EEG Representation Using Multi-Instance Framework on The Manifold of Symmetric Positive Definite Matrices

Khadijeh Sadatnejad, Mohammad Rahmati, Reza Rostami, Reza Kazemi, Saeed S. Ghidary, Andreas Müller, Fatemeh Alimardani

Abstract—The generalization and robustness of an electroencephalogram (EEG)-based system are crucial requirements in actual practices. To reach these goals, we propose a new EEG representation that provides a more realistic view of brain functionality by applying multi-instance (MI) framework to consider the non-stationarity of the EEG signal. In this representation, the non-stationarity of EEG is considered by describing the signal as a bag of relevant and irrelevant concepts. The concepts are provided by a robust representation of homogeneous segments of EEG signal using spatial covariance matrices. Due to the nonlinear geometry of the space of covariance matrices, we determine the boundaries of the homogeneous segments based on adaptive segmentation of the signal in a Riemannian framework. Each subject is described as a bag of covariance matrices of homogeneous segments and the bag-level discriminative information is used for classification. To evaluate the performance of the proposed approach, we examine it in a cultural neuroscience application for classification Iranian versus Swiss normal subjects to discover if strongly differing cultures can result in distinguishing patterns in brain electrical activity of the subjects. To confirm the effectiveness of the proposed representation, we also evaluate the proposed representation in EEG-based mental disorder diagnosis application for Attention Deficit Hyperactivity Disorder/Bipolar Mood Disorder, Schizophrenia/ normal, and Major Depression Disorder/normal diagnosis applications. Experimental results confirm the superiority of the proposed approach, which is gained due to the robustness of covariance descriptor, the effectiveness of Riemannian geometry, the benefits of considering the inherent non-stationary nature of the brain by applying bag-level discriminative information, and automatic handling the artifacts.

Index Terms— **EEG, covariance matrix, multi-instance learning framework, non-stationarity, Riemannian manifold.**

I. INTRODUCTION

In this paper, we propose a new representation of EEG signal to be used in EEG-based classification systems with the aim of increasing the robustness and generalization of the system. Automatic analysis of EEG signal based on machine learning techniques has been applied to a variety of real-world applications. For instance, analysis of EEG signals is considered as a tool for some basic research in the field of computational neuroscience which evaluate the validity of using EEG databases for diagnosis purposes by investigating the differences in EEG signals of the different nationality by different cultures [58, 59], mental disorder diagnosis [1] [2] [3], brain-computer interfacing [4] [5] [6], and emotion recognition [53].

EEG-based computer-aided diagnosis (CAD) as a quantitative method for diagnosis, which relies on analysis of electrical activity of the brain, can help clinicians to increase the confidence of diagnosis in comparison with classical diagnosis methods which relied on qualitative diagnostic criteria, such as diagnostic and statistical manual of mental disorders [1], used in clinical interview, which can be impressed by physician knowledge and attitude, with a relatively low agreement in different clinicians' decision. Recently, numerous studies have investigated this problem from different points of views. However, the generalization and robustness of those systems are still crucial bottlenecks.

The EEG-based diagnosis systems usually are based on using models or comparison with references which are captured from subjects with different cultures [60]. The influence of culture on EEG [59] may affect the robustness of diagnosis systems and its generalization cross cultures. The existence of similar patterns between normal subjects with the same culture and dissimilarity of the patterns cross cultures strongly influence the validity of quantitative diagnosis based on EEG signals.

The generalization and robustness of the model is an essential challenge for BCI systems [54].

The generalization and robustness of a classification system in different EEG-based application can be considered in each

component of the system, including preprocessing, feature extraction, and classification components. In a general categorization, EEG-based systems, for representing signals, can be divided into two linear and non-linear approaches. Some researchers have shown promising results using linear approaches [2, 3, 4, 5]; however complex, non-linear, and non-stationary nature of EEG signals inevitably leads to insufficiency of linear methods for representing EEG signals [2, 52]. Variety of nonlinear methods such as Higher Order Spectra, different entropies, fractal dimension, correlation dimension, largest Lyapunov exponent, and Hurst exponent have been used for representing EEG signals in different applications, including seizure detection, diagnosis of epilepsy [6, 7, 8, 9, 10], Alzheimer [11], Schizophrenia [12], and depression [13, 14, 15], BCI [55], and other EEG-based applications [56]. The promising results achieved by these methods confirm the suitability of non-linear approaches for representing EEG in EEG-based systems [54, 55, 56].

Recently a new approach for EEG signal analysis based on the Riemannian geometry has been developed [16], which uses the spatial covariance matrix of EEG recording as a non-linear representation of EEG signals. This method has gained promising results in EEG-based applications, especially in brain-computer interfacing (BCI). Other researches also have reported reliability, robustness, and high performance by considering the Riemannian geometry of symmetric positive definite (SPD) matrices in comparison to the usual Euclidean approaches [17]. These findings advocate us to use the spatial covariance matrix of EEG segments for representing the recordings.

The analysis of covariance matrices or in general terms, statistical data processing approaches for EEG signal analysis is based on the assumption of stationarity of the signal, while the brain activity is essentially non-stationary [18]. A common approach to resolve this conflict, divides the EEG signal into short time segments to satisfy piecewise stationary condition (i.e. fixed-size segmentation approach) [19]. Analysis of these segments may lead to some practical insight, however, it is statistically inefficient and provides an incomplete representation of the EEG signal [18]. This shortcoming results from the trade-off between the length of time segments and the stationarity assumption. For example, in the case of covariance matrices being used as descriptors, computing sample covariance matrices in very short time segments would lead to indefinite matrices, while in longer time segments it would lead to the occurrence of heterogeneities within segments with higher probability. Dividing EEG signals into the homogeneous pieces using adaptive segmentation methods [19], fairly resolve this problem by determining the length of homogeneous segments adaptively.

In spite of usual approaches that consider non-stationarity to be the result of external stimuli on brain functioning, the non-stationarity can be considered as the result of switching of the metastable state of the neural assemblies [18] or pathological changes. Ignoring these significant sources of non-stationarity can result in missing some valuable information about the brain functionality or loss of robustness in EEG-based systems [55, 56].

It is noteworthy that all derived homogeneous segments are not necessarily relevant to the subject's class label, (e.g. regions dominated by noise, different artifacts, biological functioning). In addition, the patterns which are related to a class may have multi-modal distribution in feature space or in other words, it may have different patterns related to each class. For example, for assessment of patients with epilepsy, different epileptiform EEG patterns are identified [∘ˇ]. In the case of children with different types of ADHD, the intergroup differences in EEG power provide insight into the brain function of the subjects [65]. In BCI application, patterns related to each subject may differ in different sessions (considerable inter class-variation exist) [62].

