

Multifractional Gaussian Process Based on Self-similarity Modelling for MS Subgroups' Clustering with Fuzzy C-Means

Yeliz Karaca[1(✉)] and Dumitru Baleanu[2,3]

[1] University of Massachusetts Medical School, Worcester, MA 01655, USA
yeliz.karaca@ieee.org
[2] Department of Mathematics, Çankaya University, 1406530 Ankara, Turkey
dumitru@cankaya.edu.tr
[3] Institute of Space Sciences, Magurele, Bucharest, Romania

Abstract. Multifractal analysis is a beneficial way to systematically characterize the heterogeneous nature of both theoretical and experimental patterns of fractal. Multifractal analysis tackles the singularity structure of functions or signals locally and globally. While Hölder exponent at each point provides the local information, the global information is attained by characterization of the statistical or geometrical distribution of Hölder exponents occurring, referred to as multifractal spectrum. This analysis is time-saving while dealing with irregular signals; hence, such analysis is used extensively. Multiple Sclerosis (MS), is an auto-immune disease that is chronic and characterized by the damage to the Central Nervous System (CNS), is a neurological disorder exhibiting dissimilar and irregular attributes varying among patients. In our study, the MS dataset consists of the Expanded Disability Status Scale (EDSS) scores and Magnetic Resonance Imaging (MRI) (taken in different years) of patients diagnosed with MS subgroups (relapsing remitting MS (RRMS), secondary progressive MS (SPMS) and primary progressive MS (PPMS)) while healthy individuals constitute the control group. This study aims to identify similar attributes in homogeneous MS clusters and dissimilar attributes in different MS subgroup clusters. Thus, it has been aimed to demonstrate the applicability and accuracy of the proposed method based on such cluster formation. Within this framework, the approach we propose follows these steps for the classification of the MS dataset. Firstly, Multifractal denoising with Gaussian process is employed for identifying the critical and significant self-similar attributes through the removal of MS dataset noise, by which, mFd_MS dataset is generated. As another step, Fuzzy C-means algorithm is applied to the MS dataset for the classification purposes of both datasets. Based on the experimental results derived within the scheme of the applicable and efficient proposed method, it is shown that mFd_MS dataset yielded a higher accuracy rate since the critical and significant self-similar attributes were identified in the process. This study can provide future direction in different fields such as medicine, natural sciences and engineering as a result

© Springer Nature Switzerland AG 2020
O. Gervasi et al. (Eds.): ICCSA 2020, LNCS 12250, pp. 426–441, 2020.
https://doi.org/10.1007/978-3-030-58802-1_31

of the model proposed and the application of alternative mathematical models. As obtained based on the model, the experimental results of the study confirm the efficiency, reliability and applicability of the proposed method. Thus, it is hoped that the derived results based on the thorough analyses and algorithmic applications will be assisting in terms of guidance for the related studies in the future.

Keywords: Fractional Brownian Motion · Fractional Gaussian process · Hölder regularity · Multifractal analysis · MS · Fuzzy C-means · Classification · Discrete variations · Regularity · Self-similarity

1 Introduction

Multifractional Brownian Motion (mBm) is considered to be one of the stochastic multifractal models employed to analyse and extract dissimilar patterns, images and signals. Fractal Brownian Motion (fBm) provides attention-grabbing models with various related methods for a broad range of phenomena occurring in the natural world. Multifractal analysis tackles the singularity structure of functions or signals both locally and globally. While Hölder exponent at each point provides the local information, the global information is attained by a characterization of the statistical or geometrical distribution of the Hölder exponents occurring, which is referred to as multifractal spectrum. Introduced by Mandelbrot and Van Ness [1], it is a quintessential theoretical model for the Hurst effect [1,2]; and it is regarded as a powerful model in applied mathematics and other related disciplines such as medicine, biology, physics, financial mathematics, and so forth [3–10]. The reason why it is a powerful model for long-range dependent and short-range dependent complex phenomena in practice is due to its capability of detecting and estimating the highly irregular variations of Hölder regularity of generalized multifractional Brownian motion (GMBM) [11]. As a continuous Gaussian process, GMBM extends both the classical mBm and fBm [12]. Additionally, the analyses with these methods are practical and time-saving while dealing with highly irregular signals [13].

Over the last few years, advances in new technologies have provided different methods to characterize and identify the self-similar and complex patterns in natural phenomena and related problems thereof. Accuracy stands out in these processes particularly for critical decision making. Concerning the time-saving aspect and accuracy of the aforementioned methods, Karaca et al. provided a study on Hölder regularity functions (polynomial, periodic (sine), exponential) through the use of Self-Organizing Map (SOM) algorithm for MS clustering [6]. The study by Lahmiri [14] also points out the accuracy aspect of multifractal patterns of electroencephalographic (EEG) signals of patients with epilepsy and healthy individuals. The results of the study demonstrate that the generalized Hurst exponent (GHE) could be employed in an efficient way to make the distinguishing between healthy individuals and epileptic patients. A recent study by

Tafraouti et al. [15] yields an accuracy rate of 96% with their proposed classification approach with the emphasis on fractional Brownian motion (fBm) model that is capable of characterizing natural phenomena, which is crucial for actual clinical practice. The study by David et al. [16] showed the usefulness of the Hurst exponent and fractal dimension for the analysis of EEG signals' dynamics for epileptic patients and healthy individuals. The study of Rohini et al. [17] also found out that Hölder exponent, tangent slope as well as maximum Hölder exponent proved as the most significant methods for differentiating purposes, which is critical concerning the early stage diagnosis of Alzheimer's. Adapted to complex images, signals and other patterns, multifractal denoising methods are significant to extract irregular and hidden elements in complex systems, providing certain enhancements in the noisy data observed. When studies that address such phenomena in the real world are examined, Karaca et al.'s [18] study performed the application of Diffusion Limited Aggregation (DLA), as one of the multifractal denoising methods, on the patients' MRI images for identifying the self-similar and homogenous pixels. The classification for MS subgroups was conducted by ANN algorithms (CFBP and FFBP); and the critical significance of DLA for MS classification was demonstrated through that study [18].

