently, we introduce the ESC. The latter is represented by the curve, living in the training dataset, which is the closest (based on the semi-metric FPCs) to the separation rule given by $TSC_z^{FPC}(t)$ and can be defined as follows:

$$ESC_z^{FPC}(t) = x^{(z)FPC}(t) = \underset{i}{\operatorname{argmin}} \ d_2 \left(x_i(t), TSC_z^{FPC}(t) \right). \tag{18}$$

Henceforth, $ESC_z^{FPC}(t)$ is a curve outlining the functional empirical separation rule provided by the binary separation of a node split and is very helpful to understand the separation rule over the time domain. It follows that we can identify and plot a $ESC_z^{FPC}(t)$ for every split z being in the FCT based on FPCs. Therefore, we can always identify a $ESC_z^{FPC}(t)$ associated to each $TSC_z^{FPC}(t)$. This strategy serves in explaining the various levels of separation in the FCT based on FPCs.

In the instance of a fixed basis system, for example, Equation (4) using B-splines, the TSP is given by:

$$TSC_z^{BSP}(t) = \sum_{s \in Q} c_{0s} \varphi_s(t), \tag{19}$$

where Q is the number of b-splines $\varphi_s(t)$ used in the classification rule path until the split of the zth node $(z = 1, \ldots, Z)$.

Instead, the TSC can be defined as follows:

$$ESC_z^{BSP}(t) = x^{(z)BSP}(t) = \underset{i}{\text{argmin}} \ \left\| x_i(t) - TSC_z^{BSP}(t) \right\|_2 \tag{20}$$

Hence, the distance used to look for $ESC_z^{BSP}(t)$ is provided by Equation (10) whereas the semi-metric adopted to identify $ESC_z^{FPC}(t)$ is given by Equation (12).

3.2 | Leaves quality assessment using functional tools

After pruning the FCT to look for an optimal trade-off between complexity and accuracy, the final nodes are composed of curves belonging to different classes. Therefore, using the Gini or Shannon-Weiner indexes[33,35] (see Equations (15) and (16)), it is straightforward to observe that a pruned FCT is often made of many terminal nodes that are not pure. In a classical non-functional classification tree, it is common to refer only to the classical measures of heterogeneity above-mentioned to check for the impurity of a terminal node. However, in the context of FDA and FCTs, it is possible to exploit supplementary sources of information furnished by the functional nature of the data in each node. For this reason, in this subsection, many additional criteria to evaluate the quality of the leaves in terms of variability are presented.

3.2.1 | The functional variability of a leaf

The functional variability of the leaves is introduced as extra information about the characteristics of the decision rules. In other words, the variability of the terminal nodes can be computed according to two different types of measures, and thus we can get a double source of information. The first source of variability is given by the heterogeneity computed using the Gini or the Shannon-Weiner index; therefore, we have two measures of the impurity of a node determined according to the distribution of the curves based on the different modalities of the outcome Y. From a conceptual and also practical perspective, the only difference with respect to the classical case is that individuals are represented by curves belonging to different classes rather than scalar values. The second source of variability is given by the functional dispersion of the curves in each leaf around the functional mean of the same leaf. In other words, despite the curves in each leaf can belong to different classes, we can also measure the functional deviance that examines the leaf variability over the whole reference domain. The functional deviance of a leaf is given by:

$$DEV^{(l)}(t) = \sum_{i=1}^{n^{(l)}} (x_i(t) - \bar{x}^{(l)}(t))^2 \quad i \in l, \tag{21}$$

where $DEV^{(l)}(t)$ is the functional deviance of the generic lth leaf, $n^{(l)}$ is the number of curves in the lth leaf ($l = 1, \dots, L$), $\bar{x}^{(l)}(t)$ is the functional mean of the lth leaf. Large values of $DEV^{(l)}(t)$ furnishes information on time intervals in which the curves belonging to the same terminal node are very far from the functional mean of the same leaf, and therefore the curves are diverse from each other. In other words, the latter would be evidence of a terminal node composed of different functions, regardless of the labels of the response variable Y that do not contribute to the calculation of this type of variability. Conversely, a moderate value of $DEV^{(l)}(t)$ symbolizes a leaf with very similar curves.

To display the functional variability of the terminal nodes, in terms of the total functional deviance of the root node, $DEV^{(l)}(t)$ can be depurated by the total variability as follows:

$$_{rel}DEV^{(l)}(t) = \frac{\sum_{i=1}^{n^{(l)}} (x_i(t) - \bar{x}^{(l)}(t))^2}{\sum_{i=1}^{N} (x_i(t) - \bar{x}(t))^2} \quad i \in l, \tag{22}$$

where $\sum_{i=1}^{N} (x_i(t) - \bar{x}(t))^2$ is the functional deviance of the root node and $\bar{x}(t)$ is its functional mean. The latter tool can be useful to compare the quality of the leaves and better understand their composition according to different parts of the time domain.

3.2.2 | The functional between groups variability within a single leaf

Based on the classification rule of a FCT, the perfect result would consist of getting all the curves in a terminal node with the same predicted class. Nevertheless, in a pruned FCT, the nodes are usually impure, and thus there are some leaves composed of incorrectly classified curves. To capture the variability among different groups in a leaf, the functional between groups sum of squares (FBGSS) of the leaf can be used, the latter can be defined as follows:

$$FBGSS^{(l)}(t) = \sum_{g=1}^{G} n_g^{(l)} (\overline{x}_g^{(l)}(t) - \overline{x}^{(l)}(t))^2, \tag{23}$$

where G is the total number of groups, $n_g^{(l)}$ is the number of curves of the gth group in the lth leaf, $\overline{x}^{(l)}(t)$ is the total functional mean in the lth terminal node, and

$$\overline{x}_g^{(l)}(t) = \sum_{i=1}^{n_g^{(l)}} \frac{x_i(t)}{n_g^{(l)}} \quad i \in l, \tag{24}$$

is the functional mean of the gth group in the leaf l.

To obtain a more refined functional measure, depurated by the variability of the terminal node l, $FBGSS^{(l)}(t)$ can be divided by the total variance of the leaf as follows:

$$_{rel}FBGSS^{(l)}(t) = \frac{FBGSS^{(l)}(t)}{DEV^{(l)}(t)}. \tag{25}$$

A particular interpretation of the functional between groups variability within a single leaf when the response variable is binary

Let the possible values of Y be 0 ("*Disease*") or 1 ("*Healthy*") and a classification problem where the aim is to predict the presence/absence of a specific disease in people at risk. It follows that the predicted classes for new curves in the test set can be "*Positive*" or "*Negative*" depending on the classification rule of the FCT, and thus, on the terminal node where the patient will be attributed. As in the classic non-functional case, if the class predicted by the leaf is "*Disease*" and the patient is ill, then there is a true positive (TP). If the classification rule makes a mistake because the patient is healthy, there is a false positive (FP). If the class predicted in the leaf is "*Healthy*," and the patient is healthy, then there is evidence of a true negative (TN). Conversely, if the patient is sick, a false negative (FN) is present. Starting from the above described classical approach to define FPs, FNs, TPs, and TNs, ad-hoc functional measures to get extra information on leaves' quality can follow a specific notation.

When the predicted class of a terminal node l is "*Disease*," the functional mean of the TP curves in the leaf l can be indicated as follows:

$$\overline{x}_{TP}^{(l)}(t) = \sum_{i=1}^{n_{TP}^{(l)}} \frac{x_i(t)}{n_{TP}^{(l)}} \quad i \in l, \tag{26}$$

where $n_{TP}^{(l)}$ is the total number of TP curves in the leaf l with $n_{TP}^{(l)} \leq n^{(l)}$. Instead, the functional mean of the FP functions in the leaf l is given by:

$$\overline{x}_{FP}^{(l)}(t) = \sum_{i=1}^{n_{FP}^{(l)}} \frac{x_i(t)}{n_{FP}^{(l)}} \quad i \in l, \tag{27}$$

where $n_{FP}^{(l)}$ is the total number of FP functions in the leaf l with $n_{FP}^{(l)} \leq n^{(l)}$, and $n^{(l)} = n_{TP}^{(l)} + n_{FP}^{(l)}$.

$FBGSS^{(l)}(t)$ and $_{rel}FBGSS^{(l)}(t)$, in this context, are indicated with $FBGSS_{TP-FP}^{(l)}(t)$ and $_{rel}FBGSS_{TP-FP}^{(l)}(t)$, respectively, to stress that the focus is on a leaf predicting a "*Disease*" status. The assessment and plot of $_{rel}FBGSS_{TP-FP}^{(l)}(t)$ across the entire

domain provides interesting information about the composition of the leaf and why TP and FP curves are in the same terminal node l. Effectively, given the variance of the node, the lower ${}_{rel}FBGSS^{(l)}_{TP-FP}(t)$, the more similar $\overline{x}^{(l)}_{TP}(t)$ and $\overline{x}^{(l)}_{FP}(t)$; thus, despite the actual health condition of some patients, they are badly classified because their signals are similar to diseased people in certain parts of the time domain. Evidently, the parts of the domain in which these similarities exist resulted in the same predicted class. Instead, the higher ${}_{rel}FBGSS^{(l)}_{TP-FP}(t)$, the greater the difference between $\overline{x}^{(l)}_{TP}(t)$ and $\overline{x}^{(l)}_{FP}(t)$ in some parts of the time domain. In the latter instance, there is an important indication of where the actual functional difference between TPs and FPs is. In other words, it is possible to identify those characteristics of the FPs' signals that has been ignored by the classification rule, resulting in a classification error. As an extreme case, ${}_{rel}FBGSS^{(l)}_{TP-FP}(t) = 0$ may happen even if it is quite unlikely but still possible. In fact, in the presence of an impure node and equality between $\overline{x}^{(l)}_{TP}(t)$ and $\overline{x}^{(l)}_{FP}(t)$ over the whole domain, we have a situation in which it is almost impossible to train a functional classifier that would be precise for those statistical units. In fact, it would mean that identical curves have different original labels. Training a classifier with good chances to correctly predict the class of those similar functions would be quite impossible because the latter share the same features.

Following the above reasoning, the same idea can be extended to the case in which the predicted class of a terminal node l of a FCT is "*Healthy*." The functional mean of the TN curves in the leaf l is given by:

$$\overline{x}^{(l)}_{TN}(t) = \sum_{i=1}^{n^{(l)}_{TN}} \frac{x_i(t)}{n^{(l)}_{TN}} \quad i \in l, \tag{28}$$

where $n^{(l)}_{TN}$ is the total number of TN curves in the leaf l with $n^{(l)}_{TN} \leq n^{(l)}$. Instead, the functional mean of the FN functions in the leaf l is given by:

$$\overline{x}^{(l)}_{FN}(t) = \sum_{i=1}^{n^{(l)}_{FN}} \frac{x_i(t)}{n^{(l)}_{FN}} \quad i \in l, \tag{29}$$

where $n^{(l)}_{FN}$ is the total number of FN functions in the leaf l with $n^{(l)}_{FN} \leq n^{(l)}$, and $n^{(l)} = n^{(l)}_{TN} + n^{(l)}_{FN}$.

The same reasoning described for TPs and FPs can be extended to the case of TNs and FNs. $FBGSS^{(l)}(t)$ and ${}_{rel}FBGSS^{(l)}(t)$, in this circumstance, are indicated with $FBGSS^{(l)}_{TN-FN}(t)$ and ${}_{rel}FBGSS^{(l)}_{TN-FN}(t)$, respectively, to remark we are focusing on a leaf predicting a "*Healthy*" status. In this case, high values of ${}_{rel}FBGSS^{(l)}_{TN-FN}(t)$ indicate that despite TNs and FNs have different behaviors in some parts of the time domain, they share the same predicted class. Analyzing the intervals where this difference is most marked is a clue to detect those functional features that have been neglected by the classification rule and led to the increase of the final misclassification error of the functional classifier.

In this section, there is no distinction according to the basis system. Indeed, the proposed strategy can be adopted whatever the base system used because the former is based on leaves' composition rather than how curves are represented.

3.2.3 | The functional between leaves variability of all the leaves for a single group

In the previous section, in addition to the classical impurity measures, some original procedures to evaluate the quality of a specific leaf based on the between groups variability in that specific leaf are introduced. Thus, in the latter approaches, only individual terminal nodes and their composition are taken into account. Nevertheless, common questions that would arise in this peculiar context are the following: what is the behavior of the functional mean of a subgroup of FNs in a leaf dominated by TNs (thus leading to the predicted class "*Healthy*") if compared to the trend of the functional mean of a subgroup of TPs in an impure leaf dominated by TPs (which therefore leads to the predicted class "*Disease*")? And what is the behavior of the functional mean of a subgroup of FPs in an impure leaf dominated by TPs (which therefore leads to the predicted class "*Disease*") if compared to the functional mean of a subgroup of TNs in a leaf dominated by TNs (which then leads to the predicted "*Healthy*" class)? It is possible to answer these questions by introducing the concepts of functional between leaves variability of all the leaves and functional between leaves variability of two leaves. In the following, we propose two different measures: the first one is the functional between leaves variability of one group among all the leaves; and the second one is the functional between leaves variability of a couple of leaves. The latter is a particular case of the first one, that is, when considering only two leaves and one of them is the so-called "*reference*" leaf. The next section is dedicated to the latter functional measure.

The functional between leaves sum of squares (FBLSS) among all the leaves can be defined as follows:

$$FBLSS_g(t) = \sum_{l=1}^{L} n_g^{(l)} (\overline{x}_g^{(l)}(t) - \overline{x}_g(t))^2, \tag{30}$$

where $\overline{x}_g(t)$ is the overall functional mean for the group g, that is, considering all the leaves. In other words, $FBLSS_g(t)$ takes into account only one group g to detect its variability among different leaves. Consequently, it is a valuable measure to understand how leaves lead to different prototypes of the same original group g.

A particular interpretation of the functional between leaves variability of all the leaves for a single group when the response variable is binary

Limiting the attention to the case of a binary outcome Y that can only take values ("*Disease*") or ("*Healthy*"), $FBLSS_{TP-FN}(t)$ indicates the FBLSS taking into account all the diseased patients present in all the leaves, that is, TPs and FNs. On the other hand, $FBLSS_{TN-FP}(t)$ is the FBLSS taking into account all the healthy patients present in all the leaves, that is, TNs and FPs.

3.2.4 | The functional between leaves variability of two leaves for a single group

A special case of $FBLSS_g(t)$ can be obtained by considering only a couple of leaves instead of all the terminal nodes. Particularly, we consider arranging in descending order the terminal nodes based on the relative frequency of correct predictions available in each of them. Let the best leaf in the ranking be indicated as l_*. As stressed before, in a pruned FCT, the nodes are often impure and therefore rarely lead to a prediction that is "error-free." Suppose the lth terminal node leads to the $\hat{g}$ predicted class. Hence, the lth terminal node has a majority of statistical units with original labels equal to g and a minority with an original class different from g (note that, if not, the overall prediction of the terminal node would be different from g). Using the previous example of diseased and healthy people, one would question which is the divergence in trends between diseased patients who are badly classified as "*Healthy*" and a reference curve representing diseased people who are rightly classified as "*Diseased*." In other words, which is the difference between FNs in a specific leaf and a reference curve of TPs?

The generalized version of the FBLSS of two leaves l and l_* can be given by:

$$FBLSS_g^{(l)-(l_*)}(t) = n_g^{(l)} (\overline{x}_g^{(l)}(t) - \overline{x}_g^*(t))^2 + n_g^{(l_*)} (\overline{x}_g^{(l_*)}(t) - \overline{x}_g^*(t))^2, \tag{31}$$

where $\overline{x}_g^*(t)$ is the functional mean computed considering only the leaves l and l_*, $n_g^{(l_*)}$ is the total number of curves belonging to the group g in the reference leaf l_*. In other words, $FBLSS_g^{(l)-(l_*)}(t)$ considers only one group g in a specific leaf l with respect to a reference leaf l_*. Hence, it is a useful measure to compare two leaves leading to two different prototypes of the same original group g.

To obtain a more accurate functional comparison between $\overline{x}_g^{(l)}(t)$ and $\overline{x}_g^{l_*}(t)$, we consider a functional variability measure that is depurated by the total functional between leaves variability of the group g (see Equation (30)) as follows:

$$_{rel}FBLSS_g^{(l)-(l_*)}(t) = \frac{FBLSS_g^{(l)-(l_*)}(t)}{FBLSS_g(t)}. \tag{32}$$

In summary, $_{rel}FBLSS_g^{(l)-(l_*)}(t)$ is the proposed "*gold-standard*" functional measure to provide a functional comparison between two groups of functions belonging to the same class g but in two different terminal nodes relative to the total variability of the curves belonging to the same class g between all the leaves.

A particular interpretation of the functional between leaves variability of two leaves for a single group when the response variable is binary

Focusing the attention to the case of a binary outcome with possible values ("*Disease*") and ("*Healthy*"), we can consider two different cases that help the understanding of the proposed approach. First, when focusing on diseased patients, we are interested in comparing TPs who are correctly classified in a leaf and FNs that are incorrectly classified in another leaf.

Both subgroups are composed by diseased people, and thus the most interesting aspect is to capture the similarities and dissimilarities of these subgroups. Hence, the functional between leaves mean squares to compare TPs in the reference leaf and FNs in a generic leaf l, relative to the total between variability among all diseased patient in different leaves, is indicated as:

$$_{rel}FBLSS_{FN-TP}^{(l)-(l_{TP})}(t) = \frac{FBLSS_{FN-TP}^{(l)-(l_{TP})}(t)}{FBLSS_{FN-TP}(t)}, \tag{33}$$

where the subscript "$FN - TP$" instead of "g" highlights that the subgroups of diseased patients are considered. Following the previous notation, in Equation (33), l_{TP} is the reference leaf for diseased people, that is, the leaf with the most accurate prediction of disease. However, the reference leaf can also be changed according to different reasoning and ranking, for example, looking for a trade-off between the total number of curves in the leaf and its accuracy in predicting the disease status. Evidently, the plot of $_{rel}FBLSS_{FN-TP}^{(l)-(l_{TP})}(t)$ over the whole time domain provides fascinating information about the FN curves in the generic leaf l. Indeed, now, it is possible to answer to the first out of the two above-mentioned questions. High values of the functional between variability over some intervals of the time domain can help to explain why FNs are in leaf l leading to a classification error, instead of being in the terminal node l_{TP} with the other TPs. In other words, we are able to obtain a functional reason for FNs to be "*false*" instead of being "*true*" positives.