When the spatial covariance matrices are used for describing the segments of EEG signal and a Riemannian metric is used to compute the distances [17], we can assume artifacts, noise, and any kind of brain activity that is not related to the class label to have representations that are adequately different than the representation of patterns appeared by the class label in EEG signal. Therefore, it could be expected to have a feature space with a multi-modal distribution of data points to represent a subject. If we assign the same label to all pieces extracted from the signal recorded for a person and then embed all of the derived homogeneous segments in a single instance learning framework, it would result in a complex and non-linear separable distribution of data points (i.e., with overlapping between two classes in some parts). Especially when the signals are dominated by the physiological artifacts in different time segments, the overlapping between two classes would be considerable (in single instance framework). Therefore, in such cases, a discriminative classifier such as SVM would suffer from high training error rate, which leads to increasing the upper-bound of the probability of the test errors (i.e., the low generalization of the system) [63].

In addition, in some applications, there is a dependency between the rates of some physiological artifacts and the subjects' class label [33, 62]. In such applications, the artifacts may convey important information about the class type. Therefore, removing the artifact in a preprocessing step or analyzing the EEG signals in a single instance framework without removing the artifacts, which leads to overlapping between classes and misclassification of similar patterns (i.e., representation of segments dominated by the artifacts) miss the discriminative information conveying by some artifacts.

By considering the above-mentioned facts, our aim in this study is to introduce a new representation which has three simultaneous objectives; 1) to include the non-stationarity of the EEG signal and considering the different sources which cause the non-stationarity in EEG, 2) to overcome the shortcoming of fixed-size segmentation technique by using adaptive segmentation and 3) to overcome the shortcomings of single-instance learning process using multi-instance learning framework [47].

In our new representation, we describe the extracted segments using a robust representation, e.g. covariance matrix, then the resulting positive or negative patterns are embedded in

a multi-instance (MI) framework. This representation is conceptually compatible with the heterogeneity of the EEG signal during recording, which is produced by different sources of non-stationarity. Therefore, a subject is described using a bag of concepts (representation of homogeneous intervals) in the MI framework [20]. The concepts are derived by applying an adaptive segmentation approach to avoid the shortcomings of using fixed-size segmentation of EEG signals. The robust representation of homogeneous segments and applying bag-level discriminative information lead to increasing the generalization of the system. Our proposed representation is based on applying Riemannian geometry for determining the boundaries of homogeneous segments and describing the dataset as- a similarity matrix using the MI kernel.

The remainder of the paper is organized as follows: In section II we describe the mathematical preliminaries which are required for better understanding of this paper. In section III our proposed approach for representing and analyzing EEG signal is described. Experimental setups and the results are described in section IV. Finally, the results are concluded in section V.

II. BACKGROUND

In this section, we first review some basic concepts in Riemannian geometry, then describe the preliminaries of multi-instance learning framework, which are necessary for reading the paper.

A. Riemannian geometry and the manifold of SPD matrices

A Riemannian manifold (M, g) is a differentiable manifold M, which is endowed with a smooth inner product (Riemannian metric $g(u, v)$) on each tangent space $T_P M$. The P denotes the base point of the tangent space. The inner product (Riemannian metric) in Riemannian manifolds is a metric that allows measuring similarity or dissimilarity of two points on the manifold [21, 22, 23].

A curve $\gamma: I \subset R \to M$ is a geodesic if the rate of change of $\dot\gamma$ has no component along the manifold for all $t \in I$ or $\ddot\gamma$ is 0 [21]. Given a vector v in the tangent space $T_P M$, there is a geodesic $\gamma(t)$ which is characterized by its length, where geodesic issued from $\gamma(0) = P$, and $\dot\gamma = v/\|v\|$. Two points on the manifold may have multiple geodesic between them, but the ones which have minimum length is called minimizing geodesic [21].

In this paper, we describe the inputs of our system (e.g. recorded EEG) using the covariance matrices. The space of covariance matrices or in general terms symmetric positive definite matrices, does not satisfy the scalar multiplication axiom of a vector space (i.e. the multiplication of an SPD matrix with a negative scalar value is not an SPD matrix). Since the space of $d \times d$ dimensional SPD matrices, Sym_d^+, forms a convex cone in $\mathbb{R}^{d^2}$ Euclidean space, using a Riemannian metric to analyze the geometry of the space of Sym_d^+ is more compatible with its non-linear structure in comparison with investigating it in $\mathbb{R}^{d^2}$ Euclidean space. A number of different metrics have been proposed for Sym_d^+ to capture its non-linear

structure [17, 24, 25]. For example, log-Euclidean [24] and affine-invariant Riemannian metric [25], which induce log-Euclidean and affine-invariant geodesic distances, are two popular metrics used over the manifold of SPD matrices. A geodesic that connects two SPD points using log-Euclidean metric is defined as:

$$\gamma(t) = exp\big((1-t)\,log(C_1) + t\,log(C_2)\big) \tag{1}$$

where $t \,\epsilon\, [0,1]$ and $C_1, C_2 \,\epsilon\, \text{Sym}_d^+$. Log-Euclidean geodesic distance between C_1 and C_2 (i.e. the minimizing geodesic derived from log-Euclidean metric) can be expressed as:

$$d_{LE}(C_1, C_2) = \|log(C_1) - log\,(C_2)\|_F \tag{2}$$

where $d_{LE}(C_1, C_2)$ is the log-Euclidean distance and $\|.\|_F$ denotes the Frobenius matrix norm. The affine-invariant metric is the other effectiveF metric which induces geodesic distance over the manifold of SPD matrices:

$$d_{AI}(C_1, C_2) = \left\|log\big(C_1^{-1/2} C_2 C_1^{-1/2}\big)\right\|_F \tag{3}$$

where d_{AI} denotes the affine-invariant geodesic distance between two $C_1, C_2 \,\epsilon\, \text{Sym}_d^+$ points [24].

B. Multi-instance learning framework

Multiple-instance (MI) learning [20] is a variety of inductive machine learning methods (commonly supervised learning), which instead of learning over a set of individually labeled instances, the learner receives sets of labeled bags. For example, in the multiple-instance binary classification, let χ be the instance space (or the space of feature vectors) and $\Omega = \{+, -\}$ be the binary class attributes. The aim of learning in MI framework is finding a $v_{MI}: \mathbb{N}^\chi \to \Omega$ function, using training samples, where $\mathbb{N}^\chi$ refers to the set of all functions from χ to $\mathbb{N}$. $\mathbb{N}^\chi$ is isomorphic to the set of all multi-subset χ and a function $g(x) \epsilon \mathbb{N}^\chi$ returns the number of occurrence of instance x in the multi-set [47]. In other words, v_{MI} maps a bag composed of one or more occurrences of different instances to a positive or negative label.

Standard MI assumption assumes that each instance belonging to a bag has a hidden label which is either positive or negative $l \in \Omega = \{+, -\}$. A positive label can be assigned to a bag if and only if it contains at least one positive instance. In other words, let $X = \{x_1, x_2, \ldots, x_{N_X}\}$ be a bag or multi-set with N_X instances. The label of an instance is defined by a function $h: \chi \to \Omega$. Considering the positive and negative labels in concept-level as logical "true" and "false" constants, we can state the standard MI assumption as:

$$v_{MI}(X) \Longleftrightarrow (h(x_1) \vee h(x_2) \vee \ldots \vee h(x_{N_X})) \tag{4}$$

where v_{MI} is the MI concept function and $\vee$ is the logical disjunction operator. Several single-instance learning methods such as support vector machine (SVM) [26], neural networks [27], decision trees [28, 29], and ensemble-based learning [30] have been adapted to the multi-instance framework under the standard MI assumption.