Being an autoimmune neurological disease, Multiple sclerosis (MS) is characterized by a frequently progressive degeneration of the central nervous system (CNS). Inflammatory demyelination occurs in MS cases, which damages the axons and the neurocytes of those axons [19–21]. The onset of the disease is frequently seen in young adults and its prevalence ranges from 2 to 200 in 100,000 depending on the geographical attributes. It was introduced initially by Jean Martin Charcot in 1868 [22,23]. The present study constitutes three different MS subgroups: RRMS, SPMS and PPMS [9,24,25]. Relapsing Remitting MS, the most common course in the disease, is accompanied by recurrent attacks with neurological deficits in various parts of the nervous system. They either resolve completely or nearly completely in a short time while leaving minor deficit. The second subgroup of MS is SPMS which follows an initial course of relapsing-remitting. Most patients in this group experience a progressive deterioration of neurological functions along with the accumulation of disability in time [9,24,25]. Finally, PPMS is a subgroup which is characterized by accumulation of disability, with worsening neurological functions, starting from the symptoms' occurrence [9,24,25]. There is no early relapse or remission in this subgroup.

Consisting of millions of cells, brain is an inherently complex system itself with its intricate dynamics. Unravelling the complex structure of brain and neural behaviour is critical for diagnostic processes and treatment success. At the same time, it is a challenge to understand how human diseases stem from internal neural irregularities [26]. Advances in artificial intelligence have started to address the challenges in medical settings with the functional computer programs developed which simulate expert human reasoning. Parallel to these developments, algorithms are extensively employed for clustering purposes. Fuzzy C-means (FCM) clustering algorithm is employed accordingly, and modified to do applications on the directional data with a number of advantages. The study

by Das and Das [27] proves the advantage of FCM, proposing a fast and accurate segmentation approach for mammographic images with respect to cancer disease, with the use of the kernel based FCM. Their approach yielded the resolution of imprecise and uncertain characteristics in mammograms [27]. William et al.'s study on cervical cancer classification used an enhanced Fuzzy C-means algorithm and their results demonstrate that the method proposed was capable of outperforming many of the other current algorithms [28]. The study of Sheela and Suganthi is concerned with automatic brain tumor segmentation in MRI with the integration of Fuzzy C-means and Greedy Snake Model optimization. The results show that the method used outperformed the conventional segmentation methods concerning brain tumor [29]. Karaca et al. [7] worked on stroke clustering with the application of FCM and K-means algorithms based on multifractal Bayesian denoising and the results demonstrate the higher accuracy level derived by FCM. In the literature, many studies demonstrate accurate classification results for MS and subgroups thereof. Yet, there seems to be a shortage as regards studies that make use of integrated methods.

In this study, we worked on the MS dataset (139×228), based on the MRI and EDSS data belonging to a total of 139 individuals, 120 of whom were diagnosed with MS (76 RRMS, 38 SPMS, 6 PPMS) and 19 people are healthy individuals, constituting the control group. The first contribution of this study is that the dataset is more comprehensive, which is one of the novel aspects of this paper. In addition, even though earlier works [18–25] have been done on various kinds of analyses concerning MS dataset, no study has been reported thus far which relates attributes (MRI images and EDSS scores) by the use of Multifractal denoising method (L2-norm) with Fuzzy C-means algorithm applied for the purpose of clustering. Accordingly, the principal objective of this study is to identify the similar attributes in homogeneous MS clusters and the dissimilar attributes in the different clusters of MS subgroups. Therefore, the purpose has been to demonstrate the accuracy and applicability of the proposed method which is based on such clustering. For this aim, the approach we have proposed is made up of the following steps for the classification of the MS dataset: (i) Multifractal denoising with Gaussian process, one of the Multifractal denoising method, is used in the identification of critical and significant self-similar attributes by removing the noise of the MS dataset (139 × 228). As a result, mFd_MS dataset (139 × 228) was generated. (ii) Fuzzy C-means algorithm was applied to the MS dataset and the mFd_MS dataset for the clustering purpose of MS subgroups. The experimental results through the proposed method demonstrate that mFd_MS dataset yielded higher accuracy rates compared to MS dataset since significant, self-similar and regular attributes have been characterized. This shows that characterizing the significant attributes plays a critical role in MS subgroup clustering by multifractal denoising method. Comparing our study with the aforementioned studies [18–25] it is seen that the work has a broad and comparative nature, this is because the mFd_MS dataset as obtained from Multifractal denoising with Gaussian process has been the case in point with the Fuzzy C-means algorithm application for the first time in literature.

The paper is organized as follows: Sect. 2 deals with Materials with the Patient Details subsection; and Methods with the subsections of Fractional Brownian Motion and Extensions, Hölder Regularity Analysis, Multifractal Denoising method (L2-norm) and Fuzzy C-means algorithm. Section 3 presents the Experimental Results, to conclude, Sect. 4 provides the Conclusion and Discussion.