On the contrary, if the interest is on healthy patients, we aim to compare TNs who are correctly classified and FP that are incorrectly classified. Both subgroups are composed by healthy people, and thus we aim to detect similarities and dissimilarities of these subgroups. Hence, the FBLSS to compare TNs in the reference leaf and FPs in a generic leaf l, relative to the total between variability among all healthy patient in different leaves, is given by:

$$_{rel}FBLSS_{FP-TN}^{(l)-(l_{TN})}(t) = \frac{FBLSS_{FP-TN}^{(l)-(l_{TN})}(t)}{FBLSS_{FP-TN}(t)}, \tag{34}$$

where the subscript "*FP-TN*" instead of "g" highlights that all the subgroups of healthy patients are considered, and l_{TN} is the reference leaf for healthy people, that is, the leaf with the most accurate prediction of health condition. Evidently, $_{rel}FBLSS_{FP-TN}^{(l)-(l_{TN})}(t)$ allows us to answer the second question mentioned above, that is, explaining the difference, over the whole time domain, between FPs in a leaf and TNs in a reference terminal node describing the most reliable classification rule of the FCT to predict the healthy status. High values of $_{rel}FBLSS_{FP-TN}^{(l)-(l_{TN})}(t)$ allows us to explain why healthy patients ended up in the wrong path of the FCT. In other word, we can capture the functional differences between the best prototype of TNs and the specific average curve describing FPs in another leaf.

The concept of functional between leaves variability of two different leaves can be extended to any couple of nodes l and $j \neq l_*$; however, the most interesting mean comparison, in our opinion, in that one with respect to the most accurate leaf in predicting the specific class.

3.3 | Assessing the performance of FCTs by projecting test set curves onto the FPCs space generated by the training set

To evaluate the performance of the FCT, in terms of accuracy, many possibilities exist, as in the case of the classical supervised classification. Focusing on the training set, we can consider bootstrap or cross-validation to test the misclassification error as in the classical framework. Instead. using a test set is the most reasonable approach, but it deserves attention.

When a fixed basis system is adopted, for example, b-splines, to represent the curves of the training set, the order and number of the b-splines are selected. If the same order and number of basis are used also to describe the curves in the test set, the use of the splitting rules given by the FCT is straightforward. Indeed, the coefficients of the linear combination representing the functions of the test set are associated with the same basis functions used for the coefficients of the linear combination representing the functions of the training set. Instead, when a data-driven basis system is used, for example, FPCs, computing once again FPCs but on the test set will produce a different basis system because it strictly depends on the data. Hence, the test set curves must be projected into the space spanned by the FPCs estimated on the training data. In other words, the values of the scores of the test set curves must be calculated. Using the latter strategy, the jth principal component score can be computed as follows:

$$v_{ik} = \langle x_c, \xi_k \rangle = \int_T x_i(t)\xi_k(t)\, dt \quad i = 1, \dots, M, \tag{35}$$

where the weight functions $\xi's$ are obtained performing the FPCs on the training set, x_c are the centered functions of the test set, and M is the total number of curves in the test set. After expressing the curves in the test set as a function of the FPCs system estimated on the training set, then we can apply the classification rule provided by the FCT and estimate the misclassification error and accuracy of the functional classifier.

4 | FROM FCTS TO THE ENSEMBLE: THE FUNCTIONAL RANDOM FOREST (FRF) APPROACH

Today, RF[33,34,38] is recognized as one of the most efficient supervised classification algorithm. Specifically, RF is a particular case of bagging for classification trees.[37] This study suggests extending the classical RF algorithm[39,40] to functional data with the equivalent motivations used to support the introduction of this approach in the non-functional framework, that is, improving the performance and decreasing the variance of a single FCT. To introduce the so-called FRF, the functional bagging procedure (FBG), that is, functional bootstrap aggregation, should be introduced in advance. The main reason to introduce FBG is that, in a single FCT, modest variations in the data may lead to very diverse FCTs and thus different classification rules. Accordingly, the starting idea of FBG is to train an ensemble of FCTs for building a terminal classification rule and reducing the variance of a single FCT.

Let FBG consists of H FTCs τ_h, $h = 1, \dots, H$, where H is chosen to be a large number. The hth tree τ_h is grown on a random subset of the training set, obtained from the original data $D = \{(y_i, x_i(t)), i = 1, \dots, N\}$ by drawing, with replacement, a bootstrap sample $D_h^* = \{(y_s^{(h)}, x_s^{(h)}(t)), s = 1, \dots, N\}$ of the same size N as the original data set. Thus, the set of curves $i = 1, \dots, N$ present in the hth bootstrap sample D_h^* is called, from now on, "*In-Bag Functional Data*" sample (IBFD) and is used to build the single hth FCT. Instead, the "*Out-of-Bag Functional Data*" sample (from now on, OOBFD) is composed of the remaining curves relative to those functions that are not present in D_h^*. Then, we train H FCTs using H bootstrapped functional training sets to get $\hat{f}_h^*$. Afterwards, we average all the predictions to obtain the final prediction for the ith curve as follows:

$$\hat{f}_H(x_i(t)) = \frac{1}{H}\sum_{h=1}^{H}\hat{f}_h^*(x_i^{(h)}(t)). \tag{36}$$

In other words, for a given $x_i(t) \in D$, we register the class predicted by each of the H FCT, and use the so-called "majority vote" see for example, Reference 34. The misclassification error rate can be computed as the average of the error rate of each $\hat{f}_h^*$. Therefore, the overall forecast for each new curve ith is the most usually occurring class amongst the H predictions according to the H different FCT.

Each FCT is grown deep and is not pruned, and thus each FCT has low bias, but a large variance. Hence, the advantage of FBG is that averaging these H FCTs decreases the variance. This procedure gains in accuracy with respect to a single FCT because it combines many FCTs. Increasing H will not lead to overfitting. In practice, we would use a value of H that is large enough for the test error to have settled down. Nevertheless, the limit of FBG is that the decrease in variance is limited because the FCTs are not independent. Effectively, it is trivial to observe that, in real applications, the FCTs are very correlated because almost all are dominated by the same features, that is, FPCs or b-splines in our context. In practice, in most FCTs, the first separations at the top of FDTs are always dictated by the same features that can better discriminate the classes of the outcome Y.

FRF improves FBG by way of a minor tweak that decorrelates the FCTs and reduces the variance when we average many FCTs. Each time a split in an individual FCT is considered, a random selection of m features is chosen as split candidates from the full set formed by K FPCs (see Equation (13)) (or S b-splines if we use Equation (14)). It follows that, when $m < K$ (or $m < S$ in a fixed basis system), we have FRF whereas when $m = K$ (or $m = S$ in a fixed basis system), we have that FBG = FRF. Following this strategy, the FCTs into the forest will be less correlated because the essential FPCs (or b-splines) will not always be those features on the top of the FCT, determining the first significant cut rules. Focusing on FPCs, for example, a general rule of thumb can be to select, as the size of the subset of FPCs, a value of $m \approx \sqrt{K}$.

Consequently, at each split in the FPC, the algorithm is not even allowed to consider most of the available features. Surely, on average, $\frac{m-K}{m}$ of the splits will not even contemplate some FPCs. In this way, FRF decorrelates the FCT, making the average of the FCT less variable and hence more reliable. Thus, the discrepancy between FBG and FRF depends on the selection of m.

Regarding the estimation of the misclassification error rate, in addition to using the training set, test set, or the classical cross-validation approaches, it is also possible to exploit the OOBFD as described in Equation (36). As discussed above, the latter method has the advantage of allowing all curves in the dataset to be used to train the functional classifier. In this way, it is possible to limit the loss of information due to a possible division of the starting dataset into a training and test set. However, the accuracy of the classifier would always be slightly overestimated if compared to the use of a "pure" test set.

5 | APPLICATIONS TO ECG DATASETS

Cardiovascular disease (CVD) (often used interchangeably with the term "*Hearth disease*") is one of the most significant determinants of morbidity and death among worldwide people. For this reason, the prediction of cardiovascular disorders is one of the most critical problems in medicine and biostatistics. Undoubtedly, identifying in advance patients affected by heart disease can prevent serious consequences, for example, stroke and heart attack. One of the most used techniques to monitor heart health is the ECG which provides a representation of the electrical activity of the heart muscle over time.

This section presents the results of two applications of FCT and FRF to ECG datasets and some comparisons with different functional classifiers. In the first application, the whole method in detail is illustrated while, in the second application, only results and comparisons are presented. Specifically, our approach is applied to two well-known datasets available at https://www.timeseriesclassification.com/: "*ECG200*" and "*ECG5000*."

5.1 | An application to the ECG200 dataset with a detailed description of the strategy

The ECG200 dataset was proposed by R. Olszewski at Carnegie Mellon University in 2001, as part of his work "*Generalized feature extraction for structural pattern recognition in time-series data.*"[41] The ECG200 dataset is continuously used to test new classifiers, and the current world record, in terms of classification accuracy, has been reached using the BOSS algorithm and is equal to 89.05%. Each series traces the electrical activity registered during one heartbeat. The two classes are normal heartbeat (NH) and myocardial infarction (MI). The data are composed of one-hundred signals in the training set and one-hundred in the test set. The data is freely available on the website https://www.timeseriesclassification.com/description.php?Dataset=ECG200.[42] Our goal is to predict whether a new patient, whose ECG we observe, is healthy or diseased.

Figures 1 and 2 show the smoothed versions of the original signals in the training and test sets computed using Equation (4) via the *fda* and *fda.usc* R packages.[29,43] The blue signals represent healthy patients (NH), whereas the green curves represent those who had a diagnosis of heart disease (MI).

This first part of the application focuses on the FPCs approach to build FCTs and FRF. Figure 3 considers the first fifteen FPCs. The variability explained by each FPC is shown in the legend. The first three FPCs explain about 75% of the total variability. We remark that, in this framework, the traditional ways of choosing the number of FPCs are not helpful. Indeed, FPCs explaining little variability are often decisive in discriminating the outcome classes and thus are essential in the construction of the FCT. Effectively, the first FPC, which by construction is the one that catches most of the variability (see the black curve in the central part of the time domain), is rarely crucial in FCT. As Figure 3 remarks, each FPC explains different parts of the time domain differently and can be helpful to discriminate the classes of the outcome in the context of FCTs.

A non-pruned tree (available as supplementary material) is rarely useful for practical purposes; in fact, when the FCT incurs in overfitting, it will have poor performance in classifying different datasets. For this reason, the pruning phase,[39] based on the search for an optimal trade-off between complexity and accuracy, is fundamental. The cost-complexity pruning via cross-validation is performed using the R package *rpart*.[35] Figure 4 shows the pruned FCT built using the FPCs' scores as features via the *rpart.plot* R package.[44] The cut on a specific value of a FPCs' score determines the split of a node. Among all the possible FPCs and splitting score values, the one that maximizes the decrease of impurity of the node is
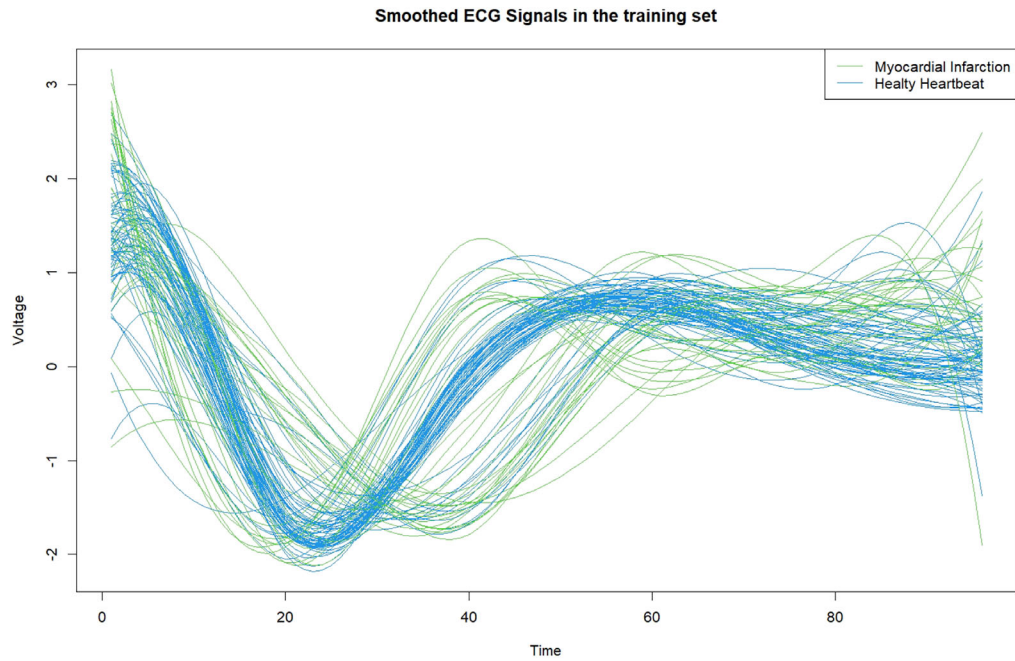
FIGURE 1 Smoothed ECG functions in the training dataset
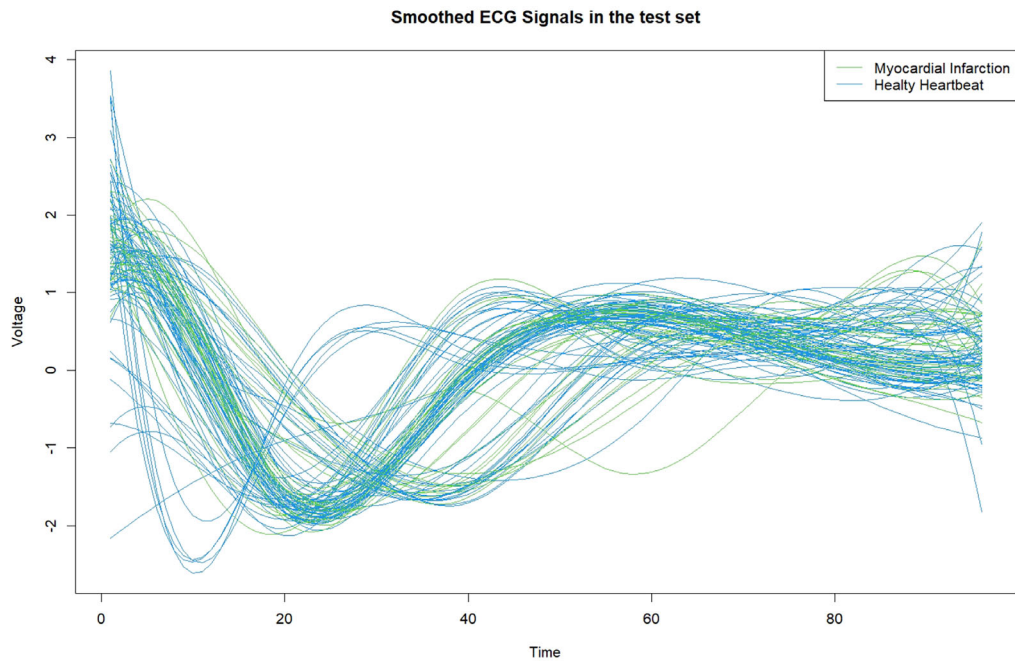
FIGURE 2 Smoothed ECG curves in the test dataset

chosen. The second FPC is the most important feature in our functional classifier. As expected, the first FPC is not essential for discriminating the classes, as it captures a variability common to many curves of different classes. Curiously, the ninth FPC appears to be important in improving the separation rule. This last result proves that, in this context, the most discriminant FPCs are not necessarily the ones explaining the highest inertia according to the classical criterion of PCs selection.

Figure 5 explains in detail the first separation rule (root node) dictated by the second FPC. Figure 5A shows the two original groups and their functional means, while Figure 5B depicts the predicted groups applying only the first
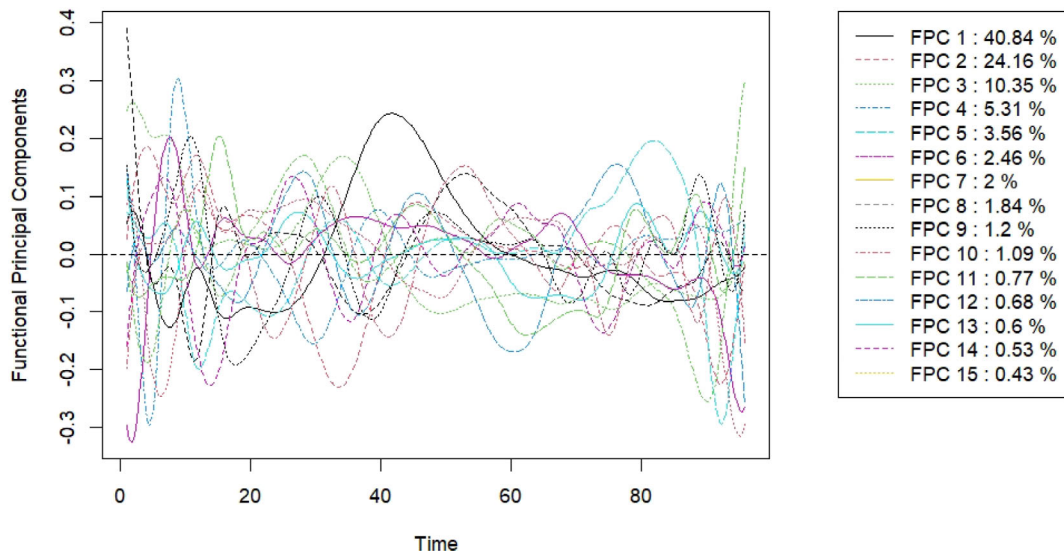
FIGURE 3 First fifteen FPCs of the original ECG curves of the training set
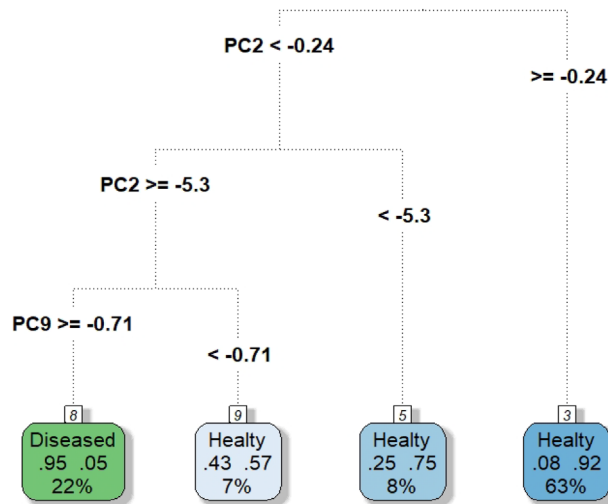
FIGURE 4 Pruned FCT[44] based on the FPCs of the ECG200 dataset. The green leaves indicate a classification as diseased, while the blue leaves lead to the classification of the subjects as healthy. The relative frequencies of subjects of the two groups that make up the terminal leaf are indicated in the boxes. The percentage indicated at the bottom of the box indicates the fraction of people classified following the decision rule that leads to that terminal leaf

separation rule and their functional means. Evidently, the second FPC has great power in distinguishing between NH and MI because the two figures appear to be similar. However, beyond the first evidence, there are many curves that are still poorly classified.