In MI learning literature, different relaxations of this hard assumption are proposed to satisfy the requirements of some other problems, which are not exactly compatible with the standard MI assumption. For example, a metadata-based approach is a simple MI learning approach which replaces each bag with a metadata feature vector derived somehow from the
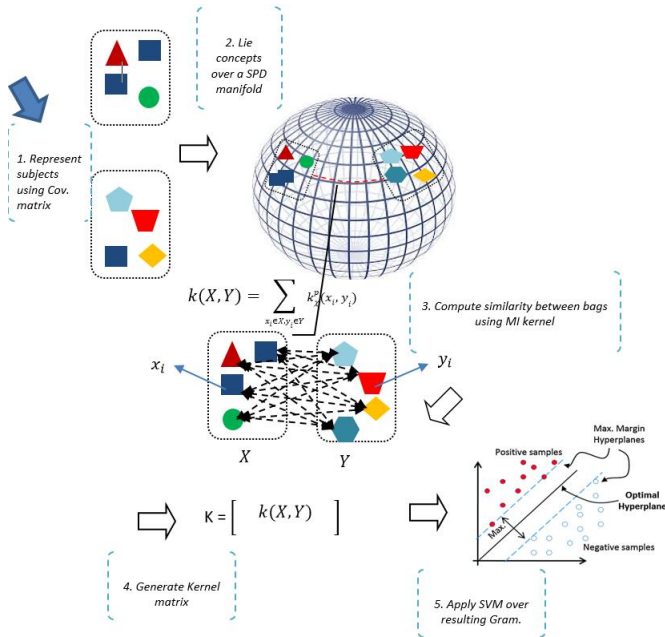
Fig. 1. The diagram of the proposed system based on MI kernel for representing the feature space.
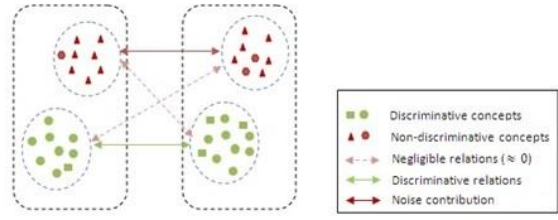

Fig. 2. Different types of relations in computing MI kernel between two bags

neural assemblies or pathological changes, ambient noise, artifacts, and biological sources), we interpret the recorded trial of subject X as a bag of N_X different concepts ($X = \{x_1,.., x_i, ..., x_{N_X}\}, x_i \in \chi$). First of all, the different EEG signals are filtered by different band-pass filters. Then the signals are divided into the homogeneous segments based on an adaptive segmentation approach described in section III.B. After that, the concepts are described by the covariance matrices of homogeneous segments. The covariance matrices are approximated using sample covariance matrices as:

$$C(t,T) = \frac{1}{T-1} E_{(t,T)} \times E_{(t,T)}^T \qquad (6)$$

where $E_{(t,T)}$ denotes an EEG segment started at the time t of the recording with the total length of T seconds, and $C(t,T)$ denotes the empirical covariance matrix of this segment. Therefore, a subject is described as a bag of $n_c \times n_c$ matrices (i.e., as 2nd-order tensors) where n_c denotes the number of channels used in the analysis.

Different sources of non-stationarity of EEG signal lead to two types of concepts in the EEG signal: the concepts which are relevant to the class label and the concepts which are not relevant to the class label. Let consider the relevant ones as positive concepts ($+$) and the irrelevant ones as negative concepts ($-$). In an EEG-based system, a subject can be labeled with a specific class if there exist a specific pattern or combination of patterns corresponding to that class in its brain electrical activity. For example, in an EEG-based diagnosis system, obviously, a subject can be labeled with a specific disorder if its EEG contains the patterns which are corresponding to that disorder (positive instances corresponding to that class). This representation and interpretation of recorded non-stationary trials are compatible with the multi-instance learning assumption. Therefore, by embedding the retrieved instances in MI framework, we can consider the discriminative information on bag-level using MI learning algorithms.

In some applications, the patterns of a specific class are easily recognized by experts via visual inspection of the subjects' brain electrical activity. However, in some others, more accurate analysis of the EEGs is required to detect the patterns of the signal affected by a specific class (ex., the disorder, the culture, or action imagery). In these cases, explicit concept-level labeling of the signals is impossible and it is required to adapt machine learning techniques for mining the underlying patterns. As the MI kernel approach is able to compute similarities without explicitly inducing the concepts [31], we use it for EEG signal representation.

In MI kernel approach, the similarity between subjects is

instances of that bag. Therefore, the MI learning problem is converted to a single instance learning problem. This transformation can be done either implicitly or explicitly. After this transformation, in the learning process, a single-instance learning algorithm can be applied to the resulting transformed feature space. This approach implicitly is based on an assumption, called metadata assumption, which states that the labels of the samples are directly related to their representation in transformed space (metadata feature space) [47]. For example, [31] proposed MI kernels that can apply any kernel-based method such as standard SVM algorithm to MI data. The kernel is defined as:

$$k_{MI}(X,Y) = \sum_{x_i \in X, y_i \in Y} k_\chi^p(x_i, y_i) \qquad (5)$$

where k_χ^p is the pth power of an instance-level kernel k_χ, x_i and y_i denote instances while X and Y denote bags. For a positive definite (PD) kernel k_I, the pth power of it generates a PD kernel and it is proved that for p values that are sufficiently large [50], if the instance-level kernel is separable its corresponding MI kernel is separable too.

III. PROPOSED METHOD

Although the information geometry in EEG processing shows the promising results in the related literature, the golden way to use it for representing the non-stationarity of EEG signal is open. We propose a new representation of an EEG signal, which intends to model the non-stationary of the signal and increase the robustness of the system (Fig. 1). In this section, we first introduce our representation and then describe the EEG segmentation method used to provide homogeneous segments.

A. MI representation of EEG signal

To represent the variations produced by different sources of non-stationarity (ex., switching of the metastable state of the

computed by a generalization of the inner product between every two concepts obtained from distinct subjects as:

$$k(X,Y) = \frac{\sum_{x_i \epsilon X, y_j \epsilon Y} k_\chi{}^p(x_i, y_j)}{N_X \times N_Y} \tag{7}$$

where $k(X,Y)$ denotes the similarity between X and Y bags that represents two subjects, x_i and y_j denote the concepts that belong to the X and Y bags, respectively, and N_X and N_Y denote the number of concepts in the X and Y bags. $k_\chi(x_i, y_j)$ is an instance-level kernel defined in instance space χ. According to the (7), computing MI kernel $k(X,Y)$ leads to computing $N_X \times N_Y$ concept-level computation of the kernel function. By considering that, each subject is represented by positive and negative concepts, as illustrated in Fig. 2. These $N_X \times N_Y$ calculations can be categorized into three groups, according to the types of the concepts involved in the computation; computing $k_\chi(x_i, y_j)$ between two positive instances, two negative instances, and a positive and a negative instances (abstracted as discriminative relation, noise contribution, and negligible relations in Fig. 2). From the machine learning point of view, the discrimination between two classes depends on the classification between their positive instances. Therefore, we consider the relationship between instances of two subjects as discriminative relations. All pairs which are dominated by other sources of non-stationarity of the signal are considered as noise contribution. In other words, $k(X,Y)$ can be considered as the sum of three types of similarities:

$$k(X,Y) = k_{+,+}(X,Y) + k_{+,-}(X,Y) + k_{-,-}(X,Y) \tag{8}$$

$$k_{+,+}(X,Y) = \sum_{x_i \epsilon X^+, y_j \epsilon Y^+} k_\chi(x_i, y_j)$$

$$k_{-,-}(X,Y) = \sum_{x_i \epsilon X^-, y_j \epsilon Y^-} k_\chi(x_i, y_j)$$

$$k_{+,-}(X,Y) = \sum_{x_i \epsilon X^+, y_j \epsilon Y^-} k_\chi(x_i, y_j) + \sum_{x_i \epsilon X^-, y_j \epsilon Y^+} k_\chi(x_i, y_j)$$

where

$$X^+ = \{\forall x_i \epsilon X | h(x_i) == '+'\}$$
$$X^- = \{\forall x_i \epsilon X | h(x_i) == '-'\}$$
$$Y^+ = \{\forall y_i \epsilon Y | h(y_i) == '+'\}$$
$$Y^- = \{\forall y_i \epsilon Y | h(y_i) == '-'\}$$

where $h(.)$ is an implicit concept level function and | and & are logical *or* and *and* operators, . By considering the negative (non-discriminative) concepts (i.e. noise, artifact,..) as any kind of concepts which are different enough to be compared to the positive (discriminative) concepts [31], $k_{+,-}(X,Y)$ would have a negligible contribution in computing the similarity between every two subjects (i.e. $k_{+,-}(X,Y) \cong 0$).

The $k_{+,+}(X,Y)$ term denotes the similarity between positive instances of two subjects and represents the discriminative information. It is expected that the subjects with the same label will show higher $k_{+,+}(X,Y)$. For example, in EEG-based disorder diagnosis, similarity of the patterns which are the result of the same psychiatric disorders or in the application of investigating the effect of culture on brain electrical activity, the existence of similar patterns evolved as the result of the similar culture leads to high similarity, while subjects with dissimilar

class will show smaller $k_{+,+}(X,Y)$. The smaller $k_{+,+}(X,Y)$ value is the result of dissimilarity of positive concepts. This results in higher/lower $k(X,Y)$ value for similar/dissimilar class (i.e. $k(X,Y) \gg k(X,Z)$ where $v_{MI}(X) == v_{MI}(Y)$ and $v_{MI}(X) \neq v_{MI}(Z)$).

The electrical activities arising from sources other than the brain are termed artifacts and can be divided into physiologic and extra-physiologic artifacts. Since the setups are the same during recording in the laboratory for all subjects in an experiment, extra-physiologic artifacts, arising from outside the body, can be considered approximately the same for all the subjects (i.e., $k_{-,-}(X,Y) \cong cte$).

It may exist correlation between the rate of some physiologic artifacts and the class type in some applications. For example, in mental disorder diagnosis application, for some disorders such as Schizophrenia, seasonal affective disorder, and depression exist dependency between eye blink rate and the disorder type [33, 62]. In such cases, the increased rate of an artifact within a class leads to increased $k_{-,-}(X,Y)$ and consequently increased $k(X,Y)$ in comparison with $k(X,Y)$ between the subjects of the other class without such dependency. Intuitively, this lead to increasing the margin size or decreasing the summation of slack variables in resulting feature space for SVM classifier and consequently decreasing the upper bound of probability of test error which means more generalization to test samples. While, in single-instance learning framework, high rate of an artifact within a class leads to increasing the probability of missing these samples by an SVM classifier, due to overlapping between classes on these concepts. By increasing the summation of slack variables for SVM classifier, the upper bound of probability of test samples is increased which means lower generalization of the system to test samples [67].

Since analyzing the covariance matrices as 2nd-order tensors in Riemannian framework leads to superior results in comparison with analysis of its vectorized equivalent in Euclidean space [32], we choose $k_\chi(x_i, y_j)$ compatible with the Riemannian geometry of the manifold of SPD matrices. Depending on the classifiers that are used, different types of the kernel can be selected from different points of view. We have studied two popular kernels which are compatible with the manifold of SPD matrices:

1) Log-Euclidean

The Log-Euclidean kernel is defined as:

$$K_{LE}: Sym_d^+ \times Sym_d^+ \to \mathbb{R}$$
$$k_{LE,\chi}(x_i, y_j) = exp(-\frac{d_{LE}^2(x_i, y_j)}{\sigma}) \tag{9}$$

where $k_{LE,\chi}(x_i, y_j)$ denotes the similarity between x_i and y_j concepts based on the Log-Euclidean distance, d_{LE} and σ is a positive value known as bandwidth parameter. Since Log-Euclidean Gaussian kernel satisfies mercer condition for all $\sigma > 0$ [32] (i.e. it is a positive semi-definite kernel), the resulting $k(X,Y)$ which would be the sum of multiple positive

semi-definite (*PSD*) kernels is a *PSD* kernel and can be used in any kernel-based learning method.

2) Affine-invariant Gaussian kernel

Using affine-invariant Riemannian distance as (10) is called affine-invariant Gaussian kernel:

$$K_{AI}: Sym_d^+ \times Sym_d^+ \to \mathbb{R} \qquad (10)$$

$$k_{AI,\chi}(x_i, y_j) = exp(-\frac{d_{AI}^2(x_i, y_j)}{\sigma})$$

Although this kernel generally is not a *PD* kernel, an empirical affine-invariant Gaussian kernel might be a *PD* kernel depending on the distribution of data points and its bandwidth parameter.

Both of these kernels are able to provide a discriminative projection of data points into the feature spaces, therefore they are appropriate choices for SVM classifier. SVM is based on minimizing a regularized combination of empirical and structural risk. This objective function, which can be interpreted as maximizing the margin between two classes in the feature space, leads to the robustness of the SVM classifier in EEG analysis [51].

B. Identifying the homogeneous segments

To determine the boundaries of homogeneous segments, we apply an adaptive segmentation approach in which the boundaries of the segments are recognized as points where the resulting segments satisfy stationarity condition (being homogeneous in the statistical sense). At first, the EEG signals are divided into elementary segments using fixed-size segmentation approach. We describe the segments using spatial covariance of the pieces (6). To detect the boundaries of homogeneous segments we compare spatial covariance matrices of the successive elementary segments in terms of geodesic distance. By applying a thresholding method, we can detect the local peaks that are identified as cut points or the boundaries of homogeneous segments of the recorded EEG. New segments will be generated from merging the successive elementary segments between two cut points. This process can be summarized as:

$$d_G\big(C(t,T), C(t+T,T)\big) > th = \begin{cases} t+T \in \mathrm{T}_{int} & if\ No \\ t+T \in \mathrm{T}_{cut} & if\ Yes \end{cases} \quad (11)$$

where the th is an empirical threshold level, d_G denotes the geodesic distance, T_{cut} denotes the set of cut points of a trial, and T_{int} denotes the set of boundary points of elementary segments of a trial, which should be considered as interior points of a segment resulted by adaptive segmentation.