2 Materials and Methods

2.1 The Details of the Patient

In our study, patients (120) diagnosed with a clinical definite MS subgroup (RRMS, SPMS, PPMS) based on McDonald criteria [30] and control group with healthy individuals (19 persons) were followed at Hacettepe University, Faculty of Medicine Neurology and Radiology. The individuals' MRI [6,9,18,25,31] images and their EDSS scores [32] based on the years their MRI images were included in the MS dataset. The individuals are aged 18–65 years (see Table 1). The study has been ethically approved by Hacettepe University, Faculty of Medicine Neurology and Radiology and Hacettepe University Ethics Commission.

Table 1. Number and ages of individuals

Gender/Age	18–25	26–30	31–35	36–40	41–45	46–65
Female	10	22	17	16	9	14
Male	6	11	4	6	4	1

The MS dataset (139×228) is made up of the number of lesion diameters and EDSS scores obtained from the MRI images that belong to 120 MS patients (diagnosed with the subgroups of 76 RRMS, 38 SPMS, 6 PPMS) as well as 19 individuals who are healthy, making up the control group.

2.2 Methods

The MS dataset in our study is made up of the Expanded Disability Status Scale (EDSS) scores and Magnetic Resonance Imaging (MRI) (taken in different years) of patients diagnosed who were diagnosed with one of the MS subgroups, which are RRMS, SPMS and PPMS, while healthy individuals constitute the control group. Through this study, mainly two contributions have been provided. Initially, we proposed the use of Multifractal Denoising method (L2-norm) for identifying the critical and significant self-similar attributes. Another contribution is the application of Fuzzy C-means (FCM) algorithm for the clustering goal. As a result of this process, the similar attributes in homogeneous MS clusters and the dissimilar attributes in the different clusters of MS subgroups were classified. Our method is reliant on the steps stated as follows:

(i) Multifractal denoising with Gaussian process (L2-norm) was employed for identifying the self-similar and significant attributes. Consequently, mFd_MS dataset (139×228) was generated.
(ii) Fuzzy C-means algorithm was applied to the two datasets, which are MS dataset and the mFd_MS dataset for the clustering purpose of the MS subgroups.
(iii) Comparisons of clustering accuracy rates were performed for both datasets (MS dataset and mFd_MS dataset).

The experimental results through the proposed method illustrate that mFd_MS dataset yielded higher accuracy rates compared to MS dataset since significant, self-similar and regular attributes have been characterized. This shows that characterizing the significant attributes plays a critical role in MS subgroup clustering by Multifractal denoising with Gaussian process (L2-norm).

All the computations and related analyses were obtained by the Matlab R2019b [33] and FracLab [34].

Fractional Brownian Motion and Extensions. It is possible to model various phenomena that occur naturally in an effective way through the use of self-similar processes for which observations distant in terms of time or space have strong correlations. This situation indicates that a long-range dependence exists. Therefore, there have been successful uses of self-similar processes to model data that shows long-range dependence emerging in wide-ranging domains, which include but are not limited to medicine ([5–9,18,35]), biology ([36,37]) and economics ([5,38]). Long-memory's empirical presence of in these kinds of series is observed in a given local version of a power spectrum, acting as $|\lambda|^{1-2H}$, as $\lambda \to 0$, in which $H \in]1/2, 1[$ refers to the long-memory parameter (see [39]) for a comprehensive monograph of self-similar and long-memory processes with a focus on their statistical and historical sides. There are simple models displaying a long-range dependence, so among them, the fractional Brownian motion (fBm) can be taken into consideration. It was introduced by Mandelbrot et al. in 1968 [1]. The process being null at the origin is defined for real $t \geq 0$ through stochastic integral (1).

$$\begin{aligned} &B_{H,C}(t) = CV_H^{1/2} \int_R f_t(s) dB(s) \ with \\ &f_t(s) = \tfrac{1}{\Gamma(H+1/2)} \{ |t-s|^{H-\frac{1}{2}} 1|_{]-\infty,t]}(s) - |s|^{H-1/2} 1|_{]-\infty,0]}(s) \}, \end{aligned} \tag{1}$$

with $B_{H,C}(0) = 0$ and $V_H = \Gamma(2H+1) \sin(\pi H)$; Γ signifies the Gamma function, while B denotes a Brownian motion that is standard. The fractional Brownian motion of index $H (0 < H < 1)$ and scale parameter C, which is Gaussian process $\{B_{H,C}(t), t \geq 0\}$, based on which the mean is 0, along with self-similar and stationary increases as follows (2) [40]:

$$E\left(\{B_{H,C}(t) - B_{H,C}(s)\}^2 \right) = C^2 |t-s|^{2H} \quad \forall s, t \in R^+. \tag{2}$$

The H index characterizes self-similar aspect of the relevant process [40]. For the estimation of H, variance, maximum likelihood methods (Whittle estimator)

as well as covariance based methods (log-periodogram, rescaled range (R/S)) are among the ones that are used mostly [39]. These methods were aimed both at the analysis of fractional Brownian motion and more extensive clusters related to stochastic processes, which is to say those of Gaussian processes which exhibit locally self-similar aspects initially.

It is considered that a zero-mean Gaussian process $\{X(t), t \geq 0\}$, that has stationary increments is self-similar at 0 locally, which is expressed by locally self-similar Gaussian Process. The semi-variance function $v(t)$ is stated by (3) [39,40]:

$$v(t) = \frac{1}{2}E((X(s+t) - X(s))^2), \tag{3}$$

$t \to 0=$ fulfills the following characteristic (4):

$$v^{2D}(t) = v^{2D}(0) + (-1)^D C|t|^{2H} + o(|t|^{2H}), \tag{4}$$

Based on (2), $0 < H < 1$, D expresses the larger integer in a way that x is 2D-times differentiable. Within the Gaussian framework, the local self-similarity, which is at 0, is equivalent to (4).