Figure 6 illustrates the TSCs obtained as a functional linear combination using the three splitting score values. The three TSCs computed according to Equation (19) are inserted in the representation of the FCT to support the understanding of the classification strategy. Although the three TSCs seem to have high variability over time, this effect is apparent and only due to the increased scale of the ordinate axis. In reality, these curves are very flat. Therefore, when plotted together with the original curves, TSCs do not help to explain the separation rule in terms of curves existing in the data. For this reason, we introduce the ESCs showed in Figure 7. Figure 7A–C exhibit the three ESCs based on the three different classification rules recommended by the FCT. The red curve indicates the ESC derived from the root node split based on the threshold $v_{02} = -0.24$. The orange function is the ESC based on the split of the subset of curves satisfying

(A) ECG original groups

(B) Groups according to the split rule of the 2nd FPC

FIGURE 5   The original ECG groups and the predicted ECG groups according to the first split rule of the root node
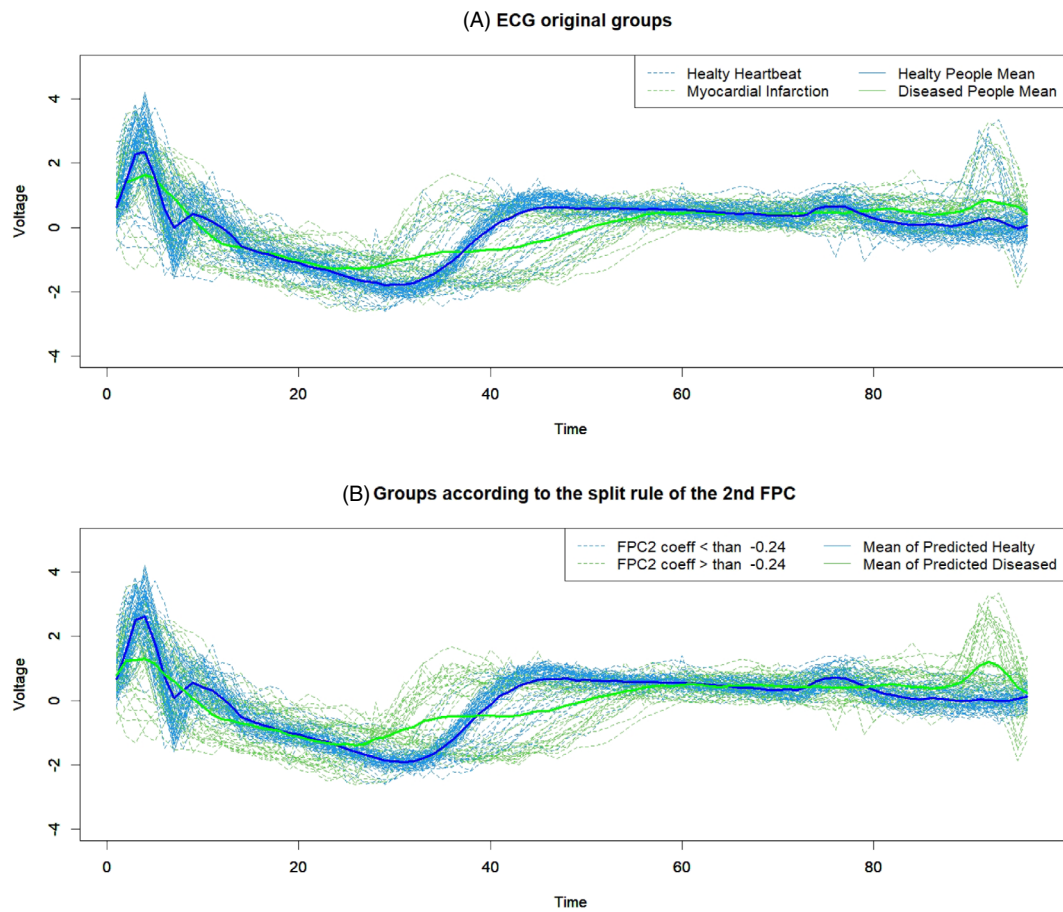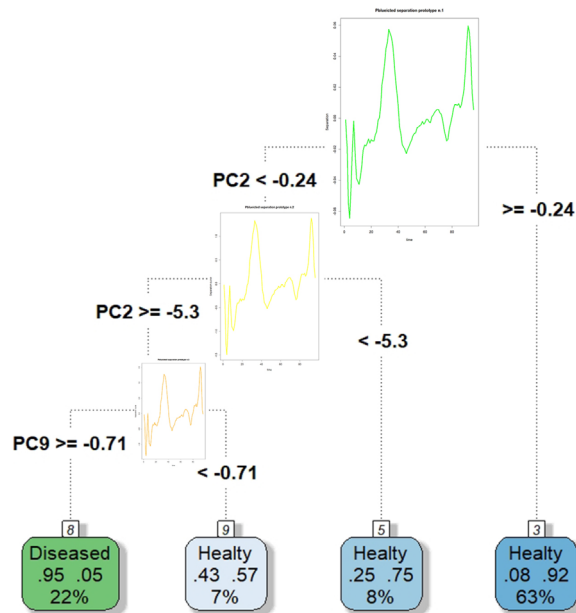

FIGURE 6   Theoretical splitting curves (TSCs) of the FCT obtained using the three splits based on FPCs
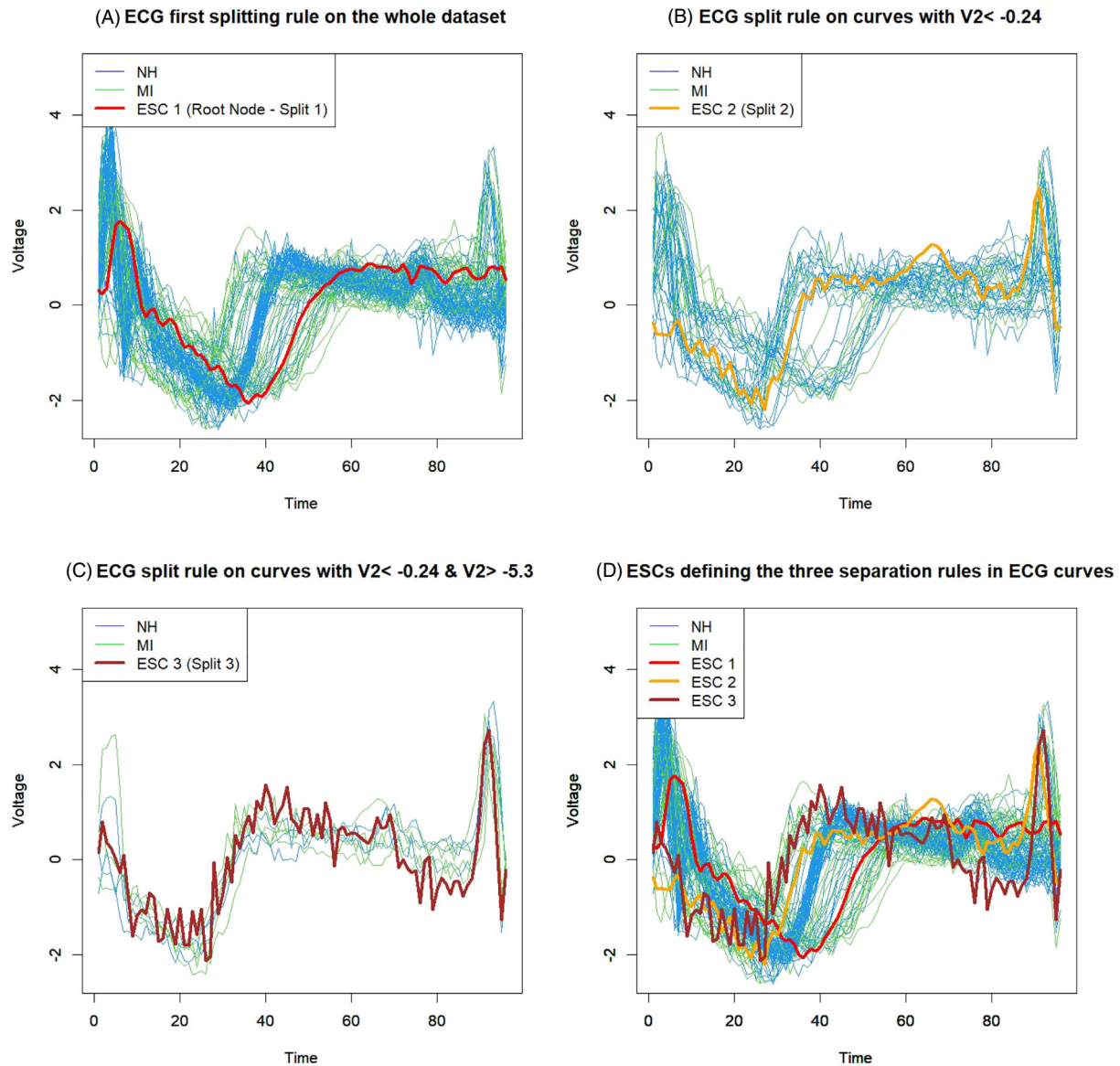
FIGURE 7 Empirical splitting curves (ESCs) of the FCT obtained using the three splits based on FPCs

the condition $v_{02} < -0.24$. Using the latter subset, the orange curve discriminates those functions based on the threshold value $v_{02} = -5.3$. The brown curve, in Figure 7C, describes the third ESC applied to the subset of curves satisfying the conditions $v_{02} < -0.24$ and $v_{02} \geq -5.3$. On the latter set of curves, the splitting rule based on the split value $v_{09} = -0.71$ is applied. Figure 7D illustrates the three ESCs overlapping the whole dataset to help to interpret the separation rules.

Figure 8 summarizes all the aspects of the splitting rules produced by the FPC in terms of ESCs. The graph also shows the structure of each terminal node displaying the curves, the functional mean, original classes, and functional deviance of each leaf. The read of Figure 8 highlights interesting results. Indeed, the first leaf (reading from left to right) is mainly composed of curves representing people with a cardiac problem, and thus the overall predicted class is "*Disease*." Effectively, 95% of the functions in that leaf benefits from an accurate prediction of MI. In other words, the first leaf is characterized by 95% of TPs and 5% of FPs. Figure 8 also shows the functional deviance of the leaves, computed using Equation (21). In addition to the classical measures to estimate the impurity of the terminal nodes, the functional deviance of the leaves can be used to evaluate the internal variability of the single leaf over the whole domain. The functional deviance of the first leaf highlights an interesting peculiarity of this subset of curves. Despite the final classification of the curves in this leaf is quite precise, there is a large functional variability in the middle part of the time domain. This circumstance suggests two different patterns present within the node and that the MI condition does not have only one
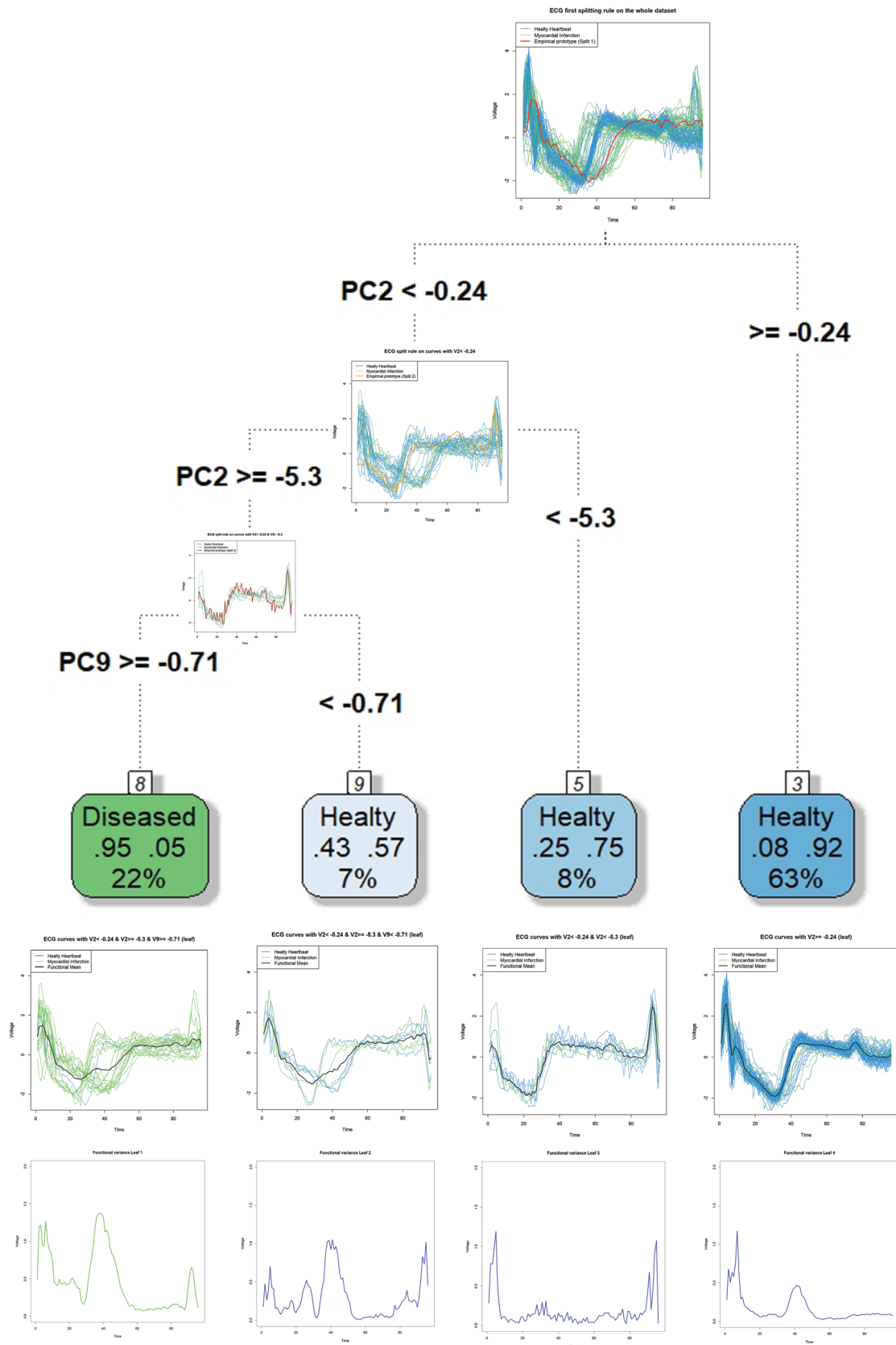
FIGURE 8   Details of the splitting rules generated by the FCT in terms of ESCs and leaves deviance. The functional deviance of the leaf is plotted below each terminal node. The blue and green colors are linked to the final predicted class of each terminal node. Green = MI; Blue = NH

type of functional form. The leaves number two and three are composed of seven and eight curves, respectively. The dominant prediction for those curves is "*Healthy*," but the impurity of those terminal nodes is high because there are high percentages of curves also representing "*Diseased*" people. For this reason, the percentages of FNs in these terminal nodes are 43% and 25%, but, in absolute numbers, the contribution to the increase of the misclassification error rate is quite low. Leaf 4 is dominated by TNs, and thus the overall predicted class is "*Healthy*." Only 8% of the leaf is composed by FNs, and thus the classification rule dictated by the first split rule based on the second FPC is quite good; in fact, it separates 63% of individuals, inserting them in a terminal node. As for leaf 1, the reading of the functional deviance of leaf 4 is quite interesting. Despite the high purity of this node, we observe an intensified variability at the beginning and in the middle part of the time domain. It could be a clue to the presence of two different patterns of healthy people that experienced cardiologists should carefully interpret to understand the meaning of these differences over the time domain. As for leaf 1, these differences could also be due to patient characteristics such as age, sex, comorbidities, or other possible covariates. Therefore, medical advice is essential to interpret this high variability in quite pure leaves.