IV. EXPERIMENTS

In this section, we evaluate the proposed representation in two different applications. These two applications are indoor problems and all the datasets are recorded in the laboratory with the same set-up for all the subjects involved in an experiment. The aim of first experiment is finding the answer of this question that whether there exists any considerable difference between brain electrical activity of Iranian versus Swiss normal subject. The second application is EEG-based mental disorder diagnosis, that includes different psychiatric disorder diagnosis
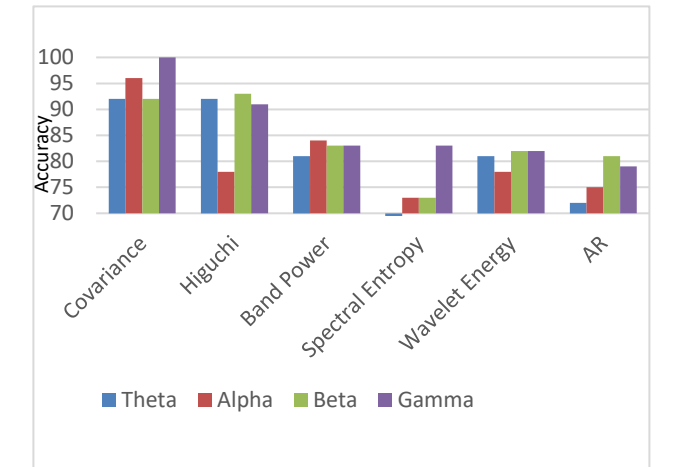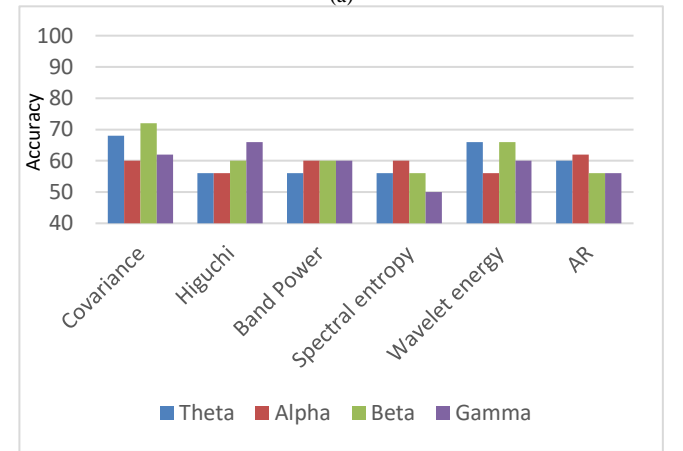

(a)

(b)

(c)
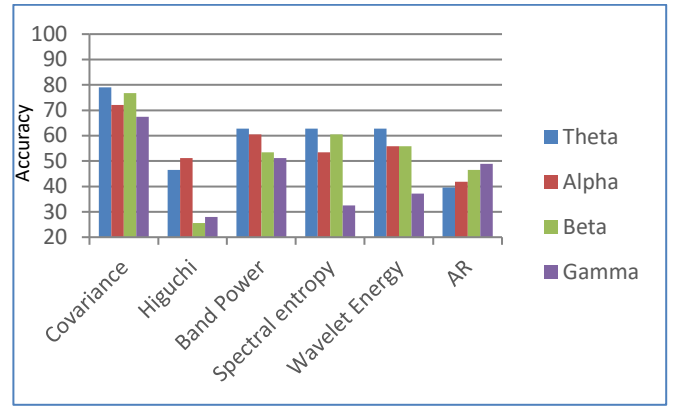
Fig. 3. Comparison of the classification accuracy of different features extracted from EEG signals of the (a) ADHD/BMD (b) Depression/Normal (c) Iranian/Swiss normal subject's datasets in the eyes-closed resting condition.

applications: Attention Deficit Hyperactivity Disorder (ADHD) / Bipolar Mood Disorder (BMD), depression/ normal, and Schizophrenia/ normal group.

We first describe the datasets and the recording conditions of EEG signals for these two datasets, then we describe different experimental setups and results, and finally discuss the results.

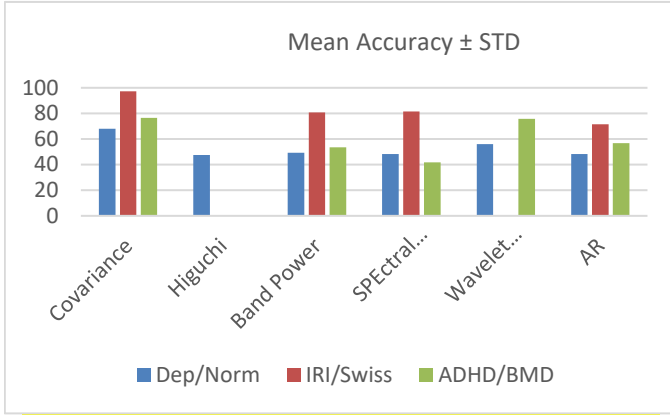*Datasets*

The specification and recording conditions of the datasets are described in details as follows:

1) Iranian/Swiss normal subjects

This dataset includes the EEG signal of 50 Iranian and Swiss healthy people recruited by an announcement. Data collection was performed in Atieh Clinical Neuroscience Center (ACNC) and Brain and Traumafoundation Grison/Switzerland. The inclusion criteria were the male and female in the range of 18-70 years old and have normal IQ. The exclusion Criteria were substance abuse, brain trauma, and having psychiatric or neurological disorders.

The SCL90 was used for the assessment of mental health and SPM for evaluation of psychometric (IQ) in people. Psychometric, demographic and socioeconomic status were measured for all participants.

The EEG data were recorded in both eyes-closed and eyes-open resting condition for five minutes.

The Mitsar 201 (Mitsar Ltd.) equipment was used for recording 19 channel EEG. The input signals referenced to the linked ears. The signals were filtered between 0.5 and 50 Hz and the sampling rate was 500 Hz. The impedance of the electrodes was under 5 kΩ. Electrodes placement were according to the 10-20 system using an electrode cap with tin electrodes (Electro-cap Inc.). The quantitative data were obtained using WinEEG software.

ADHD/BMD

ADHD and BMD are psychiatric disorders with similar clinical symptoms. The overlap between clinical symptoms of these two disorders leads to the unreliability of qualitative diagnosis approach. This dataset consists of the EEG signals of 43 children and adolescents, with 21 subjects with ADHD (age range: 10–20, age mean ± std: 14.36 ± 2.9) and 22 subjects with BMD (age range: 13–22, age mean ± std: 16.50 ± 2.50). Data acquisition was performed in the Biophysics Laboratory of Shiraz University of Medical Sciences. The children, adolescents and at least one of their parents were interviewed using DSM_IV criteria for diagnosis [36]. It should be noted

TABLE I
THE COMPARISON BETWEEN THE PROPOSED DESCRIPTOR (MI-ADAPTIVE-COV) WITH SOME COVARIANCE-BASED DESCRIPTORS, IN TERMS OF CLASSIFICATION ACCURACY ON (A) ADHD/BMD (B) DEPRESSION/NORMAL (C) SCHIZOPHRENIA/NORMAL (D) IRANIAN/ SWISS NORMAL SUBJECTS DATASETS IN THE EYES-OPEN AND EYES-CLOSED RESTING CONDITIONS.