Considering $\{V^a(t), t \geq 0\}$ to be the process which is derived by filtering $\{X(t), t \geq 0\}$ with α filter that has the length $l+1$ and order $p \geq$ Based on this $\sum_{q=0}^{l} a_q q^r = 0$ for $r = 0, ..., p-1$. Being observed at times $\left\{0, \frac{1}{N}, ... \frac{N-1}{N}\right\}$ $\{X(t), t \geq 0\}$ is seen and variations of F of $\{X(t), t \geq 0\}$ and the F variations of X are expressed as follows (5) [39,40]:

$$V_N(F, a) = \frac{1}{N-l} \sum_{i=l}^{N-1} F\left(V^a\left(\frac{i}{N}\right)\right). \tag{5}$$

Numerous estimation procedures employed recently are based on these kinds of F-variations with a specific choice of (6) [39,40]:

$$F(t) = H^k(t) \frac{1}{E\left(|V^a(0)|^k\right)} |t|^k - 1. \tag{6}$$

Hölder Regularity Analysis. The Fractional Brownian motion with Gaussian process $B_H(t)$ produces zero mean. The covariance function is seen as in (7) ([7–9,41]).

$$\text{cov}\left\{B_H(s), B_H(t)\right\} = \frac{1}{2}\left\{|s|^{2H} + |t|^{2H} - |s-t|^{2H}\right\} \tag{7}$$

H Hurst exponent, $H = \frac{a}{2}$ and $a \in (0,1)$ [20]. Fractional Brownian Motion is stated based on (8) with a stochastic integral (see the following references for further details and fomula steps: [3,8,42–44]).

For the most critical components of the irregular and singular signals regarding 2-D analysis, it should be noted that the Hölder exponent supplies required information. For this reason, the Hölder base measures the variations in characterizing

of the multiple areas that belong to the signal processor. The Hölder exponent, or singularity exponent, characterizes the singularity force at $t = t(0)$ [8,9] to attain a positive measurement which is expressed by a signal $X(t)$ [41–44].

Multifractal Analysis. Multifractals' measurement refers preliminarily to the measuring of a statistic distribution and outcomes based on such measurement provide beneficial information although the essential structure does not display self-affine as well as self-similar behaviours [45,46]. A signal's multifractal analysis comprises the measurement of its regularity at every sample point, the grouping of the points that have the same irregularity as well as the estimating of the fractal dimension, namely the Hausdorff dimension each iso-regularity set subsequently [46].

Multifractal analysis is performed on statistical basis through the analysis of a large deviation, which refers to a stochastic procedure. Here, it is considered that any generality that $T = [0, 1]$ remains. $N_n^{\in}(\alpha)\#\{s : \alpha - \varepsilon \leq \alpha_n^s \leq \alpha + \varepsilon\}$ in which α_n^s s the coarse-grained Hölder exponent corresponding to a dyadic interval $I_n^s = [s2^{-n}, (s+1)2^{-n}]$ and this is denoted according to (8) [46]:

$$\alpha_n^s = \frac{\log |Y_n^s|}{-\log n} \tag{8}$$

Y_n^s is observed to be a number which calculates the variation of X in the interval of I_n^s. $Y_n^s := X((s+1)2^{-n}) - X(s2^{-n})$ yields a fundamental level of analytical calculations. It is also possible to derive it in different ways by taking as the wavelet coefficient $x_{n,s}$. For data displaying a large deviation $f_g(\alpha)$ is expressed based on (9) [47,48].

$$f_g(\alpha) = \lim_{\varepsilon \to 0} \lim_{n \to \infty} \sup \frac{\log N_n^{\varepsilon}(\alpha)}{\log n} \tag{9}$$

It seems obviously evident that no matter what the choice of Y_n^s, f_g shows a range in $\Re^+ \cup \{-\infty\}$ every time [47,48].

Accordingly, multifractal processes have been expanded on by [49,50]. In particular, when the noise characteristics display complex properties, multifractal approach is employed.

Multifractal Denoising Method (L2-norm). Regularization is considered to be an important method in computational processes in order to eliminate the overfitting problem, avoiding the coefficients to fit so that is flawless for overfitting [47,48].

During the application of a relevant computational process, it is possible to experience the unknown option. At that point, L2-norm generates non-sparse coefficients. L2 regularization on least squares is expressed as in (10) [51,52].

$$w^* = \arg\min_w \left(\sum_j \left(t\left(x_j\right) - \sum_i w_i h_i\left(x_j\right) \right) \right)^2 + \lambda \sum_{i=1}^{k} w_i^2 \tag{10}$$

The least squares error (LSE) is another term for the L2 norm loss function. While (y_i) refers to the total number of square of the differences belonging to the target value, $f(x_i)$ expresses the estimated values, shown in (12). They are kept to the minimum value [51,52].

$$S = \sum_{i=1}^{n} \left(y_i - f\left(x_i\right)\right)^2 \tag{11}$$

L2 regularization is considered to be efficient in terms of computational processes owing to their analytical solutions. Accordingly, in our study, Multifractal denoising with Gaussian process (L2-norm) is applied to the MS dataset (139×228). As a consequence of regularity-based enhancement, mFd_MS dataset (139×228) has been generated. Subsequently, Fuzzy C-means algorithm was applied to the two datasets, namely MS dataset and mFd_MS dataset to attain an accurate and efficient clustering for the classification of subgroups of MS (RRMS, SPMS and PPMS) and the individuals who are healthy.