Figure 9 provides details of the relative FBGSS of the four leaves. Particularly, Figure 9A,B provides details of leaf 1. Because the final prediction of leaf 1 is "*Disease*," this terminal node is composed of two types of curves, that is, TPs and FPs. TPs in leaf 1 are those curves describing ECGs of people that are really diseased and thus are correctly classified. FPs are those individuals whose curves should be classified as "*Healthy*" but they are not; thus, the latter functions are in the terminal node 1 for a mistake. Now, the most critical question is the following. Which are the similarities and dissimilarities between the functions of these two subgroups? In other words, which are the similarities that led to putting those people into the same terminal node? Furthermore, which are the dissimilarities that have been neglected, leading to a misclassification error? Figure 9A highlights the functional means of FPs and TPs in leaf 1. Instead, Figure 9B presents the plot of the relative FBGSS between FPs and TPs computed using Equation (25). According to the chart, we can state that the classification rule that led to assign these curves in the same terminal node neglected that these functions are slightly diverse in the first half of the time domain. Evidently, the similarity between FPs and TPs curves in the second part of the time domain led to uncertainty in identifying the labels of the curves. Thus, the pruning of the FCT pushed to neglect some features able to discriminate some curves in the first half of the time domain. Figure 9C–F depict the subgroups of TNs and FNs, their functional means, and the relative FBGSSs, respectively. Figure 9G,H show the details of leaf 4. The last terminal node directly comes from the split of the root node (see Figure 8). 92% of the curves in leaf 4 are TNs and describe "*Healthy*" patients that are correctly classified. However, 8% of the curves in leaf 4 are FNs. The plot of the relative FBGSS in Figure 9H helps us to explain why FNs are in the same terminal node of TNs. Effectively, the plot of FBGSS stresses that the groups are very similar, and it is very hard to discriminate between them. The only minor difference between TNs and FNs in leaf 4 can be found in the first quarter of the time domain, but their divergence is too small to be detected.

Figure 10 provides a detailed view of the cross-comparison between leaves 1 and 4. The latter leaves contain the highest number of curves; indeed, the leaf 1 considers 22% and leaf 4 contains 63% of the total number of curves. Moreover, they can both used as reference leaves because they are also the most accurate in predicting the "*Disease*" and "*Healthy*" classes, respectively (see Figure 8). Therefore, we omit discussing all the possible cross-comparisons of terminal nodes, and we focus only on leaves 1 and 4. Specifically, Figure 10A,B presents FPs of leaf 1 and TNs of leaf 4 functional means, and their relative FBLSS. Figure 10C,D shows FNs of leaf 4 and TPs of leaf 1 functional means, and their relative FBLSS. The latter charts allow us to try answering the most challenging issues. The first problem (Figure 10A,B) is to capture the difference between FPs in leaf 1 and TNs in leaf 4. In other words, we aim to explain the functional variability between healthy people that are correctly predicted and healthy people that are not correctly predicted by considering the most dominant decision rules in terms of the number of individuals in the leaves with opposite class predictions. The charts stress that the bad predictions of FPs in leaf 1 is probably due to a peculiar shapes of these curves in three parts of the domain (red curve's peaks). Indeed, the gray curve (FPs' functional mean in leaf 1) is very different from the blue function (TNs' functional mean in leaf 4), especially in the central area of the domain, which appears to be decisive for getting why FPs are badly assigned to leaf 1 instead of leaf 4. The second issue is the opposite, that is, trying to explain why some diseased patients (FNs in leaf 4) are assigned to a leaf that is mainly composed by healthy people and which is their functional between variability with respect to diseased people that are correctly classified in leaf 1 (TPs in leaf 1). Effectively, leaf 1 mainly predicts the disease status with 95% accuracy and thus understanding the difference with respect to this subgroup of TPs can help to detect peculiarities/anomalies in some ill patients, and probably also different patterns of disease. It is evident that, from a functional point of view, the sick and poorly classified patients of leaf 4 are very different from the well classified sick patients of leaf 1. The biggest difference is evidenced by the FBLSS peak exactly in the center of the time domain, where we observe a constant high value of FBLSS. Effectively, also looking at the green curve
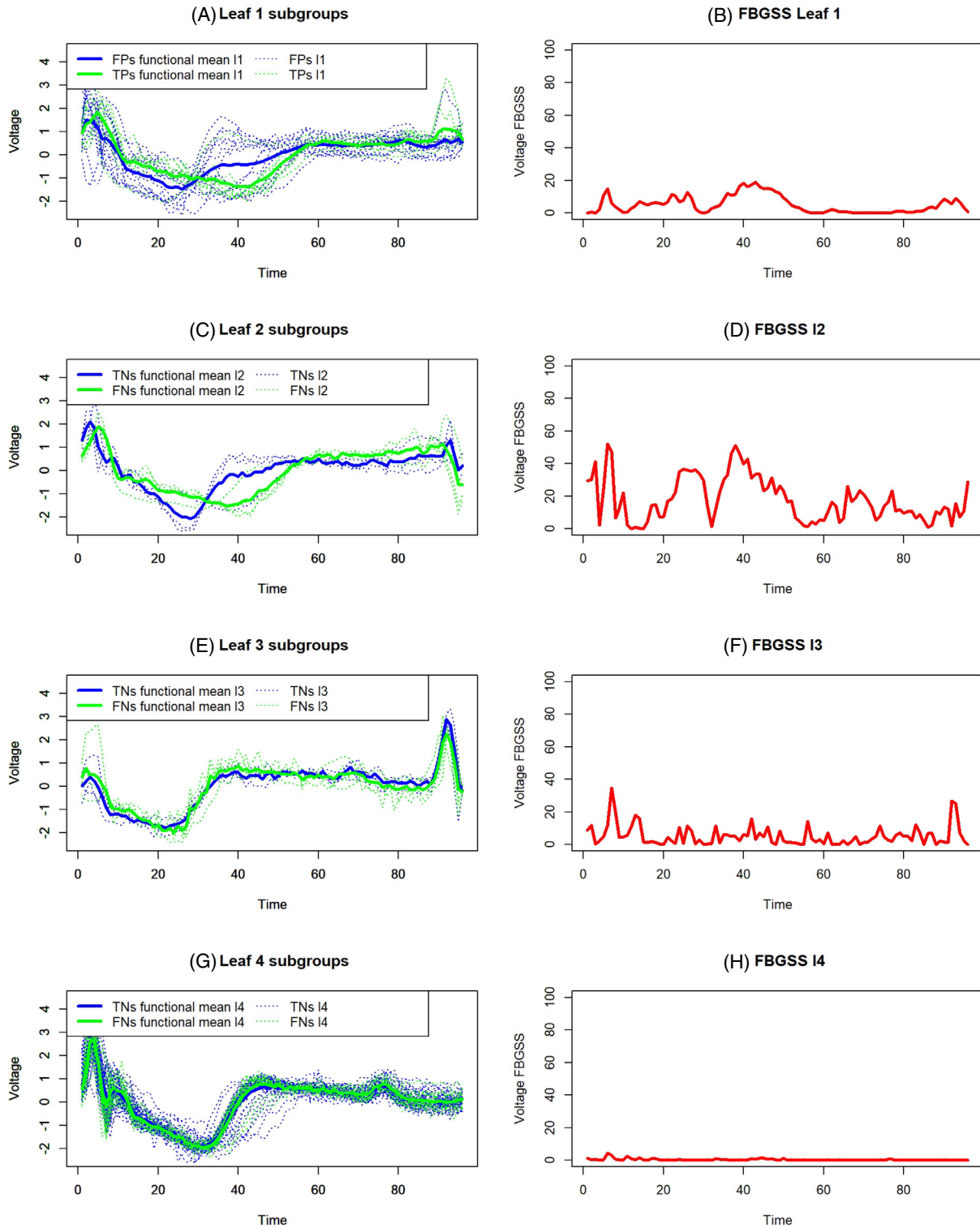
FIGURE 9   Details of leaves composition: Functional group means and relative FBGSS between TP-FP or TN-FN for the four leaves (FBGSS = functional between groups sum of squares)

(the prototype of TPs in leaf 1) in Figure 10C, it appears to be much lower than the gray curve (FNs in leaf 4) in the central part of the domain. It would seem that two different patterns characterize MI patients, and one of these patterns is difficult to recognize because it resembles the pattern of healthy people (the gray curve).

Because FRF provides an improvement over FBG, we present only the results of the FRF. Indeed, in the application of FBG with the ECG200 dataset, the FCTs are all dominated by the second FPC and, therefore, very correlated.
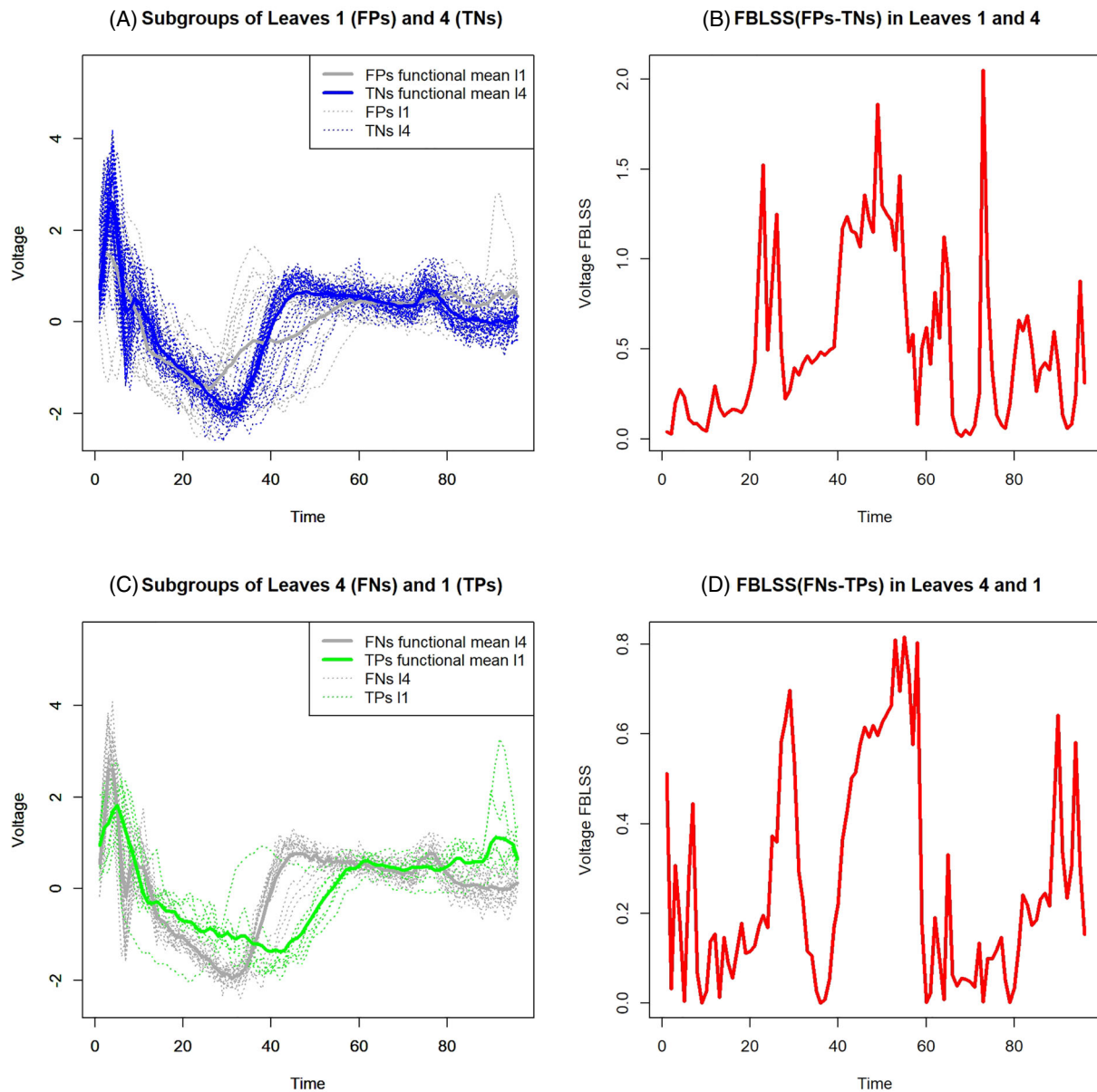
FIGURE 10   FBLSS 1 vs 4, and vice versa, that is, FPs vs TNs, and FNs vs TPs (FBGLS = functional between leaves sum of squares)

As discussed in Section 4, in FBG, FCTs are always very similar because at the top of the FCT there are always the same more discriminating features. Thus, we focus on FRF, which also produces the best results in accuracy and variance reduction. Using the FRF algorithm, the FCTs are not all governed by the second FPC. Indeed, at the top of the FCTs in the forest, there are often different FPCs (a picture of the first nine FCTs obtained is attached as supplementary material).

Figure 11 illustrates the influence of the FPCs in the FRF classifier based on the mean decrease accuracy index.[33,35] Unquestionably, the second FPC is the most powerful in distinguishing the classes of the outcome; nevertheless, the FPCs n. 1, 3, 5, and 9 also play an influential role in enhancing the accuracy of the functional classifier.

We implemented the FRF algorithm for different sizes of the forest and different numbers of FPCs by testing the accuracy both applying the OOBFD misclassification error estimation and using the test set (see Tables 1 and 2). A comparison with other competitive functional classification techniques is presented in Tables 3–5. The best performance computed using the OOBFD strategy on the training set is 89.00% with 12 FPCs and 9 trees. Instead, Table 2 exhibits the accuracy of the FRF classifier using the test set. The best accuracy on the test set is 93.00% with 9 FPCs and 9 FCTs. The world top accuracy reached for this dataset was 89.05% with the BOSS algorithm.[42] Hence, the FRF classifier based on FPCs beats the previous world record set by the BOSS algorithm.
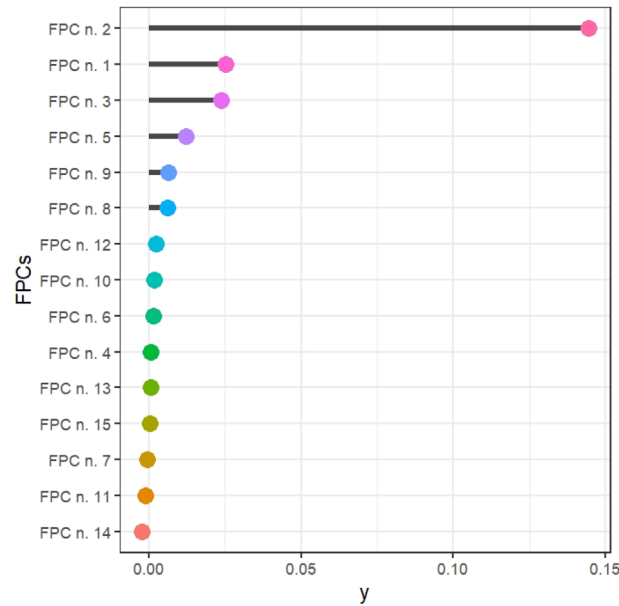
FIGURE 11 Interpreting the importance of FPCs for the FRF classifier based on FPCs (mean decrease accuracy index)

To compare the FRF based on FPCs to the FRF based on b-splines and the most recent and widespread methods to classify functional data, we provide the results of different approaches implemented in the fda.usc R package.[29] Tables 3–5 highlight that the other methods are not able to reach the performance of the FRF classifier. Indeed, the best accuracy of the FRF based on b-splines gives an accuracy of 87.00% on the training set using 29 FCTs, and the same result is achieved using the test set with 8 FCTs. Moreover, Table 4 shows that, using the functional K-NN classifier of the R package fda.usc, the best outcome on the training set is 92.00% with 3 nearest neighbors whereas the best result on the test set is 91.00% with 3 or 5 nearest neighbors. Finally, using the functional depth classifiers (Table 5), the best accuracy on the training set is 83.00% with the depth measure "*RPD*," and the best accuracy on the test set is 81.00% with the depth measure "*RP*." In summary, the FRF classifier based on FPCs shows exciting results and proves to be the best in terms of accuracy with this data when focusing on the test set.

A second application on the ECG200 dataset is performed using a fixed basis system. For brevity, in this case, we just show some essential results. The b-spline fixed basis system is composed of 98 b-splines of order 4. The most essential b-splines in discriminating the classes are the numbers 33, 32, and 34. The accuracy of the b-splines-based FRF classifier gives interesting results. The best accuracy on the training set is 87.00% using twenty-nine trees. The best accuracy on the test set is 87.00% with eight trees. This approach is interesting to reduce the dimensionality and get good accuracy. However, the former is inferior to the data-driven approach and the BOSS algorithm, which reached the previous record on this dataset. In summary, the interpretation is relatively complicated compared to the data-driven approach, and the classifier's performance is not up to par. For this reason, we consider the FRF based on FPCs as the "gold standard" of our proposal. Detailed information regarding the pruned FRF using b-splines are the b-splines importance in building FRF procedure are available as supplementary material.