	Theta	Alpha	Beta	Gamma	3-50 Hz
ADHD/BMD (Eye-open)					
SI-Cov+SVM	79.07	83.72	83.72	83.72	79.07
Mean-Cov+SVM	76.74	86.05	86.05	79.07	79.07
MI-fixed-Cov+SVM	79.07	86.05	88.37	81.40	83.72
MI-Adaptive-Cov+SVM	88.37	88.37	90.70	88.37	83.72
ADHD/BMD (Eye-closed)					
SI-Cov+SVM	79.07	72.09	76.74	67.44	74.42
Mean-Cov+SVM	72.09	72.09	72.09	69.77	62.79
MI-fixed-Cov+SVM	79.07	76.74	79.07	81.40	79.07
MI-Adaptive-Cov+SVM	83.72	83.72	88.37	76.74	88.37

(a)

	Theta	Alpha	Beta	Gamma	3-50 Hz
Depression/normal (Eye-open)					
SI-Cov+SVM	70.00	62.00	72.00	62.00	60.00
Mean-Cov+SVM	62.00	56.00	66.00	60.00	56.00
MI-fixed-Cov+SVM	72.00	76.00	76.00	64.00	82.00
MI-Adaptive-Cov+SVM	72.00	80.00	80.00	78.00	86.00
Depression/normal (Eye-closed)					
SI-Cov+SVM	68.00	60.00	72.00	62.00	58.00
Mean-Cov+SVM	62.00	68.00	72.00	68.00	56.00
MI-fixed-Cov+SVM	70.00	70.00	80.00	82.00	86.00
MI-Adaptive-Cov+SVM	74.00	74.00	80.00	84.00	90.00

(b)

	Theta	Alpha	Beta	Gamma	3-50 Hz
Schizophrenia/normal (Eye-open)					
SI-Cov+SVM	68.00	72.00	72.00	84.00	86.00
Mean-Cov+SVM	74.00	76.00	76.00	90.00	82.00
MI-fixed-Cov+SVM	90.00	86.00	88.00	88.00	96.00
MI-Adaptive-Cov+SVM	92.00	88.00	88.00	90.00	96.00
Schizophrenia/normal (Eye-closed)					
SI-Cov+SVM	78.00	66.00	74.00	80.00	80.00
Mean-Cov+SVM	84.00	86.00	80.00	70.00	88.00
MI-fixed-Cov+SVM	84.00	84.00	78.00	90.00	88.00
MI-Adaptive-Cov+SVM	90.00	88.00	88.00	90.00	94.00

(c)

	Theta	Alpha	Beta	Gamma	3-50 Hz
Iranian/Swiss (Eye-open)					
SI-Cov+SVM	97.00	96.00	91.00	97.00	**97.00**
Mean-Cov+SVM	95.00	٨٨.00	88.00	٩٧.00	93.00
MI-fixed-Cov+SVM	100	93.00	٩٠.00	٩٩.00	94.00
MI-Adaptive-Cov+SVM	100	٩٩.00	94.00	100	98.00
Iranian/Swiss (Eye-closed)					
SI-Cov+SVM	92.00	96.00	92.00	97.00	97.00
Mean-Cov+SVM	87.00	93.00	92.00	94.00	90.00
MI-fixed-Cov+SVM	93.00	98.00	95.00	98.00	96.00
MI-Adaptive-Cov+SVM	94.00	99.00	96.00	100	97.00

(d)

that most of the subjects were followed up by the psychiatrist for at least 6 months and the psychiatrist was assured about their

type of illness. For each patient, the EEG signals were recorded in two eyes-open and eyes-closed resting conditions for 3 minutes, in order to analyze the natural behavior of their EEGs. The EEG signals were recorded using the Neuroscan-LT setup (equipped with a 16bit A/D card). The signals were recorded using 22 electrodes according to the 10–20 international recording system. The scalp channels were located in the following positions: G, Fz, Cz, Pz, C3, T3, C4, T4, Fp1, Fp2, F3, F4, F7, F8, P3, P4, T5, T6, O1, O2, A1, and A2. The average value of the A1and A2 electrodes, which were attached to the earlobes, was used as the reference. The ground electrode (G) was placed on the forehead center. The impedances of the electrodes were lower than 10 kΩ through the recording and the sampling rate of the EEG was 250 Hz.

Depression/Normal

Depression is a common disorder with a high prevalence of critical outcomes, such as a suicidal attempt. The World Health Organization (WHO) has predicted that the depression would be the second largest burden of disease by 2020. Fifty patients with bipolar depression and major depressive disorder (MDD) were recruited for this study after their disorder was confirmed according to the DSM IV-TR criteria through a clinical interview. The patients were examined at the Atieh Clinical Neuroscience Center during a period from February to September 2017. The inclusion criteria were the following: outpatient subjects whose age ranged from 18 to 65, diagnosed as having major depression or bipolar mood disorder, according to the DSM IV-TR, had a Beck depression inventory (BDI-II) score of more than 14. The exclusion criteria were: personality disorder in Axis II, seizures, epilepsy in the first degree relatives, pregnancy and head trauma [37].

For recordings, a 19 ElectroCap (ElectroCap, Inc; OH) was used, with its electrodes placed on Fp1, Fp2, F7, F3, Fz, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6, O1, O2 by a 10–20 system. An A1 + A2 electrode was used for reference. The impedance of the electrodes was under 5 kΩ. EEG was recorded in 5 minutes, while the subjects were in the resting state with their eyes closed and in 5 minutes for eyes open resting condition. The recording was carried out using a Mitsar system in an acoustic room. The data were converted into numbers with 500 sample rates and with a high-frequency filter of 50 Hz, a low-frequency filter of 0.3.

Schizophrenia/ Normal

This dataset is composed of 25 Schizophrenic and 25 normal subjects. The patients participated in the study (mean age 25) who were diagnosed with schizophrenia based on clinical interview by a psychiatrist using the criteria in the fifth edition of diagnostic and statistical manual of mental disorders (DSM-V). Mean age of the patients in the time of onset of the disorder was 18 years. Seven of them have been hospitalized and two have received ECT. The recording performed in two eyes-open and eyes-closed resting condition in Atieh center. The set-up of EEG recording is the same as depression/normal dataset.