Fuzzy C-Means Algorithm. It is a well-known fact that conventional clustering methods work better when homogeneous attributes are handled, while their performance declines when the methods are implemented on inhomogeneous attributes since they are sensitive to noise. At such a stage, Fuzzy C-means, as an unsupervised clustering technique, is employed for medical analyses as well as image processing.

Fuzzy C-means algorithm assigns information to each section by the use of fuzzy memberships [53–56]. If $X = (x_1, x_2, ..., x_n)$, n d-dimensional information shows information needed to divide into c clusters, $\hat{x}_i$ expresses the features data. The indication of the steps of the optimization algorithm are as follows (see Fig. 1) [7,53–56]:

The principal aim of our study is to identify the similar attributes in homogeneous MS clusters and the dissimilar attributes in the different clusters of MS subgroups. Fuzzy C-means algorithm was applied to both the MS dataset $X = (x_1, x_2, ..., x_{139\times228})$ and the mFd_MS dataset $\hat{X} = (\hat{x}_1, \hat{x}_2, ..., \hat{x}_{139\times228})$ in line with the steps mentioned above for the clustering of the MS subgroups (RRMS, SPMS, and PPMS) and the control group of healthy individuals.

3 Experimental Results

The MS dataset of our study includes Expanded Disability Status Scale (EDSS) scores and Magnetic Resonance Imaging (MRI) (taken in different years) of patients diagnosed with one MS subgroup (RRMS, SPMS and PPMS) and the healthy individuals are involved in the control group. Two contributions have been provided by this study. Firstly, we proposed the use of Multifractal Denoising method (L2-norm) to identify the critical and significant self-similar attributes. Secondly, the application of Fuzzy C-means (FCM) algorithm was

Algorithm 1: FCM algorithm

Input: Dataset with n number of samples $X = (x_1, x_2, \ldots, x_n)$

Method:

1. At the beginning, the construction of the U membership matrix is made with random values. Afterwards, we determine the fuzzy weight (m), the threshold value on the condition stopping (ε) and the number of clusters (c), which shown in (13).

$$v_j = \frac{\sum_{i=1}^{N} \mu_{ij}^{m} x_i}{\sum_{i=1}^{N} \mu_{ij}^{m}}, 1 \le j \le c \tag{12}$$

2. The determination of the cluster centres (v_j) is based on Eq. (14),

$$X_{ijG}^2 = \| x_i - v_j \|_G^2 = (x_i - v_j) G (x_i - v_j)^T \tag{13}$$

3. We calculate the distance between cluster centres based on Eq. (14). In this way, each sample is calculated.

4. The updating of the membership matrix is done based on Eq. (15).

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{c} \left(D_{ijG} / D_{ikG} \right)^{2/(m-1)}}, 1 \le i \le N, 1 \le j \le c \tag{14}$$

5. If it is observed that the difference between the membership matrix in the former step and the updated membership matrix is more than the specified threshold value, we repeat the same calculations stated in Step 2 $\| U^{(t+1)} - U^{(t)} \| < \varepsilon$

Output:

Clustering information concerning the n number of samples in the whole dataset is yielded.

Fig. 1. FCM algorithm.

performed for the clustering purpose. As a result, the classification of the similar attributes in homogeneous MS clusters and the dissimilar attributes in the different clusters of MS subgroups was done. To attain the proposed contributions, with the integration of the multifractal approach and Fuzzy C-means application, our method in this study is reliant on the subsequent steps:

(i) Multifractal denoising with Gaussian process (L2-norm) was employed for identifying the critical and significant self-similar attributes. Consequently, mFd_MS dataset (139×228) was generated.

(ii) In order to classify the similar attributes in homogeneous MS clusters and the dissimilar attributes in the different clusters of MS subgroups, FCM algorithm was applied to the MS dataset and the mFd_MS dataset (see Fig. 2).

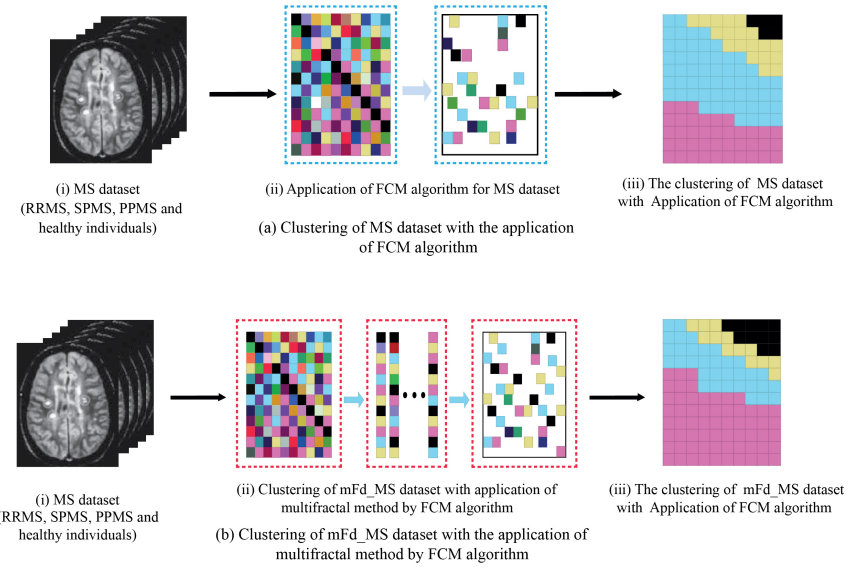

Fig. 2. MS dataset clustering with multifractal method application by FCM algorithm.