5.2 | Results of the application to the ECG5000 dataset

The ECG5000 dataset[42] was initially used in the article "*A general framework for never-ending learning from time series streams.*"[45] After that, 5000 heartbeats were randomly selected. The training set is composed of 500 patients, and the test set contains 4500 persons. The dataset is made up of five groups of patients with different types of heart disease. The best accuracy on the training set is 97.28% with 4 FPCs and 20 trees. On the other hand, the best accuracy on the test set is 93.64% with 6 FPCs and 40 trees. The world top accuracy reached for this dataset is still 94.61% with the COTE algorithm. Using b-splines, the best accuracy on the training set is 95.20% with 29 trees, and the best accuracy on the test set is 93.87% with 24 trees. The best accuracy of the functional K-NN classifier on the training set is 95.00% with nine nearest neighbors. Instead, using the test set is 93.64% with nine nearest neighbors. Regarding the functional depth

TABLE 1 Accuracy for the ECG200 dataset using the training set: Accuracy of the FRF classifier according to different sizes of the forest and different numbers of FPCs

No. of trees	Number of functional principal components involved in the classifier FRF													
	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2	77.78	78.57	90.16	77.05	77.78	80.65	85.96	76.92	72.41	80.00	80.00	74.24	72.13	82.26
3	76.54	80.28	82.43	73.08	76.39	81.16	83.33	78.26	75.71	76.32	79.17	76.32	77.46	71.83
4	84.71	74.07	72.29	80.49	80.25	78.65	79.52	79.76	76.14	75.95	76.19	75.00	80.25	79.31
5	82.80	73.91	86.81	79.78	80.00	79.35	76.40	77.78	78.13	74.44	74.44	74.74	75.58	80.00
6	77.42	83.33	82.11	75.79	80.21	74.49	78.95	84.44	79.57	74.74	80.85	79.35	70.97	82.11
7	85.42	82.65	84.69	78.79	75.51	84.38	79.79	74.47	75.79	79.38	81.63	79.79	72.34	83.16
8	77.55	83.51	77.78	78.72	75.51	78.72	75.51	78.12	81.63	77.32	79.17	80.41	72.83	72.73
9	82.65	83.33	85.86	82.65	73.20	80.41	78.57	76.00	82.65	75.76	82.65	79.80	79.59	80.41
10	77.55	79.00	79.80	80.61	77.78	81.63	74.49	77.55	80.00	76.53	76.53	81.82	87.76	75.51
11	83.00	77.00	83.00	75.00	81.82	73.00	78.00	82.00	74.00	73.00	79.80	82.00	86.00	78.00
12	78.00	82.83	87.00	81.82	75.00	75.00	77.00	75.51	80.00	80.00	82.65	83.84	86.00	70.71
13	86.00	82.00	81.00	77.78	76.29	85.00	77.00	79.00	83.00	79.00	80.00	84.85	80.00	82.00
14	78.79	78.00	83.00	81.00	77.00	80.00	80.81	81.00	84.00	77.00	84.00	78.00	78.00	84.00
15	82.00	80.00	80.00	80.00	81.00	81.00	82.00	83.00	82.00	76.00	79.00	79.00	80.00	80.00
16	80.00	85.00	81.00	80.00	82.00	84.00	78.79	78.00	80.00	79.00	80.81	82.00	78.00	80.00
17	82.00	87.00	85.00	79.00	79.00	76.77	78.00	77.00	80.00	75.00	81.00	82.00	73.00	87.88
18	84.00	80.00	87.00	76.00	84.00	79.00	81.00	84.00	84.00	86.00	82.00	76.00	77.00	78.00
19	84.00	83.00	78.00	82.00	83.00	80.00	78.00	80.00	84.00	83.00	86.00	83.00	82.00	80.00
20	84.00	84.00	86.00	85.00	83.00	81.00	80.00	78.00	82.00	82.00	81.00	83.00	77.00	84.00
21	81.00	86.00	80.00	86.00	79.00	82.00	80.00	78.00	80.00	74.00	88.00	82.00	81.00	83.00
22	83.00	85.00	83.00	82.00	82.00	76.00	80.00	77.00	83.00	82.00	83.00	80.00	82.00	81.00
23	84.00	85.00	83.00	80.00	85.00	84.00	81.00	78.00	84.00	81.00	81.00	82.00	77.00	79.00
24	85.00	85.00	83.00	84.00	84.00	83.00	86.00	84.00	88.00	86.00	83.00	77.00	78.00	83.00
25	79.00	89.00	85.00	81.00	85.00	78.00	82.00	83.00	84.00	81.00	79.00	80.00	83.00	80.00
26	81.00	83.00	85.00	84.00	79.00	83.00	81.00	84.00	84.00	85.00	83.00	83.00	79.00	81.00
27	84.00	82.00	83.00	82.00	84.00	81.00	82.00	83.00	83.00	80.00	79.00	80.00	85.00	84.00
28	87.00	84.00	84.00	85.00	82.00	83.00	84.00	82.00	79.00	82.00	83.00	83.00	82.00	80.00
29	83.00	86.00	86.00	82.00	82.00	84.00	84.00	83.00	81.00	79.00	83.00	80.00	80.00	84.00
30	88.00	84.00	83.00	81.00	87.00	85.00	79.00	82.00	84.00	82.00	79.00	83.00	82.00	80.00
31	84.00	86.00	85.00	85.00	84.00	84.00	76.00	81.00	80.00	83.00	77.00	80.00	83.00	84.00
32	82.00	79.00	85.00	83.00	79.00	82.00	82.00	84.00	84.00	83.00	87.00	87.00	82.00	83.00
33	84.00	80.00	79.00	85.00	83.00	84.00	85.00	80.00	81.00	84.00	86.00	84.00	81.00	85.00
34	84.00	84.00	81.00	85.00	84.00	83.00	79.00	81.00	82.00	80.00	80.00	81.00	81.00	81.00
35	84.00	82.00	85.00	82.00	84.00	81.00	84.00	82.00	84.00	82.00	80.00	81.00	88.00	83.00
36	84.00	84.00	83.00	85.00	86.00	82.00	80.00	81.00	83.00	85.00	82.00	83.00	83.00	87.00
37	80.00	83.00	83.00	78.00	84.00	81.00	84.00	82.00	83.00	83.00	82.00	84.00	83.00	78.00
38	83.00	83.00	85.00	85.00	86.00	85.00	79.00	83.00	85.00	85.00	82.00	84.00	80.00	85.00
39	87.00	83.00	84.00	85.00	84.00	84.00	78.00	83.00	81.00	82.00	85.00	83.00	81.00	84.00
40	85.00	86.00	83.00	86.00	83.00	80.00	80.00	83.00	82.00	82.00	83.00	80.00	85.00	83.00
45	86.00	82.00	83.00	82.00	83.00	82.00	81.00	81.00	85.00	83.00	81.00	81.00	83.00	81.00
50	84.00	80.00	81.00	84.00	79.00	81.00	81.00	79.00	81.00	82.00	79.00	82.00	83.00	83.00

Note: The best accuracy on the training set is 89.00% with 12 FPCs and 9 trees.

TABLE 2 Accuracy for the ECG200 dataset using the test set: Accuracy of the FRF classifier according to different sizes of the forest and different numbers of FPCs

| No. of trees | Number of functional principal components involved in the classifier FRF | | | | | | | | | | | | | |
	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2	74.00	73.00	83.00	79.00	68.00	85.00	68.00	74.00	75.00	74.00	72.00	75.00	74.00	74.00
3	78.00	76.00	81.00	84.00	82.00	82.00	74.00	82.00	82.00	71.00	78.00	78.00	73.00	73.00
4	78.00	79.00	78.00	72.00	81.00	79.00	83.00	84.00	83.00	81.00	86.00	76.00	78.00	84.00
5	74.00	75.00	84.00	82.00	83.00	87.00	81.00	81.00	82.00	82.00	86.00	83.00	82.00	75.00
6	78.00	82.00	70.00	84.00	84.00	79.00	82.00	84.00	81.00	79.00	86.00	84.00	85.00	84.00
7	74.00	80.00	82.00	82.00	86.00	83.00	82.00	84.00	83.00	82.00	85.00	82.00	82.00	82.00
8	79.00	81.00	85.00	79.00	84.00	78.00	84.00	89.00	84.00	86.00	85.00	90.00	78.00	87.00
9	73.00	82.00	83.00	85.00	84.00	78.00	83.00	93.00	85.00	89.00	86.00	82.00	88.00	82.00
10	77.00	82.00	80.00	84.00	84.00	79.00	81.00	83.00	85.00	87.00	88.00	85.00	87.00	80.00
11	80.00	75.00	84.00	85.00	85.00	83.00	86.00	85.00	88.00	82.00	84.00	84.00	81.00	81.00
12	79.00	82.00	80.00	84.00	83.00	79.00	82.00	80.00	84.00	92.00	82.00	86.00	81.00	82.00
13	79.00	81.00	84.00	88.00	77.00	83.00	83.00	83.00	86.00	87.00	87.00	84.00	82.00	85.00
14	78.00	85.00	83.00	81.00	85.00	80.00	83.00	87.00	85.00	85.00	86.00	81.00	89.00	81.00
15	79.00	80.00	83.00	87.00	85.00	82.00	84.00	86.00	84.00	85.00	81.00	85.00	88.00	82.00
16	78.00	81.00	82.00	84.00	85.00	81.00	84.00	87.00	87.00	88.00	90.00	85.00	85.00	83.00
17	80.00	82.00	85.00	82.00	86.00	83.00	81.00	84.00	86.00	88.00	83.00	85.00	82.00	86.00
18	79.00	80.00	81.00	85.00	84.00	85.00	87.00	86.00	89.00	91.00	87.00	88.00	87.00	82.00
19	81.00	78.00	80.00	82.00	84.00	82.00	84.00	84.00	86.00	86.00	82.00	83.00	87.00	88.00
20	79.00	78.00	79.00	84.00	84.00	81.00	82.00	87.00	85.00	89.00	88.00	81.00	86.00	84.00
21	77.00	78.00	79.00	86.00	82.00	84.00	86.00	88.00	85.00	84.00	89.00	89.00	83.00	81.00
22	81.00	79.00	84.00	81.00	83.00	84.00	79.00	84.00	87.00	88.00	89.00	82.00	85.00	87.00
23	80.00	74.00	84.00	81.00	86.00	83.00	88.00	83.00	85.00	85.00	89.00	83.00	85.00	88.00
24	76.00	80.00	82.00	82.00	81.00	85.00	88.00	87.00	86.00	85.00	90.00	85.00	85.00	85.00
25	78.00	79.00	85.00	85.00	85.00	83.00	84.00	85.00	87.00	87.00	87.00	82.00	86.00	86.00
26	79.00	77.00	82.00	86.00	80.00	82.00	82.00	87.00	88.00	85.00	86.00	88.00	86.00	90.00
27	80.00	76.00	80.00	85.00	81.00	84.00	84.00	87.00	83.00	89.00	82.00	86.00	88.00	85.00
28	79.00	77.00	79.00	82.00	85.00	84.00	84.00	85.00	86.00	89.00	86.00	87.00	87.00	89.00
29	79.00	81.00	79.00	86.00	82.00	84.00	84.00	87.00	87.00	89.00	84.00	82.00	89.00	86.00
30	78.00	81.00	83.00	84.00	82.00	86.00	86.00	86.00	89.00	88.00	84.00	88.00	87.00	84.00
31	78.00	81.00	80.00	86.00	85.00	82.00	89.00	88.00	86.00	85.00	85.00	89.00	88.00	87.00
32	77.00	77.00	85.00	84.00	82.00	82.00	85.00	86.00	88.00	93.00	87.00	90.00	87.00	89.00
33	78.00	80.00	80.00	84.00	85.00	83.00	81.00	83.00	88.00	86.00	88.00	83.00	84.00	90.00
34	80.00	84.00	83.00	81.00	84.00	84.00	85.00	86.00	85.00	87.00	83.00	84.00	86.00	87.00
35	79.00	79.00	82.00	84.00	83.00	86.00	82.00	87.00	83.00	88.00	86.00	87.00	87.00	85.00
36	80.00	81.00	82.00	82.00	85.00	83.00	83.00	87.00	83.00	87.00	89.00	88.00	88.00	88.00
37	79.00	80.00	83.00	87.00	84.00	83.00	83.00	85.00	88.00	90.00	88.00	84.00	84.00	84.00
38	77.00	81.00	81.00	83.00	83.00	84.00	86.00	84.00	88.00	87.00	88.00	86.00	88.00	85.00
39	78.00	80.00	80.00	83.00	84.00	82.00	86.00	88.00	87.00	88.00	88.00	86.00	85.00	87.00
40	79.00	80.00	81.00	81.00	85.00	84.00	85.00	88.00	84.00	89.00	89.00	87.00	86.00	85.00
45	79.00	83.00	81.00	83.00	85.00	86.00	84.00	86.00	90.00	89.00	86.00	87.00	86.00	86.00
50	78.00	79.00	83.00	85.00	86.00	85.00	83.00	86.00	87.00	87.00	87.00	86.00	85.00	86.00

Note: The best accuracy on the test set is 93.00% with 9 FPCs and 9 trees, or 11 FPCs and 32 trees. FRF classifier performs better than the other methods. The world top accuracy reached for this dataset was 89.05% with the BOSS algorithm.
https://www.timeseriesclassification.com/description.php?Dataset=ECG200

TABLE 3 Accuracy for the ECG200: Accuracy of the FRF-B-spline classifier according to different sizes of the forest and a fixed number of B-splines

No. of trees	Accuracy on training set	Accuracy on test set
2	83.33	86.00
3	81.82	81.00
4	80.25	81.00
5	77.42	80.00
6	81.05	86.00
7	82.47	83.00
8	82.65	76.00
9	78.57	81.00
10	80.00	86.00
11	77.78	82.00
12	83.00	81.00
13	85.00	83.00
14	80.00	86.00
15	81.00	81.00
16	84.00	84.00
17	82.00	83.00
18	81.00	80.00
19	81.00	79.00
20	82.00	82.00
21	86.00	83.00
22	87.00	81.00
23	80.00	81.00
24	87.00	83.00
25	87.00	83.00
26	83.00	83.00
27	85.00	80.00
28	83.00	86.00
29	83.00	83.00
30	83.00	82.00
31	84.00	79.00
32	84.00	81.00
33	84.00	83.00
34	82.00	81.00
35	81.00	82.00
36	83.00	83.00
37	85.00	82.00
38	85.00	83.00
39	85.00	81.00
40	84.00	83.00
45	84.00	83.00
50	85.00	82.00

Note: The best accuracy on the training set is 87.00% with 29 trees. The best accuracy on the test set is 87.00% with 8 trees.

TABLE 4 Accuracy for the ECG200 dataset: Functional classification using the K-NN classifier of the R package fda.usc

No. of NNs in functional KNN	Accuracy on training set	Accuracy on test set
1	86.00	89.00
3	92.00	91.00
5	88.00	91.00
7	86.00	88.00
9	81.00	86.00
11	78.00	88.00
13	77.00	85.00
15	76.00	82.00
17	75.00	82.00
19	71.00	80.00
21	72.00	76.00
23	74.00	75.00
25	75.00	76.00

Note: The best accuracy on the training set is 92.00% with 3 nearest neighbors. The best accuracy on the test set is 91.00% with 3 or 5 nearest neighbors.

TABLE 5 Accuracy for the ECG200: Functional classification using depth classifiers of the R package fda.usc

Functional depth measure	Accuracy on the training set	Accuracy on the test set
	Train	Test
RP	78.00	81.00
mode	80.00	79.00
RT	44.00	43.00
FM	76.00	74.00
RPD	83.00	80.00

Note: The best accuracy on the training set is 83% with the depth measure "RPD." The best accuracy on the test set is 81% with the depth measure "RP." depth.RP computes the random projection depth (see Cuevas et al[17]). depth.mode implements the modal depth (see Cuevas et al[17]). depth.RT implements the random tukey depth (see Cuesta-Albertos and Nieto-Reyes[52]). depth.FM computes the integration of an univariate depth along the axis x (see Fraiman and Muniz[53]). It is also known as integrated depth. depth.RPD implements a depth measure based on random projections possibly using several derivatives (see Cuevas et al[17]).

classifiers of the R package fda.usc, the best result on the training set is 93.6% with the depth measure "mode," whereas the best accuracy on the test set is 91.87% with the same depth measure. The FRF based on FPCs reaches very high accuracy in this second application by testing the misclassification error rate via the OOBFD strategy. Using the test set, the performances in terms of accuracy are pretty similar if considering FRF with FPCs, FRF with b-splines, or functional K-NN. The interpretation described for the ECG200 dataset with two classes can be extended to cases where the number of classes is greater than two. More details about the application on the ECG5000 are available as supplementary materials.

6 | DISCUSSION AND CONCLUSIONS

Today, thanks to technological progress, we can collect vast amounts of biomedical data. These data often come from medical devices, apps, and different sensors that produce high-frequency observations, for example, for monitoring the heart with ECGs, controlling brain activities via EEGs, assessing lung function using pulmonary function tests (PFTs), recordings of various types of signals to study sleep disorders. Consequently, recently, many approaches have been developed to deal with this data because traditional statistical methods can fail. One of the most critical issues is, for example, the

so-called curse of dimensionality that is linked with many problematical drawbacks when working with high-dimensional biological data.

This study offers a classification strategy for high-dimensional biomedical data that merges FDA and tree-based procedures. FCTs, functional bagging, and FRF are considered possible extensions of the classical classification trees, bagging, and random forest approach to deal with functional data. The idea of combining FDA and decision trees is relatively novel. Few articles dealt with this topic in recent years, although the literature on supervised classification in the FDA context is very lively. For example, Moller et al[25] recommended selecting features based on the mean of the function within determined intervals of the domain, whereas ElHaouij et al[46] and Gregorutti et al[24] concentrated on the wavelet basis representation with particular applications.

Our approach is quite different and suggests two possible strategies on a data-driven basis and fixed-basis systems. However, the objective of the proposed method is multiple. The first goal is connected with the introduction of the FDA because it permits a robust dimensionality reduction with an interpretation linked to different parts of the time domain and quick perception.[2,29,47] In addition, FDA also provides extra information on the behavior and variability of the curves over time.[7] Finally, treating functions as single objects permits us to employ some concepts of similarity between statistical units, which are very charming because they only take into account the primary features of the functions.[1] The second purpose is linked to the introduction of FCTs combined with FDA. In particular, by introducing functional principal components as possible features to train a classification tree, it is possible to apply a dimensionality reduction that offers uncorrelated predictors and reduces unnecessary noise for classification purposes. The third and most innovative element of this research is undoubtedly linked to the functional interpretation of FCTs and their decision rules. By pooling FPCs and CTs, and including the notions of empirical separation and theoretical separation curves, a straightforward reading of the classification rules in the functional field is obtained. Besides, the article offers a series of innovative functional measures and tools to evaluate and interpret the terminal nodes of a functional classification tree. Indeed, a fourth original aim of this article is the presentation of different supplementary criteria for assessing the quality of the leaves and interpret the patterns of FNs, FPs, TNs, and TPs in the context of FCTs. We present different measures such as the FBGSS and FBLSS with this purpose. Ultimately, in addition to the original component related to the interpretation and new functional measures, the results of the functional classifiers FCT and FRF, in terms of accuracy, are excellent. In fact, using one of the two datasets, the FRF based on FPCs breaks the classification world record. In examining the results, we also compare the FRF scheme with a classification based on b-splines. The latter is not as efficient as the one based on a data-driven basis. The causes are essentially three. First, a data-driven basis system better accommodates, by definition, to our data, catching the variability and, hence, the primary information we need. Second, the interpretation of a fixed basis system is aseptic, intricate, and not engaging in considering the time domain. Finally, in terms of accuracy, the performance of the functional classifier based on a fixed basis system is slightly lower than the strategy using a data-driven basis.