B. Experiments & discussion

To evaluate the proposed approach, we designed several experiments in two different applications. As the first experiment, we examined the effectiveness of covariance in against a wide range of descriptors commonly used EEG-based applications, including Higuchi [39], descriptor autoregressive [40], power spectrum [41], spectral entropy [42], and wavelet energy [43] methods. Iranian/Swiss normal subjects, ADHD/BMD, and depression/normal datasets in the eyes-closed resting conditions are used in this experiment. We first filter the signal into the Theta (3-8Hz), Alpha (8-12 Hz), Beta (12-28 Hz), and Gamma (28-50 Hz) sub-bands using the 5th–order band-pass Butterworth filter. Then the above methods are used for representing the signal in segments resulted from applying fixed-size segmentation over the recorded signals. The signals are divided into overlapping windows (50% overlapping between segments) and the length of the segments is determined empirically by applying the leave- one subject-out cross-validation and investigating a wide range of values (up to 20 seconds) as window length. For each candidate window size, we apply leave-one subject-out, seek the parameters of optimal classifier for that selected window size, compare the result with the result of optimal models in other window sizes, and choose the best window size for further analysis on that dataset. In this experiment, we consider these features as the vectors in Euclidean space. These vectors are the result of the application of above-mentioned methods on all recorded channels except A1, A2, and G channels [12]. In order to have a quantitative comparison between different descriptors, an SVM classifier [44] with RBF kernel is used for classification. We used the LIBSVM package [45] for implementing SVM classifier. Tuning the bandwidth parameter of RBF kernel and C parameter of SVM classifier is done by applying leave- one subject-out cross-validation and examining a wide range of values (i.e. $C \epsilon \{0.1, 10, \dots, 100000\}$, $\sigma \epsilon \{0.1, 1, 10, 20, \dots, 1000\}$) for selecting the best value over the validation set.

The covariance matrix as a descriptor was estimated empirically by applying (6) on time segments of 19 recorded channels. The setup of this experiment was the same as the above-mentioned conditions. With the exception that the analysis was based on the Riemannian geometry of the manifold of SPD matrices and affine-invariant Gaussian kernel [32] was used for computing similarities. Fig. 3 illustrates a comparison between above-mentioned descriptors. It is obvious that in most sub-bands covariance matrix as a descriptor leads to superior results. To assess the standard deviation of classifier on different descriptor, we compute mean accuracy and standard deviation on 3-50 Hz sub-band for ADHD/BMD, Depression/Normal, and Irannian/Swiss normal subjects data sets, illustrated in Fig. 4.

In the second experiment, we compared the proposed MI– based representation (MI-Adaptive-Cov) with other covariance-based approaches. In the case of adaptive segmentation, the signals are divided into segments with 2 seconds length, without overlapping between windows, and the threshold value for merging or splitting consequent segments are set by the mean value of the differences between every two successive segments, in terms of geodesic distance, for each subject. The goal of these comparisons is to evaluate the

TABLE II

EXAMINING THE STATISTICAL SIGNIFICANCE OF THE PROPOSED METHOD IN COMPARISON WITH ITS COMPETITORS REPORTED IN TABLE. I, *: P < 0.05, **: P < 0.0125 AFTER BONFERRONI CORRECTION

	Depression/Normal		ADHD/BMD		Swiss/Iran		Schizophrenia/normal	
	Eyes-open	Eyes-closed	Eyes-open	Eyes-closed	Eyes-open	Eyes-closed	Eyes-open	Eyes-closed
MI-Adaptive-Cov/SI-Cov	0.0279*	0.0267*	0.0029**	0.0029**	0.0728	0.0830	0.0031**	0.0005**
MI-Adaptive-Cov/ Mean-Cov	0.0057**	0.0502	0.0190*	0.0129*	0.0115**	0.0032**	0.0046**	0.0187*
MI-Adaptive-Cov/ MI-fixed-Cov	0.0897	0.0249*	0.0705	0.1197	0.1036	0.2126	0.5262	0.0282*
MI-fixed-Cov/ SI-Cov	0.0915	0.0423*	0.2414	0.0972	0.5000	0.2229	0.0054**	0.0125**

different components involved in our representation. The representations of the EEG signals for methods involved in our comparisons are as follows:

1) Representing subjects in MI framework using MI-kernel, each subject is a bag of covariances of the segments generated by fixed-size segmentation (MI-fixed-Cov).

2) Representing the feature space using a group of covariance matrices [19]. The covariances are computed in segments generated by a fixed-size segmentation approach and the segments are labeled the same as the subject's label (Representing in single instance framework called SI-Cov).

3) Representing each subject using the geometric mean [46] of the covariance matrices of the segments generated by the fixed-size segmentation of the trial (Mean-Cov).

We applied SVM classifier with an affine-invariant Gaussian kernel for discrimination over SI-Cov and Mean-Cov. In addition, in this experiment, the basic kernel in MI-fixed-Cov and MI-Adaptive-Cov is affine-invariant Gaussian kernel. The σ and C parameters, as described in the first experiment, were set using cross-validation. The p parameter in equation (7) is set to 1 in our experiments. We evaluate the methods using leave-one subject-out cross-validation approach. The experimental evidences confirm the superiority of the MIL-Adaptive-Cov+SVM for the Iranian/Swiss normal subjects, ADHD/BMD, Depression/normal, Schizophrenia/normal datasets, in two eyes-open and eyes-closed resting conditions (Table I).

The superiority of the proposed method is further investigated by:

1) Considering the non-stationarity of the EEG signal by applying the MI framework in the proposed representation. It is evaluated by (SI-Cov+SVM/ MI-fixed-Cov+SVM) and (SI-Cov+SVM/ MI-Adaptive-Cov+SVM) comparisons (Table I and Table 2).

2) Determining the boundaries of the segments adaptively, which is evaluated by (MI-fixed-Cov+SVM / MI-Adaptive-Cov+SVM) comparison (Table I and Table 2).

3) Doing all of the analysis in Riemannian framework. We did some analysis over covariance matrices using Euclidean geometry which led to the overfitting to the training samples [66].

4) Applying appropriate representation for basic elements (concepts) in the proposed representation, confirmed by the first experiments (Fig. 3 and Fig. 4).

The significance of the superiority of the proposed approach is examined in comparison with other covariance-based

TABLE III. COMPARISON BETWEEN AFFINE-INVARIANT AND LOG-EUCLIDEAN GAUSSIAN KERNELS USED IN THE PROPOSED APPROACH

	Depression/Normal		ADHD/BMD	
Kernels	Eyes-open	Eyes-closed	Eyes-open	Eyes-closed
AIGK	86.00	90.00	83.72	88.37
LEGK	76.00	80.00	74.42	76.74

TABLE IV. PERFORMANCE OF MI-ADAPTIVE-COV+SVM ON ADHD/BMD, DEPRESSION/NORMAL, IRAN/SWITZERLAND NORMAL SUBJECTS' DATASETS IN THE EYES-OPEN AND EYES-CLOSED RESTING CONDITIONS, THE DATASETS ARE CONTAMINATED BY WHITE GAUSSIAN NOISE WITH DIFFERENT SNR LEVELS.