As the first step, in order to classify the similar attributes in homogeneous MS clusters and the dissimilar attributes in the different clusters of MS subgroups, Fuzzy C-means algorithm was applied to the MS dataset (Fig. 2(a)). Secondly, Multifractal Denoising with Gaussian process (L2-norm) was applied to the MS dataset for the identification of the critical and significant self-similar attributes; and the mFd_MS dataset (139×228) was generated (Fig. 2(b))). As the next step, Fuzzy C-means was applied to the mFd_MS dataset (139×228).

(iii) Accuracy rates as a consequence of clustering based on the proposed method were compared for both datasets (MS dataset and mFd_MS dataset) (see Table 3 for the results).

The results of characteristic parameters for Fuzzy C-means algorithm are provided in Table 2 below:

Table 2. The characteristic parameters distribution of the FCM algorithm

Parameters	The value of the parameters
Exponent for the partition matrix U	2.0
Maximum number of iterations	1000
Minimum amount of improvement	1e−3

Computation results of FCM clustering algorithm for 1000 iterations are shown in Fig. 3(a) for the MS dataset and Fig. 3(b) for the mFd_MS dataset.

The clustering computation is finalized at the 1000th iteration since there is not any change in the membership matrix according to the previous iteration, which can be seen in Fig. 3(a).

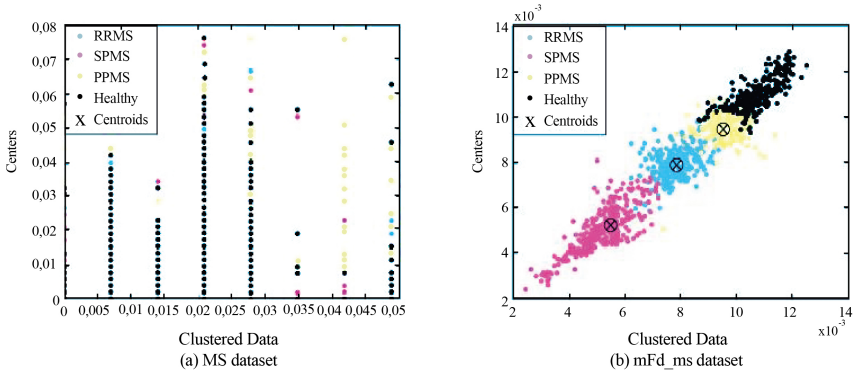

Fig. 3. The results of the clustering analyses of FCM algorithm application for (a) MS dataset (b) mFd_MS dataset.

As can be seen in Fig. 3, the distance between data centres of the MS dataset (139 × 228) as well as of the mFd_MS dataset (139 × 228) is calculated (the distance between the data centre and the MS patient within the feature domain). The results which were calculated are stored in matrix μ_{ij} After the 1000th iteration, the MS dataset objective function vector for the MS dataset [1.10, 2.4, 0.4] and for the mFd_MS dataset objective function vector for MS dataset [0.10, 0.14, 0.32] was obtained.

Regarding the calculation of the μ_{ij} for 1000 iterations (see Fig. 3), it has been shown that the clustering performance result for the mFd_MS dataset is higher than that of the MS dataset.

3.1 The Results Based on the Method Proposed

The method we have proposed by multifractal Denoising with Gaussian process (L2-norm) as applied to the MS dataset enabled identifying the critical and significant self-similar attributes; and thus, with Fuzzy C-means, accuracy rate performance was attained high.

The accuracy rate results of classification with Fuzzy-C means algorithm based on our proposed method are presented in Table 3.

As it can be seen from Table 3, the mFd_MS dataset yielded higher accuracy rates compared to MS dataset as a result of the clustering of MS subgroups by Fuzzy C-means. This shows that the analyses done by a Multifractal Denoising method (L2) enables the identifying of the critical and significant self-similar attributes, which contributes to accurate classification that is critical for diagnosis, decision-making and timely management of diseases in medicine. This also

Table 3. The accuracy rate results of classification with Fuzzy-C means algorithm

MS subgroups	MS dataset (%)	mFd_MS dataset (%)
RRMS	71	88.1
SPMS	71.5	88.2
PPMS	81.2	85.4
Healthy	90.1	96.5

ensures and maintains the high quality of life on the patients' side. For physicians, having an early accurate diagnosis certainly assists them.

4 Conclusion and Discussion

Our study has aimed to employ a multifractal approach for obtaining an accurate diagnostic classification for the MS subgroups while tackling the characterization of critical and significant self-similar attributes in the MS dataset. Accordingly, since multifractal analysis proves to be beneficial way to characterize the heterogeneous nature of fractals by dealing with the singularity structure of functions or signals, we have employed it in our study. The use of multifractal analysis is particularly observed concerning wide-ranging phenomena in the natural world. This kind of analysis saves time and is efficient while dealing with highly irregular signals, which is also the case in medicine. As different from previous works ([10–17,27–29,52–56]), this study has used Fuzzy C-means for MS dataset for the first time in line with the proposed multifractal approach through which it has been possible to identify the attributes that are critical, regular and self-similar. The experimental results for the MS_dataset and mFd_MS dataset demonstrate that the mFd_MS dataset yielded higher accuracy rates through the multifractal denoising method (L2) with Fuzzy C-means algorithm. Consequently, the approach we have proposed contributes to accurate diagnostic classification which plays a highly decisive role for the diagnostics and timely management of diseases in medicine. This asset also secures and maintains the high quality of life on the patients' side. Regarding future direction, it can be stated that the multifractal approach can be employed in natural sciences as well as engineering apart from medicine.