In the application, we have used two datasets regarding ECG data. We are aware of the limitations of this study because the ECG literature is huge and has a long history. To read, analyze, and interpret ECGs, many methods have been proposed in the literature for many decades in the most important international scientific journals, and books.[6,48-50] For this reason, we are conscious that the reading and interpretation of the results for medical purposes requires an accurate interdisciplinary collaboration with experts. Nevertheless, the main goal of this research is to provide methodological tools to aid the interpretation of biomedical signals for professionals who are experts in specific pathologies. Effectively, our method, even if we used ECG signals as an example, can be applied to any biomedical signal, for example, EEG and PFTs. Naturally, this research focuses on medical data, but the methodological approach can be adapted to any time series and also to functions whose domain is different from time. This study shows the application on different datasets not affected by the missing data problem. However, in the functional context, the problem of the existence of some missing data in the time domain does not significantly affect the analysis because the functions and FPCs can still be reconstructed with smoothing techniques. On the other hand, the most severe problem could arise when there are truncated data in the temporal domain, that is, observations that are definitively interrupted at a specific instant of time. There is lively and engaging literature on the problem of performing FPCA in such cases, for example, Shi et al.[51] Once FPCs have been reconstructed, the proposed approach remains easily implementable.

The proposed line of research, that is, combining FDA and tree-based supervised classification methods, is appealing and promising. Several potential future extensions and applications may be introduced. The most direct future improvement is to extend this proposal to different types of basis systems such as Fourier or Wavelets basis or create other types of distance or new interpretative tools. From an applicative perspective, as mentioned before, many interesting applications may be done, for example, for monitoring brain activities via EEGs, assessing lung function using pulmonary function tests (PFTs), or studying the different signals adopted to analyze sleep disorders.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID

Fabrizio Maturo https://orcid.org/0000-0002-2362-4970

REFERENCES

1. Ramsay J, Silverman B. *Functional Data Analysis*. 2nd ed. New York, NY: Springer; 2005.
2. Ferraty F, Vieu P. *Nonparametric Functional Data Analysis*. New York, NY: Springer; 2006.
3. Zhou Y, Sedransk N. Functional data analytic approach of modeling ECG T-wave shape to measure cardiovascular behavior. *Ann Appl Stat*. 2009;3(4):1382-1402. doi:10.1214/09-aoas273
4. Sørensen H, Goldsmith J, Sangalli L. An introduction with medical applications to functional data analysis. *Stat Med*. 2013;32(30):5222-5240. doi:10.1002/sim.5989
5. Ieva F, Paganoni AM. Risk prediction for myocardial infarction via generalized functional regression models. *Stat Methods Med Res*. 2016;25(4):1648-1660. doi:10.1177/0962280213495988
6. Matabuena M, Vidal JC, Hayes PR, Saavedra-Garcia M, Trillo FH. Application of functional data analysis for the prediction of maximum heart rate. *IEEE Access*. 2019;7:121841-121852. doi:10.1109/access.2019.2938466
7. Cuevas A. A partial overview of the theory of statistics with functional data. *J Stat Plan Infer*. 2014;147:1-23. doi:10.1016/j.jspi.2013.04.002
8. Aguilera A, Aguilera-Morillo M. Penalized PCA approaches for B-spline expansions of smooth functional data. *Appl Math Comput*. 2013;219(14):7805-7819. doi:10.1016/j.amc.2013.02.009
9. Bongiorno E, Goia A. Describing the concentration of income populations by functional principal component analysis on Lorenz curves. *J Multivar Anal*. 2019;170:10-24.
10. Maturo F, Ferguson J, Di Battista T, Ventre V. A fuzzy functional K-means approach for monitoring Italian regions according to health evolution over time. *Soft Comput*. 2019;24(18):13741-13755. doi:10.1007/s00500-019-04505-2
11. Zanin Zambom A, Collazos J, Dias R. Functional data clustering via hypothesis testing K-means. *Comput Stat*. 2018;34(2):527-549. doi:10.1007/s00180-018-0808-9
12. Maturo F. Unsupervised classification of ecological communities ranked according to their biodiversity patterns via a functional principal component decomposition of Hill's numbers integral functions. *Ecol Indic*. 2018;90:305-315. doi:10.1016/j.ecolind.2018.03.013
13. Fortuna F, Maturo F, Di Battista T. Clustering functional data streams: unsupervised classification of soccer top players based on Google trends. *Qual Reliab Eng Int*. 2018;34(7):1448-1460. doi:10.1002/qre.2333
14. Maturo F, Migliori S, Paolone F. Measuring and monitoring diversity in organizations through functional instruments with an application to ethnic workforce diversity of the U.S. federal agencies. *Comput Math Organ Theory*. 2019;25(4):357-388. doi:10.1007/s10588-018-9267-7
15. Maturo F, Balzanella A, Di Battista T. Building statistical indicators of equitable and sustainable well-being in a functional framework. *Soc Indic Res*. 2019;146(3):449-471. doi:10.1007/s11205-019-02137-5
16. Carcenac M, Redif S. Application of the sequential matrix diagonalization algorithm to high-dimensional functional MRI data. *Comput Stat*. 2019;35(2):579-605. doi:10.1007/s00180-019-00925-8
17. Cuevas A, Febrero M, Fraiman R. Robust estimation and classification for functional data via projection-based depth notions. *Comput Stat*. 2007;22(3):481-496. doi:10.1007/s00180-007-0053-0
18. Aguilera-Morillo M, Aguilera A, Escabias M, Valderrama MJ. Penalized spline approaches for functional logit regression. *TEST*. 2012;22(2):251-277. doi:10.1007/s11749-012-0307-1
19. Escabias M, Aguilera AM, Aguilera-Morillo MC. Functional PCA and base-line logit models. *J Classif*. 2014;31(3):296-324. doi:10.1007/s00357-014-9162-y
20. Yu Y, Lambert D. Fitting trees to functional data, with an application to time-of-day patterns. *J Comput Graph Stat*. 1999;8(4):749-762. doi:10.1080/10618600.1999.10474847
21. Balakrishnan S, Madigan D. Decision trees for functional variables. Paper presented at: Proceedings of the 6th International Conference on Data Mining (ICDM'06); 2006. 10.1109/icdm.2006.49
22. Nerini D, Ghattas B. Classifying densities using functional regression trees: applications in oceanology. *Comput Stat Data Anal*. 2007;51(10):4984-4993. doi:10.1016/j.csda.2006.09.028
23. Fan G, Cao J, Wang J. Functional data classification for temporal gene expression data with kernel-induced random forests; Paper presented at: Proceedings of the 2010 IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology; 2010; IEEE.
24. Gregorutti B, Michel B, Saint-Pierre P. Grouped variable importance with random forests and application to multiple functional data analysis. *Comput Stat Data Anal*. 2015;90:15-35. doi:10.1016/j.csda.2015.04.002
25. Möller A, Tutz G, Gertheiss J. Random forests for functional covariates. *J Chemom*. 2016;30(12):715-725. doi:10.1002/cem.2849
26. Haouij NE, Poggi JM, Ghozi R, Sevestre-Ghalila S, Jaïdane M. Random forest-based approach for physiological functional variable selection for driver's stress level classification. *Stat Methods Appl*. 2018;28(1):157-185. doi:10.1007/s10260-018-0423-5
27. Belli E, Vantini S. Measure inducing classification and regression trees for functional data; 2020. arXiv preprint arXiv:2011.00046.
28. Ferraty F, Vieu P. Curves discrimination: a nonparametric functional approach. *Comput Stat Data Anal*. 2003;44(1–2):161-173. doi:10.1016/S0167-9473(03)00032-x

29. Febrero-Bande M, de la Fuente MO. Statistical computing in functional data analysis: the R package fda.usc. *J Stat Softw*. 2012;5(4):1-28. doi:10.18637/jss.v051.i04

30. Jacques J, Preda C. Model-based clustering for multivariate functional data. *Comput Stat Data Anal*. 2014;71:92-106. doi:10.1016/j.csda.2012.12.004

31. Hyafil L, Rivest RL. Constructing optimal binary decision trees is NP-complete. *Inf Process Lett*. 1976;5(1):15-17. doi:10.1016/0020-0190(76)90095-8

32. Quinlan JR. Induction of decision trees. *Mach Learn*. 1986;1(1):81-106. doi:10.1023/A:1022643204877

33. Hastie T, Tibshirani R, Friedman J. *The Elements of Statistical Learning - Data Mining, Inference, and Prediction*. Springer Series in Statistics. New York, NY: Springer; 2009.

34. James G, Witten D, Hastie T, Tibshirani R. *An Introduction to Statistical Learning*. New York, NY: Springer; 2013.

35. Therneau T, Atkinson B, Ripley B. rpart: recursive partitioning and regression trees. R package version 4.1–15; 2019.

36. Ripley B. *Pattern Recognition and Neural Networks*. Cambridge, UK: Cambridge University Press; 1996.

37. Breiman L. Bagging predictors. *Mach Learn*. 1996;24:123-140. doi:10.1007/bf00058655

38. Ho TK. The random subspace method for constructing decision forests. *IEEE Trans Pattern Anal Mach Intell*. 1998;20(8):832-844. doi:10.1109/34.709601

39. Breiman L. Random forests. *Mach Learn*. 2001;45(1):5-32. doi:10.1023/a:1017934522171

40. Liaw A, Wiener M. Classification and regression by randomForest. *R News*. 2002;2(3):18-22.

41. Olszewski R. Generalized feature extraction for structural pattern recognition in time-series data. Technical report, Carnegie-Mellon Univ. Pittsburgh PA School of Computer Science; 2001.

42. Bagnall A, Lines J, Vickers W, Keogh E. The UEA & UCR time series classification repository; 2021.

43. Ramsay JO, Graves S, Hooker G. fda: functional data analysis. R package version 5.1.9; 2020.

44. Milborrow S. rpart.plot: plot 'rpart' models: an enhanced version of 'plot.rpart'. R package version 3.1.0; 2021.

45. Chen Y, Hao Y, Rakthanmanon T, Zakaria J, Hu B, Keogh E. A general framework for never-ending learning from time series streams. *Data Min Knowl Disc*. 2015;29(6):1622-1664.

46. El Haouij N, Poggi JM, Ghozi R, Sevestre-Ghalila S, Jaïdane M. Random forest-based approach for physiological functional variable selection for driver's stress level classification. *JISS*. 2019;28(1):157-185. doi:10.1007/s10260-018-0423-5

47. Maturo F, Fortuna F, Battista TD. Testing equality of functions across multiple experimental conditions for different ability levels in the IRT context: the case of the IPRASE TLT 2016 survey. *Soc Indic Res*. 2018;146(1–2):19-39. doi:10.1007/s11205-018-1893-4

48. Guo M, Huang ML, Bai Z, Hsieh K. Important ECG diagnosis-aiding indices of ventricular septal defect children with or without congestive heart failure. *Stat Med*. 2001;20(7):1125-1141. doi:10.1002/sim.748

49. Zhou Y, Sedransk N. A new functional data-based biomarker for monitoring cardiovascular behavior. *Stat Med*. 2012;32(1):153-164. doi:10.1002/sim.5518

50. Xie S, Lawniczak A. Feature extraction of epileptic EEG in spectral domain via functional data analysis; 2019; SCITEPRESS - Science and Technology Publications

51. Shi H, Dong J, Wang L, Cao J. Functional principal component analysis for longitudinal data with informative dropout. *Stat Med*. 2021;40(3):712-724.

52. Cuesta-Albertos JA, Nieto-Reyes A. The random Tukey depth. *Comput Stat Data Anal*. 2008;52(11):4979-4988.

53. Fraiman R, Muniz G. Trimmed means for functional data. *Test*. 2001;10(2):419-440.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

Supplementary material of the paper "*Pooling random forest and functional data analysis for biomedical signals supervised classification: theory and application to electrocardiogram data*"

Abbreviations: FDA, Functional Data Analysis; FPCs, Functional Principal Components; FCTs, Functional Classification Trees; FRF, Functional Random Forest; FBG, Functional Bagging; OOBFD, Out of Bag Functional Data; IBFD, In-Bag Functional Data; ESC, Empirical Splitting Curve; TSC, Theoretical Splitting Curve; FBGSS, Functional Between Groups Sum of Squares; FBLSS, Functional Between Leaves Sum of Squares.
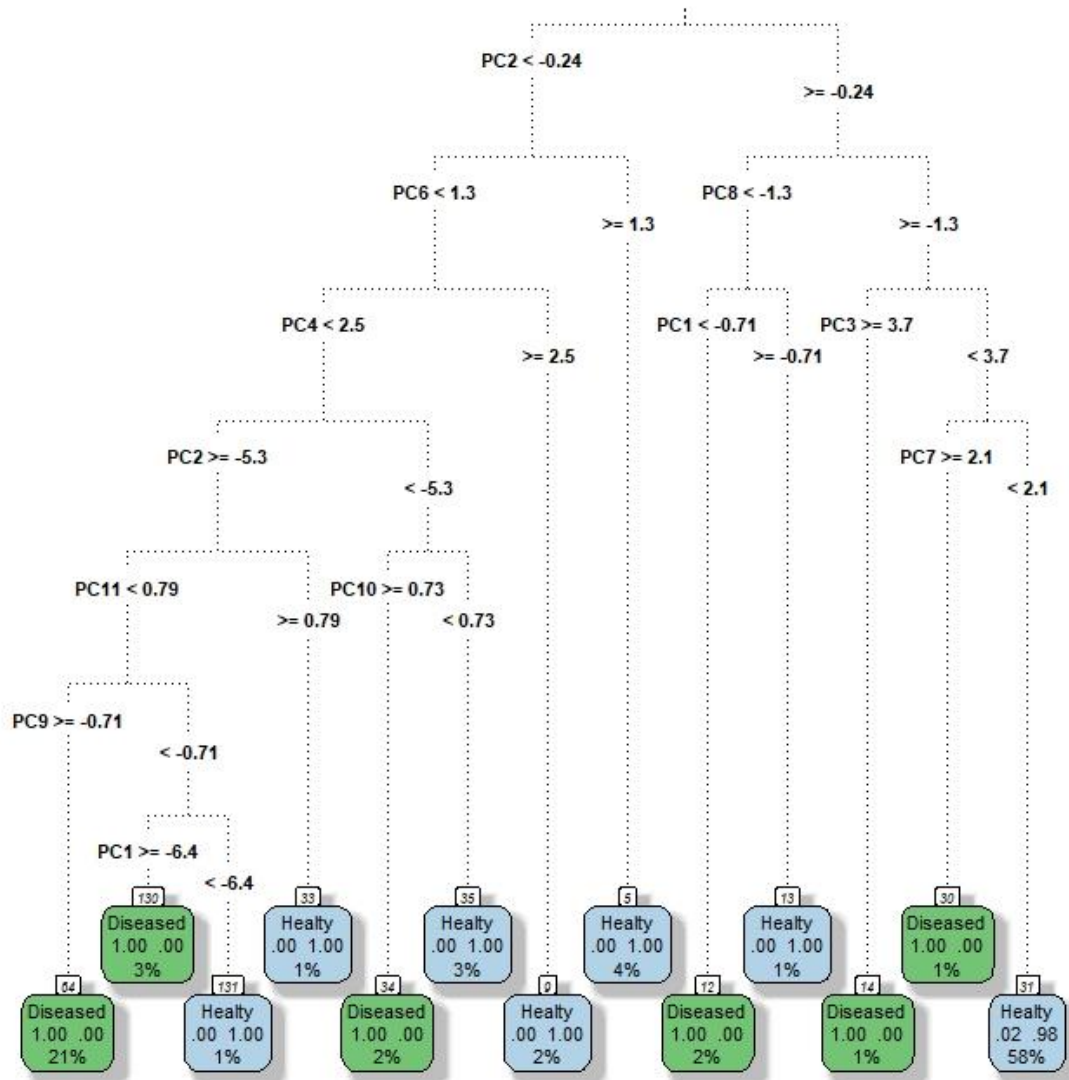
Figure 1. Non-pruned FCT based on the FPCs of the ECG200 dataset. The green leaves indicate a classification as diseased, while the blue leaves lead to the classification of the subjects as healthy. The relative frequencies of subjects of the two groups that make up the terminal leaf are indicated in the boxes. The percentage indicated at the bottom of the box indicates the fraction of people classified following the decision rule that leads to that terminal leaf.
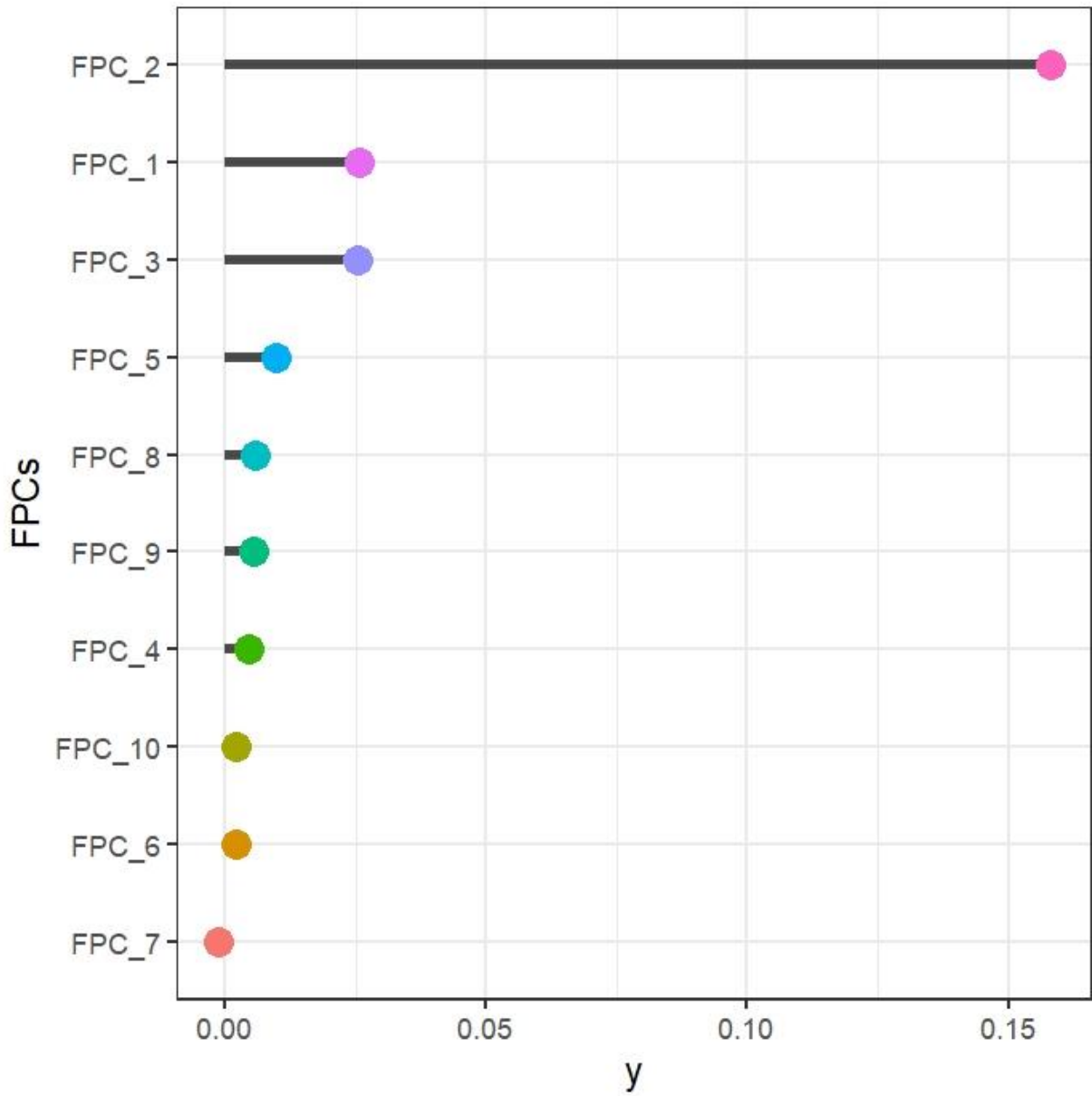
Figure 2. FPCs importance in a single FCT based on FPCs.
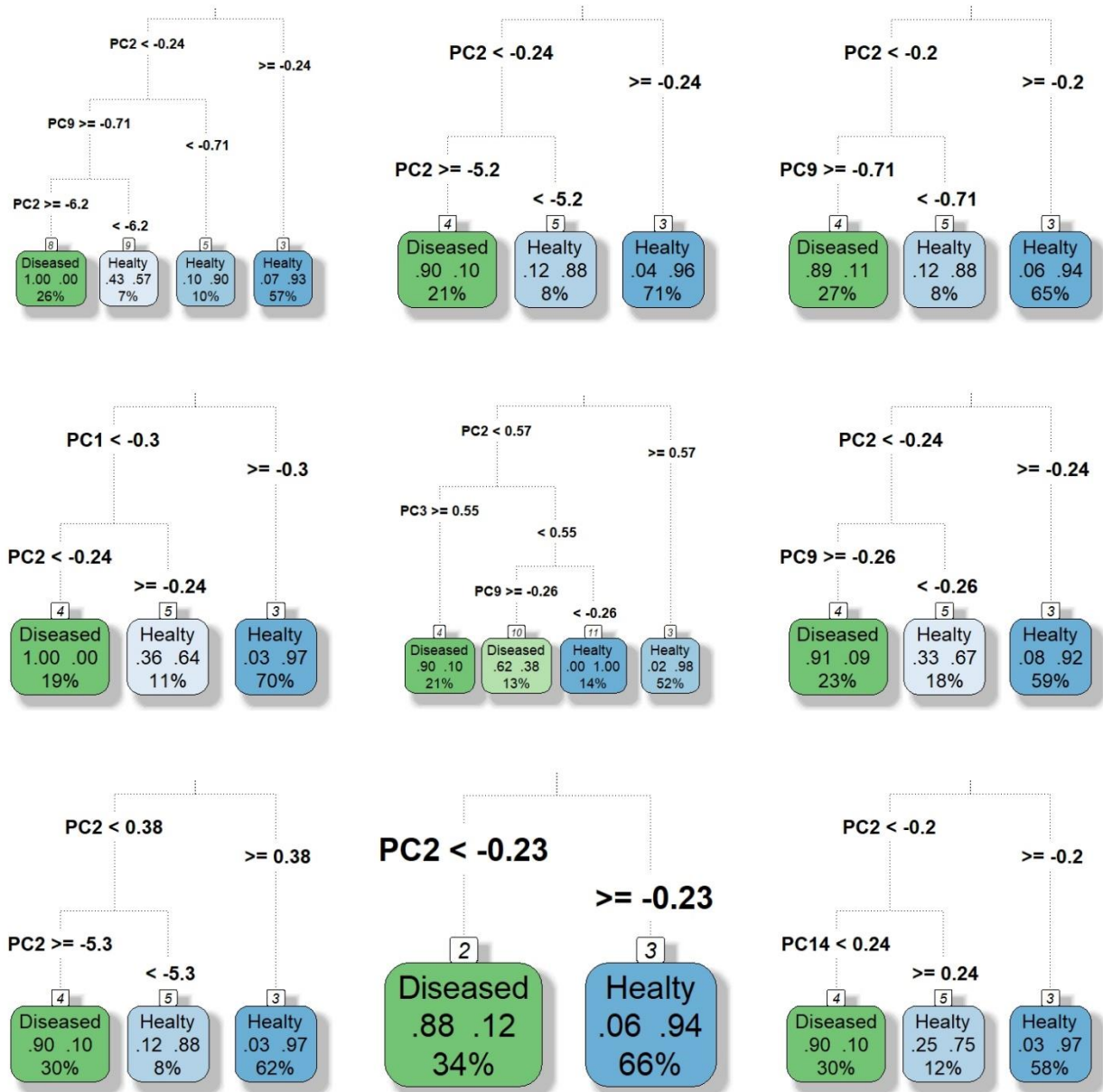
Figure 3. The first nine FCTs obtained using the functional bagging procedure based on the FPCs decomposition. The green leaves indicate a classification as diseased, while the blue leaves lead to the classification of the subjects as healthy. The relative frequencies of subjects of the two groups that make up the terminal leaf are indicated in the boxes. The percentage indicated at the bottom of the box indicates the fraction of people classified following the decision rule that leads to that terminal leaf.

Figure 4. The first nine FCTs obtained using the FRF algorithm based on the FPCs decomposition.
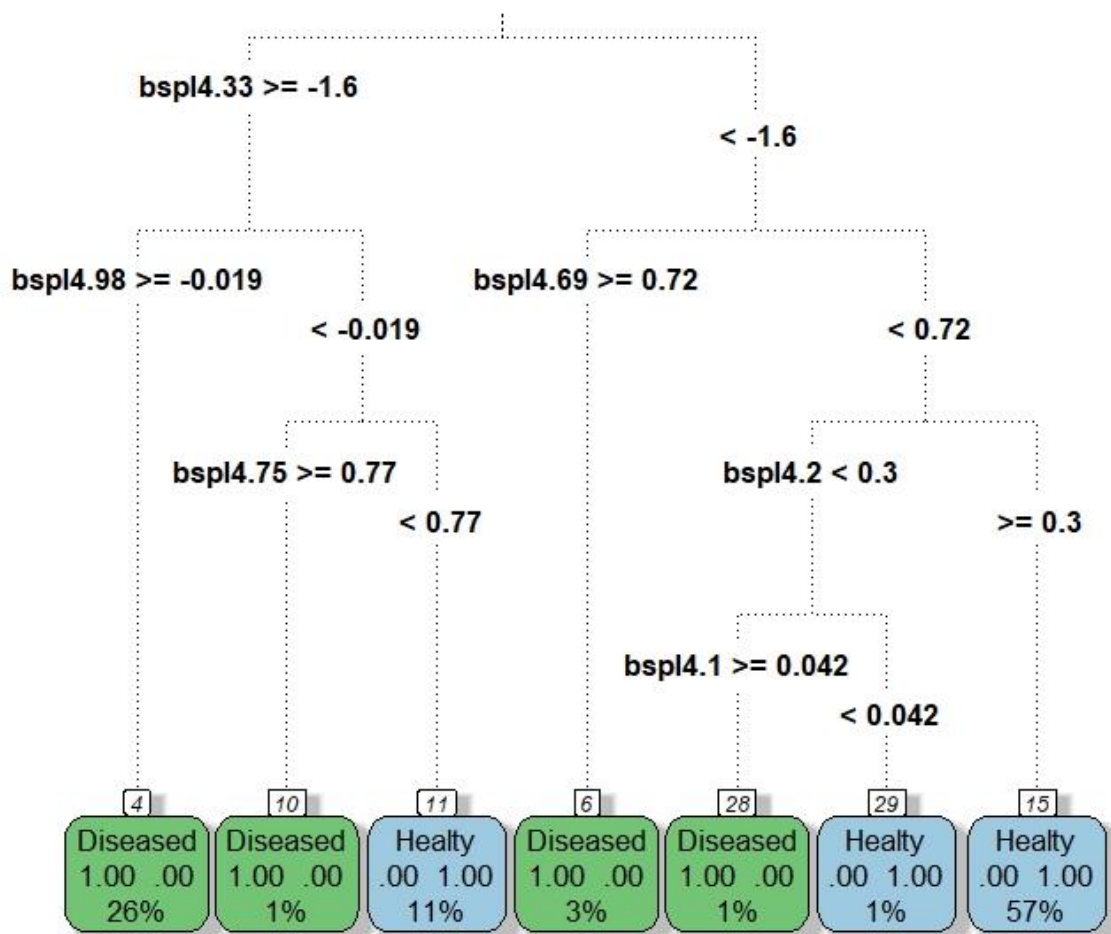
Figure 5. Pruned FRF using b-splines. The green leaves indicate a classification as diseased, while the blue leaves lead to the classification of the subjects as healthy. The relative frequencies of subjects of the two groups that make up the terminal leaf are indicated in the boxes. The percentage indicated at the bottom of the box indicates the fraction of people classified following the decision rule that leads to that terminal leaf.
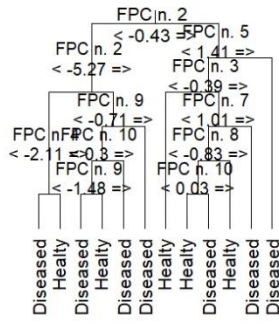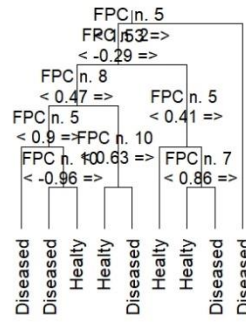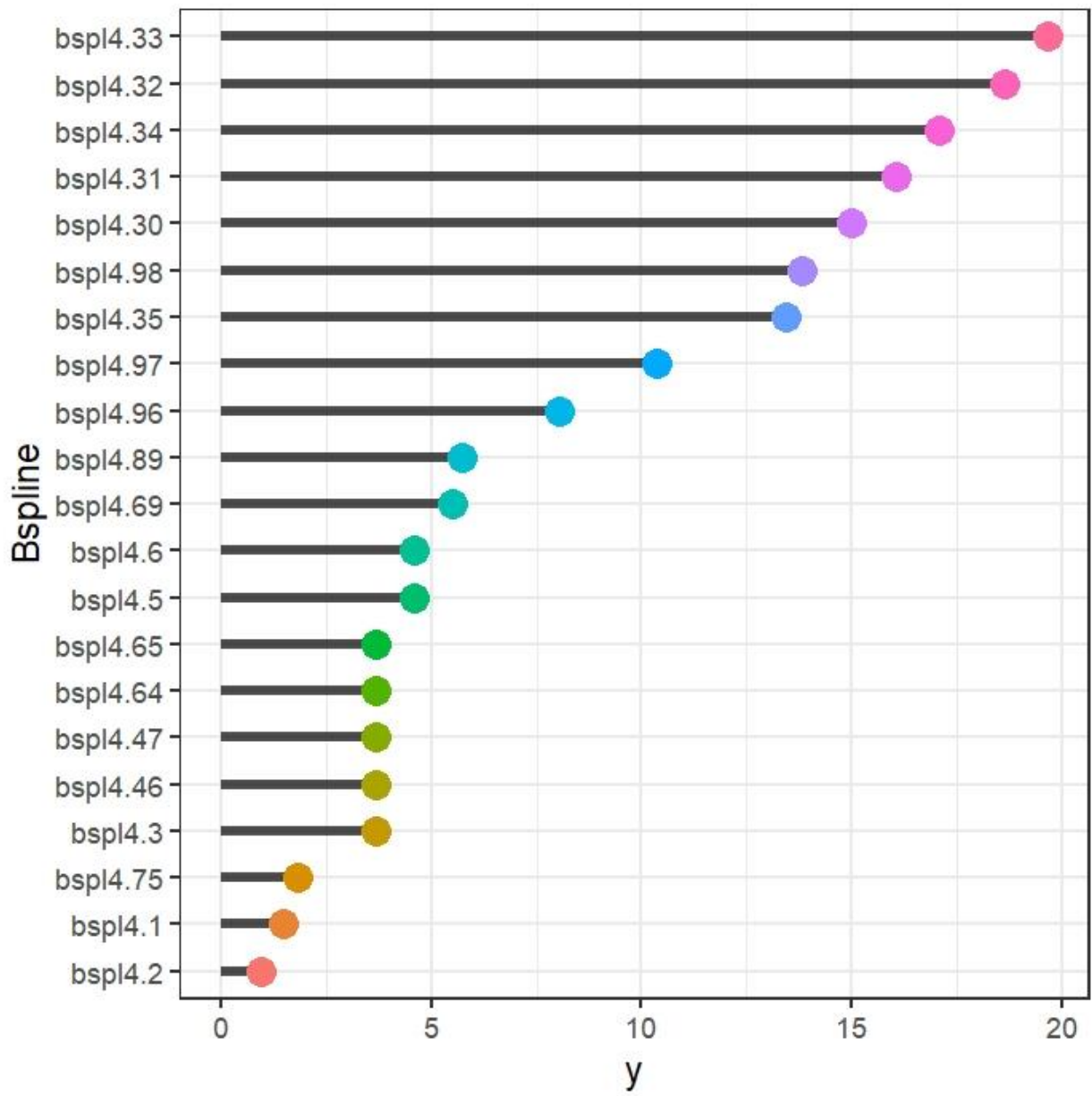
Figure 6. FPCs importance in a single FCT based on b-splines.
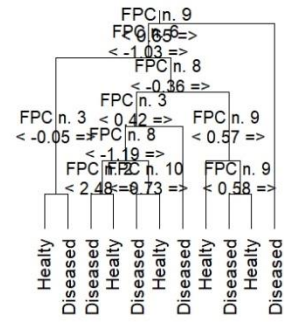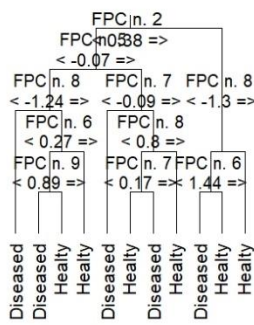
Tree n. 1 Accuracy 0.76
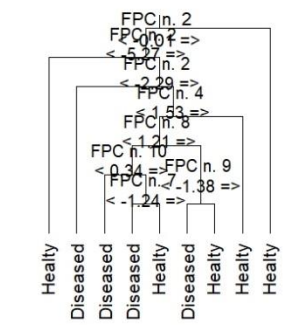
Tree n. 2 Accuracy 0.78
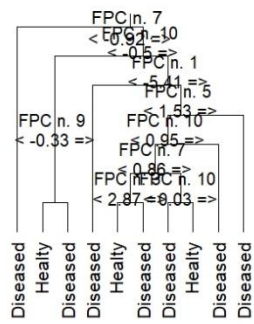
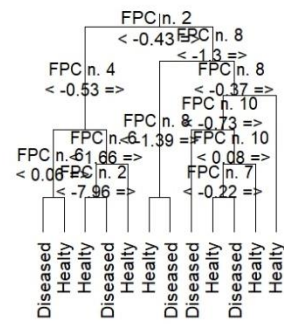**Tree n. 3 Accuracy 0.8**

Tree n. 4 Accuracy 0.78
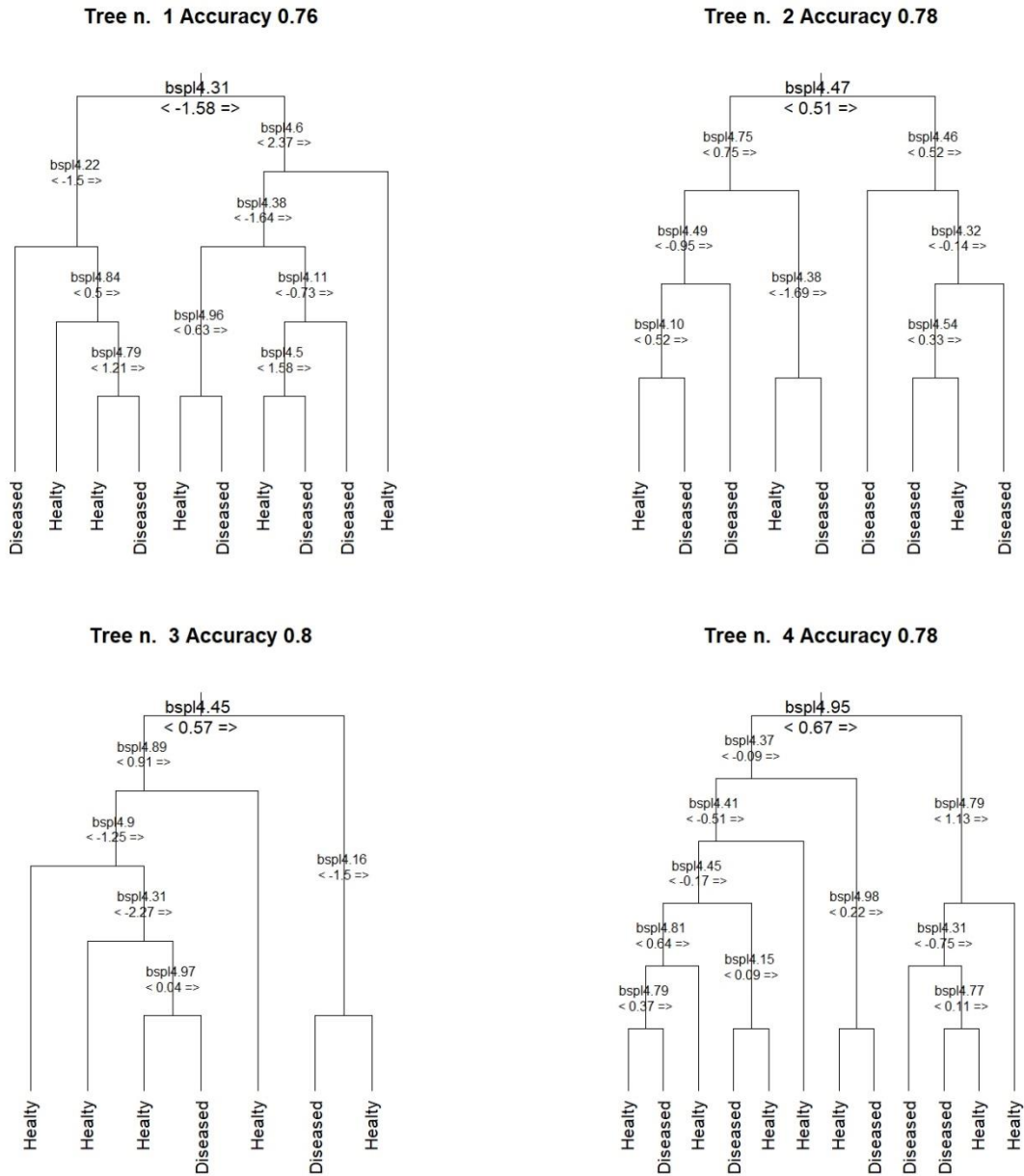
Figure 7. The first four FCTs obtained using the FRF algorithm based on the b-spline decomposition.