References

1. Mandelbrot, B.B., Van Ness, J.W.: Fractional Brownian motions, fractional noises and applications. SIAM Rev. **10**(4), 422–437 (1968)
2. Setty, V.A., Sharma, A.S.: Characterizing detrended fluctuation analysis of multifractional Brownian motion. Physica A: Stat. Mech. Appl. **419**, 698–706 (2015)
3. Dumitru, B., Kai, D., Enrico, S.: Fractional Calculus: Models and Numerical Methods, vol. 3. World Scientific (2012)

4. Karaca, Y.: Wavelet-based multifractal spectrum estimation in hepatitis virus classification models by using artificial neural network approach. In: Shapshak, P., et al. (eds.) Global Virology III: Virology in the 21st Century, pp. 73–96. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-29022-1_4
5. Karaca, Y., Zhang, Y.D., Muhammad, K.: Characterizing complexity and self-similarity based on fractal and entropy analyses for stock market forecast modelling. Expert Syst. Appl. **144**, 113098 (2020)
6. Karaca, Y., Cattani, C.: Computational Methods for Data Analysis. Walter de Gruyter GmbH, Berlin (2018). (978–3110496352)
7. Karaca, Y., Cattani, C., Moonis, M., Bayrak, Ş.: Stroke subtype clustering by multifractal Bayesian denoising with Fuzzy C means and K-Means algorithms. Complexity (2018)
8. Karaca, Y., Cattani, C.: A comparison of two Hölder regularity functions to forecast stock indices by ANN algorithms. In: Misra, S., et al. (eds.) ICCSA 2019. LNCS, vol. 11620, pp. 270–284. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-24296-1_23
9. Machado, J.A.T., Baleanu, D., Luo, A.C.: Nonlinear and Complex Dynamics: Applications in Physical, Biological, and Financial Systems. Springer, Berlin (2011). https://doi.org/10.1007/978-1-4614-0231-2
10. Zhang, Y., et al.: Fractal dimension estimation for developing pathological brain detection system based on Minkowski-Bouligand method. IEEE Access **4**, 5937–5947 (2016)
11. Ayache, A., Véhel, J.L.: Generalized multifractional Brownian motion: definition and preliminary results. In: Dekking, M., Véhel, J.L., Lutton, E., Tricot, C. (eds.) Fractals, pp. 17–32. Springer, London (1999). https://doi.org/10.1007/978-1-4471-0873-3_2
12. Ayache, A., Cohen, S., Véhel, J.L.: The covariance structure of multifractional Brownian motion, with application to long range dependence. In: 2000 IEEE International Conference on Acoustics, Speech, and Signal Processing, Proceedings (Cat. No. 00CH37100), vol. 6, pp. 3810–3813. IEEE (2000)
13. Ayache, A., Véhel, J.L.: On the identification of the pointwise Hölder exponent of the generalized multifractional Brownian motion. Stoch. Processes Appl. **111**(1), 119–156 (2004)
14. Magin, R., Ortigueira, M.D., Podlubny, I., Trujillo, J.: On the fractional signals and systems. Sig. Process. **91**(3), 350–371 (2011)
15. Lahmiri, S.: Generalized Hurst exponent estimates differentiate EEG signals of healthy and epileptic patients. Physica A: Stat. Mech. Appl. **490**, 378–385 (2018)
16. El Hassouni, M., Jennane, R., Tafraouti, A.: Evaluation of fractional Brownian motion synthesis methods using the SVM classifier. Biomed. Sig. Process. Control **49**, 48–56 (2019)
17. David, S.A., Machado, J.A.T., Inácio Jr., C.M.C., Valentim Jr., C.A.: A combined measure to differentiate EEG signals using fractal dimension and MFDFA-Hurst, Commun. Nonlinear Sci. Numer. Simul. **84**, 105170 (2020)
18. Rohini, P., Sundar, S., Ramakrishnan, S.: Differentiation of early mild cognitive impairment in brainstem MR images using multifractal detrended moving average singularity spectral features. Biomed. Sig. Process. Control **57**, 101780 (2020)
19. Karaca, Y., Cattani, C., Karabudak, R.: ANN classification of MS subgroups with diffusion limited aggregation. In: Gervasi, O., et al. (eds.) ICCSA 2018. LNCS, vol. 10961, pp. 121–136. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-95165-2_9