	Number of functional principal components involved in the classifier FRF													
N. of Trees	2	3	4	5	6	7	8	9	10	11	12	13	14	15
2	89.38	89.27	89.37	89.87	90.17	90.88	88.18	90.58	87.25	87.23	86.01	87.10	85.23	87.32
3	89.06	88.89	89.66	90.13	90.24	89.09	89.18	88.10	87.25	87.90	86.02	87.57	90.56	89.56
4	90.07	89.79	89.67	90.31	87.71	88.92	90.26	90.82	87.79	90.14	86.30	89.95	89.76	89.17
5	90.48	91.15	91.21	91.90	89.84	90.09	91.24	90.71	90.09	88.77	90.91	89.11	85.52	91.50
6	90.34	89.48	92.24	92.22	92.01	89.63	90.75	92.26	90.11	91.40	89.44	89.96	91.13	90.99
7	90.25	90.17	93.49	91.49	91.68	89.05	91.56	91.56	90.48	87.86	88.84	91.02	90.68	87.50
8	90.89	90.98	91.38	90.49	91.38	90.31	90.80	91.60	91.08	92.42	91.58	89.69	90.78	89.88
9	91.36	90.74	92.51	92.28	92.23	93.12	92.76	91.52	90.71	90.28	92.29	91.06	91.24	90.04
10	91.33	91.50	92.31	92.32	91.52	92.91	92.43	92.74	91.18	91.90	91.94	92.37	90.47	90.71
11	90.95	91.18	90.54	91.52	92.37	91.95	92.34	91.38	91.58	91.00	90.93	91.11	91.58	92.77
12	91.15	92.35	90.78	92.40	91.80	93.95	91.77	92.96	91.97	92.14	92.40	91.77	91.78	91.95
13	91.80	91.35	92.20	92.59	93.17	93.55	92.97	93.19	92.57	91.77	93.19	92.00	92.59	92.38
14	92.00	90.98	93.80	94.00	92.99	93.78	94.78	93.20	93.57	95.20	91.60	93.79	92.59	93.78
15	91.60	91.78	92.37	92.37	93.20	93.59	93.37	93.59	91.58	91.97	93.80	93.59	93.80	93.60
16	91.58	91.40	93.20	92.99	92.38	93.20	93.39	93.00	93.99	93.39	93.17	93.40	93.20	93.20
17	91.80	91.20	92.60	93.80	92.80	93.40	92.79	92.80	93.00	93.79	92.20	93.20	92.40	92.20
18	92.00	93.20	92.80	94.60	92.60	93.80	93.00	93.60	92.60	93.20	92.80	93.00	92.60	92.80
19	93.00	91.18	93.40	93.20	92.60	93.40	93.40	93.40	92.40	94.00	94.00	93.60	94.60	93.60
20	92.20	91.58	92.80	94.20	91.98	93.80	93.60	93.20	93.40	93.40	92.59	93.80	93.60	92.18
21	92.80	92.20	93.60	92.20	93.60	93.60	93.20	93.40	94.00	94.60	94.00	93.20	93.00	93.60
22	92.20	92.40	92.80	93.40	94.20	94.00	93.20	93.80	94.60	93.80	93.20	93.40	94.00	94.00
23	91.60	92.60	93.80	94.20	93.40	93.00	93.40	94.20	92.60	94.20	93.40	93.80	93.40	94.00
24	92.00	92.80	93.40	93.80	93.20	93.40	93.00	93.00	93.60	93.60	93.60	93.40	93.60	93.40
25	91.80	92.40	93.20	94.40	93.80	93.20	93.80	94.20	93.40	93.20	94.20	93.40	94.00	93.60
26	92.60	91.80	94.00	93.80	93.40	93.60	93.60	93.60	94.00	93.40	93.80	94.40	93.60	93.60
27	92.20	92.40	93.00	93.00	94.00	93.80	93.60	94.20	94.00	94.80	93.60	93.20	92.60	94.40
28	92.00	91.80	93.60	94.00	94.20	94.40	93.60	93.80	94.00	93.80	94.20	94.00	94.00	93.20
29	92.60	92.60	93.60	93.40	94.20	93.20	94.00	94.00	93.60	94.40	93.40	94.00	94.20	94.20
30	91.60	93.00	93.80	93.80	93.80	94.40	93.60	93.40	93.60	94.00	94.20	93.60	94.80	94.20
31	92.00	92.60	93.00	93.40	94.20	94.40	93.40	94.20	93.60	94.00	93.60	94.80	93.80	93.60
32	92.00	91.80	92.80	93.40	94.00	94.20	94.00	93.40	94.60	94.40	95.00	93.60	93.20	93.60
33	92.00	93.20	93.60	94.00	94.20	93.20	93.20	94.20	93.60	93.60	94.00	94.40	93.80	94.00
34	91.60	92.80	94.00	93.80	93.40	93.60	93.00	94.00	94.60	94.00	93.80	93.40	93.20	93.80
35	92.20	92.60	93.40	94.00	94.20	93.80	94.20	94.00	93.80	93.20	94.00	94.00	93.80	93.80
36	92.00	92.80	93.20	93.60	94.00	94.40	93.40	94.00	94.60	94.40	94.00	94.60	93.60	94.00
37	92.00	92.40	94.00	93.20	93.80	94.20	94.40	93.40	93.80	94.40	93.80	93.40	93.60	94.20
38	92.00	92.60	93.80	94.00	93.40	93.80	93.40	94.60	94.20	93.80	94.20	94.40	94.40	94.60
39	92.20	92.40	93.80	93.20	92.80	93.80	94.60	94.20	94.00	94.40	94.00	93.20	94.60	93.40
40	92.40	92.60	93.40	93.40	94.40	93.80	94.20	94.20	93.20	93.20	93.60	94.00	93.60	94.20
45	92.40	93.00	93.80	94.20	94.40	93.60	94.00	93.60	94.00	94.00	94.20	94.40	93.80	94.00
50	92.40	93.20	93.80	93.80	94.20	94.00	94.00	93.80	93.40	94.40	94.40	94.20	93.80	94.40

TABLE 1 Accuracy for the ECG5000 dataset using the training set: Accuracy of the FRF classifier according to different sizes of the forest and different numbers of FPCs. The best accuracy on the training set is 97.28% with 4 FPCs and 20 trees.

N. of Trees	2	3	4	5	6	7	8	9	10	11	12	13	14	15
					Number of functional principal components involved in the classifier FRF									
2	89.22	90.07	90.27	88.76	88.64	88.98	89.47	88.73	89.11	88.67	87.58	90.00	88.36	89.56
3	90.29	91.42	91.22	90.96	91.73	91.16	91.58	91.71	90.51	90.93	90.07	89.96	91.40	91.76
4	91.24	90.13	91.56	91.64	92.13	91.40	91.91	91.18	91.07	91.53	91.89	90.82	91.60	90.38
5	91.56	91.22	91.96	92.38	92.49	92.36	92.16	92.00	92.42	92.27	92.11	91.18	91.51	91.89
6	90.80	91.27	91.69	92.07	91.89	92.44	91.84	92.73	91.82	92.40	92.89	92.42	91.53	92.00
7	91.02	91.53	92.67	92.29	92.91	92.73	92.89	91.69	91.82	91.42	92.18	92.64	91.80	92.20
8	90.91	92.07	92.56	92.80	92.58	92.38	92.51	91.93	92.78	92.58	92.22	92.64	93.02	92.53
9	91.87	91.58	92.49	92.67	93.00	93.02	91.98	92.80	92.29	91.78	92.82	92.47	92.76	92.51
10	91.73	92.13	92.42	92.42	92.82	93.07	92.64	92.42	92.33	92.60	92.38	92.51	92.82	92.56
11	92.04	91.82	92.33	92.91	92.56	93.24	92.29	92.62	93.02	92.89	92.38	93.04	92.87	92.96
12	91.67	91.47	91.87	92.71	92.98	92.76	92.93	93.40	93.11	93.20	92.60	92.67	92.04	92.89
13	91.53	92.22	92.53	92.82	92.84	93.16	93.13	92.76	92.29	92.91	92.98	92.78	93.00	92.64
14	91.71	91.93	92.51	93.07	92.91	92.62	92.98	92.89	92.98	93.11	92.91	92.16	92.93	92.76
15	90.96	91.87	92.47	92.71	92.84	92.89	93.02	92.76	93.11	92.87	93.49	92.78	93.47	93.36
16	92.09	92.16	92.07	92.76	92.87	92.71	92.87	92.89	92.13	92.91	93.00	92.02	92.91	93.00
17	91.62	91.93	92.62	92.33	92.91	93.13	93.13	93.22	92.89	92.89	93.16	92.42	93.13	93.09
18	91.67	92.11	92.51	92.80	93.07	93.24	93.16	93.20	92.67	93.16	92.96	93.04	92.60	93.09
19	91.87	92.44	92.40	92.91	92.78	93.18	92.96	93.04	92.87	92.91	93.29	93.36	92.69	92.47
20	91.76	91.96	92.93	92.73	93.11	93.02	93.04	93.09	92.71	92.78	93.18	92.76	92.89	92.91
21	91.96	92.22	92.56	93.29	92.89	93.38	93.04	93.13	92.73	92.87	92.96	92.96	93.22	92.78
22	91.71	91.89	92.73	92.93	92.93	93.20	92.87	93.04	92.71	93.11	93.27	92.82	93.16	92.93
23	92.04	92.36	92.84	92.73	93.18	93.16	93.24	92.60	92.78	92.84	93.18	92.89	92.98	93.09
24	91.78	91.96	92.96	92.89	93.22	92.71	93.09	93.36	92.96	93.09	92.98	93.00	93.47	93.18
25	91.80	92.13	92.58	93.16	92.96	92.93	93.44	92.87	93.16	93.24	93.42	93.00	93.07	93.29
26	91.78	92.16	92.51	93.00	93.13	93.13	92.78	93.13	93.09	93.24	93.22	93.00	92.84	93.07
27	92.16	91.91	92.60	93.07	93.18	92.89	93.24	93.27	93.27	93.31	93.38	92.36	93.24	92.98
28	92.07	92.20	92.51	93.04	93.18	92.98	93.07	93.27	93.11	93.09	93.07	92.98	93.42	92.96
29	91.87	92.02	92.60	92.91	92.80	93.07	92.80	93.27	92.80	93.31	92.87	93.07	92.91	93.07
30	91.89	92.31	92.42	92.98	92.80	92.96	93.07	93.44	93.00	93.11	93.20	93.13	93.04	92.91
31	92.13	92.20	92.82	93.11	93.18	92.98	93.33	93.00	93.09	93.24	93.00	92.82	93.09	93.22
32	91.76	92.18	92.76	92.80	93.27	93.07	92.93	93.27	93.27	93.18	93.09	93.02	93.20	92.84
33	91.98	92.11	92.96	92.78	93.38	93.00	93.22	92.71	92.78	92.82	93.29	93.16	93.04	93.29
34	92.16	91.82	92.56	92.89	92.87	93.38	92.69	93.18	93.22	92.64	92.98	92.91	93.27	93.18
35	91.87	91.87	93.04	93.00	93.29	92.91	93.07	92.96	93.27	93.07	92.71	93.11	93.51	93.11
36	91.89	92.29	92.49	93.16	93.16	93.20	93.22	93.00	93.27	92.91	93.13	92.98	93.31	92.84
37	91.78	92.00	92.51	93.18	93.18	93.18	92.98	93.20	93.18	93.13	93.11	93.00	93.24	92.93
38	91.73	92.13	93.02	93.04	93.07	92.91	93.16	93.33	93.09	92.69	93.16	92.98	93.22	93.22
39	92.00	92.00	92.73	93.36	93.31	93.20	93.24	93.36	93.18	93.13	93.29	92.93	93.24	93.51
40	92.11	92.07	92.53	93.24	93.64	93.22	93.16	93.11	93.02	93.16	93.18	92.76	93.18	93.22
45	91.80	92.11	92.96	92.93	93.11	93.24	92.93	92.96	93.16	92.69	93.29	92.93	93.51	93.16
50	91.82	92.09	92.89	93.07	93.18	93.00	93.29	93.20	93.11	93.04	93.27	93.24	93.02	92.91

TABLE 2 Accuracy for the ECG5000 dataset using the test set: Accuracy of the FRF classifier according to different sizes of the forest and different numbers of FPCs. The best accuracy on the test set is 93.64% with 6 FPCs and 40 trees. The world top accuracy reached for this dataset is still 94.61% with the COTE algorithm. https://www.timeseriesclassification.com/description.php?Dataset=ECG5000

N. of Trees	Accuracy on Training Set	Accuracy on Test Set
2	89.25	89.49
3	88.35	91.87
4	88.32	92.24
5	91.84	93.18
6	91.54	92.20
7	93.21	93.29
8	90.89	92.82
9	91.43	93.22
10	92.77	93.00
11	92.54	93.29
12	92.20	92.80
13	92.99	93.27
14	92.38	93.38
15	94.40	93.38
16	93.19	93.24
17	93.00	93.53
18	94.00	93.31
19	93.60	93.49
20	93.80	93.29
21	94.40	93.84
22	94.60	93.62
23	95.40	93.38
24	94.60	93.87
25	94.00	93.60
26	94.20	93.44
27	93.80	93.38
28	94.40	93.67
29	95.20	93.51
30	94.80	93.53
31	94.80	93.49
32	94.80	93.49
33	94.60	93.47
34	94.60	93.67
35	94.40	93.47
36	94.80	93.36
37	93.80	93.69
38	93.60	93.58
39	94.80	93.49
40	95.20	93.62
45	94.60	93.62
50	94.80	93.58

TABLE 3 Accuracy for the ECG5000: Accuracy of the FRF-B-spline classifier according to different sizes of the forest and a fixed number of B-splines. The best accuracy on the training set is 95.20% with 29 trees. The best accuracy on the test set is 93.87% with 24 trees.

N. of NNs in functional KNN	Accuracy on Training Set	Accuracy on Test Set
1	93.60	92.31
3	94.60	93.42
5	94.80	93.68
7	94.80	93.62
9	95.00	93.64
11	93.20	93.44
13	93.00	92.69
15	93.00	92.04
17	93.00	92.09
19	93.00	92.09
21	93.00	92.07
23	93.00	92.02
25	93.00	92.02

TABLE 4 Accuracy for the ECG5000 dataset: Functional classification using the k-nn classifier of the R package fda.usc. The best accuracy on training set is 95.00% with 9 nearest neighbours. The best accuracy on the test set is 93.64% with 9 nearest neighbours.

Functional Depth Measure	Accuracy on the Training Set	Accuracy on the Test Set
	Train	Test
RP	61.60	63.16
mode	93.60	91.87
RT	77.60	73.27
FM	65.20	68.04
RPD	65.20	68.04

TABLE 5 Accuracy for the ECG5000: Functional classification using depth classifiers of the R package fda.usc. The best accuracy on the training set is 93.60% with the depth measure "mode". The best accuracy on the test set is 91.87% with the depth measure "mode". depth.RP computes the Random Projection depth (see Cuevas et al. 2007). depth.mode implements the modal depth (see Cuevas et al 2007). depth.RT implements the Random Tukey depth (see Cuesta–Albertos and Nieto–Reyes 2008). depth.FM computes the integration of an univariate depth along the axis x (see Fraiman and Muniz 2001). It is also known as Integrated Depth. depth.RPD implements a depth measure based on random projections possibly using several derivatives (see Cuevas et al. 2007).