20. Compston, A., Coles, A.: Multiple sclerosis. The Lancet **372**, 1502–1517 (2008)
21. Tallantyre, E.C., et al.: Ultra-high-field imaging distinguishes MS lesions from asymptomatic white matter lesions. Neurology **76**(6), 534–539 (2011)
22. Goodwin, S.J.: Multiple sclerosis: integration of modeling with biology, clinical and imaging measures to provide better monitoring of disease progression and prediction of outcome. Neural Regen. Res. **11**(12), 1900 (2016)
23. Multipl Skleroz Tanı ve Tedavi Kılavuzu 2016 Basım Tarihi: Galenos Yayınevi, İstanbul (2016)
24. Sand, I.K.: Classification, diagnosis, and differential diagnosis of multiple sclerosis. Curr. Opin. Neurol. **28**(3), 193–205 (2015)
25. Waxman, S.: Multiple Sclerosis as a Neuronal Disease. Elsevier, Amsterdam (2005)
26. Karaca, Y., Osman, O., Karabudak, R.: Linear modeling of multiple sclerosis and its subgroubs. System **2**, 7 (2015)
27. Siegelmann, H.T.: Complex systems science and brain dynamics. Front. Comput. Neurosci. **4**, 7 (2010)
28. Das, P., Das, A.: A fast and automated segmentation method for detection of masses using folded kernel based fuzzy C-means clustering algorithm. Appl. Soft Comput. **85**, 105775 (2019)
29. William, W., Ware, A., Basaza-Ejiri, A.H., Obungoloch, J.: Cervical cancer classification from Pap-smears using an enhanced Fuzzy C-mean algorithm. Inform. Med. Unlocked **14**, 23–33 (2019)
30. Sheela, C.J.J., Suganthi, G.: Automatic brain tumor segmentation from MRI using greedy snake model and fuzzy C-means optimization. J. King Saud Univ.-Comput. Inf. Sci. (2019)
31. Thompson, A.J., et al.: Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. Lancet Neurol. **17**, 162–173 (2017)
32. Karaca, Y., Zhang Y.D., Cattani, C., Ayan, U.: The differential diagnosis of multiple sclerosis using convex combination of infinite kernels. CNS Neurol. Disord.-Drug Targets (Formerly Curr. Drug Targets-CNS Neurol. Disord.) **16**(1), 36–43 (2017)
33. Bushnik, T.: Expanded disability status scale. In: Kreutzer, J.S., DeLuca, J., Caplan, B. (eds.) Encyclopedia of Clinical Neuropsychology. Springer, New York (2011). https://doi.org/10.1007/978-0-387-79948-3
34. The MathWorks, MATLAB (R2019b). The MathWorks, Inc., Natick (2019)
35. Vehel, J.L.: FracLab (2000). https://project.inria.fr/fraclab/
36. Kuklinski, W.S., Chandra, K., Ruttimann, U.E., Webber, R.L.: Application of fractal texture analysis to segmentation of dental radiographs SPIE. Med. Imaging III: Image Process. **1092**, 111–117 (1989)
37. Collins, J.J., De Luca, C.J.: Upright, correlated random walks: a statistical-biomechanics approach to the human postural control system. Chaos **5**(1), 57–63 (1994)
38. Baleanu, D., Jajarmi, A., Mohammadi, H., Rezapour, S.: A new study on the mathematical modelling of human liver with Caputo-Fabrizio fractional derivative. Chaos Solitons Fractals **134**, 109705 (2020)
39. Granger, C.W.J.: The typical spectral shape of an economic variable. Econometrica **34**, 150–161 (1966)
40. Beran, J.: Statistics for Long Memory Processes. Chapman and Hall, London (1994)
41. Coeurjolly, J.F.: Estimating the parameters of a fractional Brownian motion by discrete variations of its sample paths. Stat. Infer. Stoch. Process. **4**(2), 199–227 (2001)

42. Garcin, M.: Estimation of time-dependent Hurst exponents with variational smoothing and application to forecasting foreign exchange rates. Physica A: Stat. Mech. Appl. **483**, 462–479 (2017)
43. Misawa, M.: Local Hölder regularity of gradients for evolutional p-Laplacian systems. Annali di Matematica **181**(4), 389–405 (2002)
44. Aizenman, M., Burchard, A.: Hölder regularity and dimension bounds for random curves. Duke Math. J. **99**(3), 419–453 (1999)
45. Franchi, B., Lanconelli. E.: Hölder regularity theorem for a class of linear nonuniformly elliptic operators with measurable coefficients. Annali della Scuola Normale Superiore di Pisa-Classe di Scienze **10**(4), 523–541 (1983)
46. Tarquis, A., Giménez, D., Saa, A., Diaz, M.: Scaling and multiscaling of soil pore systems determined by image analysis. Scal. Methods Soil Phys. **19** (2003)
47. Lutton, E., Grenier, P., Vehel, J.L.: An interactive EA for multifractal Bayesian denoising. In: Rothlauf, F., et al. (eds.) EvoWorkshops 2005. LNCS, vol. 3449, pp. 274–283. Springer, Heidelberg (2005). https://doi.org/10.1007/978-3-540-32003-6_28
48. Pu, Y.F., et al.: Fractional partial differential equation denoising models for texture image. Sci. China Inf. Sci. **57**(7), 1–19 (2014). https://doi.org/10.1007/s11432-014-5112-x
49. Yau, A.C., Tai, X., Ng, M.K.: Compression and denoising using l_0-norm. Comput. Optim. Appl. **50**, 425–444 (2011). https://doi.org/10.1007/s10589-010-9352-4
50. Feder, J.: Fractals (Physics of Solids and Liquids). Plenum, New York (1987)
51. Li, L., Chang, L., Ke, S., Huang, D.: Multifractal analysis and lacunarity analysis: a promising method for the automated assessment of muskmelon (Cucumis melo L.) epidermis netting. Comput. Electron. Agric. **88**, 72–84 (2012)
52. Kalavathy, S., Suresh, R.M.: Analysis of image denoising using wavelet coefficient and adaptive subband thresholding technique. Int. J. Comput. Sci. (IJCSI) **8**, 6 (2011)
53. Karaca, Y., Moonis, M., Zhang, Y.-D.: Multifractal analysis with L2 norm denoising technique: modelling of MS subgroups classification. In: Misra, S., et al. (eds.) ICCSA 2019. LNCS, vol. 11620, pp. 257–269. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-24296-1_22
54. Alruwaili, M., Siddiqi, M.H., Javed, M.A.: A robust clustering algorithm using spatial Fuzzy C-mean for brain MR images. Egypt. Inform. J. (2019)
55. Jianzhong, W., Kong, J., Lu, M., Qi, M., Zhang, B.: A modified FCM algorithm for MRI brain image segmentation using both local and non-local spatial constraints. Comput. Med. Imaging Graph. **32**(8), 685–698 (2008)
56. Ahmed, M.N., Yamany, S.M., Mohamed, N., Farag, A.A., Moriarty, T.: A modified Fuzzy C-mean algorithm for bias field estimation and segmentation of MRI data. IEEE Trans. Med. Imaging **21**(3), 193–199 (2002)