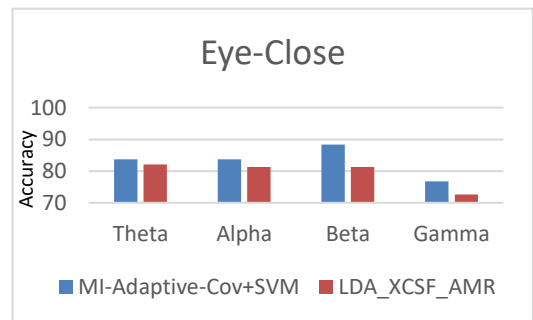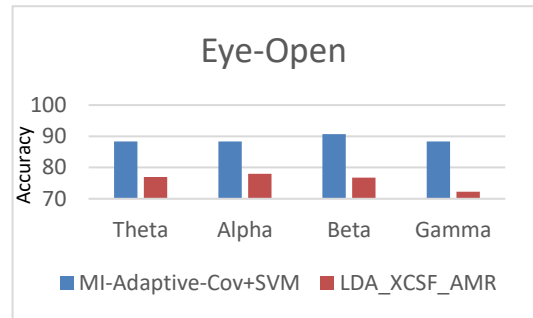|	10db	5db	0db
Depression/normal (eyes-closed)	88.00	88.00	86.00
Depression/normal (eyes-open)	84.00	82.00	82.00
ADHD/BMD (eyes-closed)	86.05	86.05	86.05
ADHD/BMD (eye-open)	81.40	81.40	79.07
Iranian/Swiss (eyes-closed)	٩٦٫٠٠	٩٦٫٠٠	٩٦٫٠٠
Iranian/Swiss(eyes-open)	96.00	95.00	95.00
Schizophrenia/normal (eyes-closed)	94.00	94.00	92.00
Schizophrenia/normal (eyes-open)	96.00	96.00	96.00

Fig. 5. Comparison of the classification accuracy of MI-Adaptive-Cov+SVM and LDA_XCSF_AMR [36] on ADHD/BMD dataset.

methods (Table II). We apply paired t-test over the classification rates resulted from applying MI-Adaptive-Cov for describing subjects versus SI-Cov, Mean-Cov, and MI-fixed-Cov in different sub-bands for Iranian/Swiss normal subjects, ADHD/BMD, depression/normal, Schizophrenia/normal datasets. The resulting p-values for

comparison between MI-Adaptive-Cov versus SI-Cov, reported in Table II, confirm the significance of the superiority of the proposed approach (p-value < 0.05) against classic covariance-based representation using single instance framework (SI-Cov) in most cases. Considering the results reported in table II, confirm that the strength of the proposed representation is especially in the cases where the classes are not linear discriminative in single instance framework and there is correlation between some artifacts and the class type (ex., Schizophrenia/normal dataset). In the cases, two classes are fairly separable in single instance framework the improvement resulted by our proposed approach is not statistically significant (Iranian/Swiss dataset).

Our proposed method also has significantly superior results in most cases (Table II) in comparison with Mean-Cov, which misses a large amount of information by describing each subject using the geometric mean point of the covariance of segments resulted by segmenting EEG of each subject.

To show the importance of adaptive segmentation, we compared non-adaptive MI-fixed-Cov method against MI-Adaptive-Cov. As shown in Table II, the adaptive segmentation is not the reason of significant superiority of our proposed approach. Adaptive segmentation for determining the concepts besides the analyzing the resulting concepts in MI framework leads to the significant superiority of MI-Adaptive-Cov+SVM against SI-Cov+SVM in most cases.

The resulting similarity matrix of our proposed representation has to satisfy the *PSD* constraint if it is going to be used as an empirical kernel matrix for kernel-based learning systems. Using non-*PSD* kernels, when the learning system minimizes the empirical risk (such as SVM) leads to a non-convex optimization problem. In such cases, the minimization of risk function is not guaranteed [49]. The positive definiteness of our proposed representation is controlled by the type of concept-level kernel and its parameters.

Similar to any kernel-based method, the type of kernel plays an important role in the classification results. We have studied the performance of two well-known kernels, including Log-Euclidean Gaussian kernel (LEGK) and affine-invariant Gaussian kernel (AIGK) at concept-level (Table III). These comparisons are performed on the ADHD/BMD and depression/normal datasets in two eyes-open and eyes-closed resting conditions. The EEG signals are filtered in 3-50 Hz.

The concept-level positive definiteness of Log-Euclidean Gaussian kernel, leads to positive definiteness of its sum over different pairs of concepts in MI framework. Therefore, it provides a *PD* representation. Gaussian kernel based on the affine-invariant Riemannian metric is not *PD* in general.

Although the positive definiteness of the similarity matrix, guarantees the convergence of the kernel-based method to an optimal solution, but it does not necessarily lead to superior classification performance in comparison with non-*PSD* kernels. The superior results achieved by affine-invariant Gaussian kernel in this experiment confirm this fact. One interesting observation in this experiment is the positive definiteness of empirical affine-invariant Gaussian kernel achieved on ADHD/BMD and depression/normal datasets. This

observation is confirmed by a recent work done by Feragen and Hauberg [48], which states that the affine-invariant Gaussian kernel has a high probability of being *PD* for a given data set in a large range of values.

To confirm the usefulness of the proposed representation, we compare the result of the MI-Adaptive-Cov+SVM in comparison with the results of another research [36] which is applied on the same ADHD/BMD dataset (Fig. 5). The experimental results confirm the superiority of our proposed approach. In addition, comparison between our proposed method with the state of the arts of analysis over the manifold of symmetric positive definite matrices in Riemannian framework, including MDM, TSLDA, FGMDM [64], and RKSVM [68] methods are reported in Table V. In this experiment, the schizophrenia/normal dataset in the eyes-open resting condition is used. The significant superiority of MI-Adaptive-Cov+SVM in comparison with competitors is confirmed by applying paired t-test, reported in table VI.

Robustness of the proposed approach against noise was examined by adding the white Gaussian noise with different signal-to-noise ratios (SNR) to signal (10db, 5db, 0db). We tried this experiment on Iranian/Swiss normal subjects, ADHD/BMD, depression/normal, and schizophrenia /normal datasets in the eyes-closed and eyes-open resting conditions. The signals involved in this experiment are filtered in 3-50Hz and leave-one subject-out cross-validation is used for evaluation. The performances of the proposed method on signals which are contaminated by white Gaussian noise with different SNR levels are reported in Table IV. As the experimental evidences confirm, the proposed approach is relatively robust in noisy conditions. The performance of the method decreases about 2 to 4 percent where the noise is considerable (0db) and about 2 percent where the SNR is equal to 10db.

TABLE V. THE COMPARISON BETWEEN THE PROPOSED DESCRIPTOR (MI-ADAPTIVE-COV) WITH MDM, TSLA, AND FGMDM ON SCHIZOPHRENIA/NORMAL DATASET

Schizophrenia/normal (Eye-open)				
	Theta	Alpha	Beta	Gamma	3-50 Hz
MI-Adaptive-Cov	**92.00**	**88.00**	**88.00**	**90.00**	**96.00**
MDM	74.00	67.00	66.00	88.00	84.00
TSLDA	82.00	86.00	86.00	86.00	86.00
FGMDM	76.00	86.00	86.00	76.00	84.00
RKSVM	90.00	84.00	86.00	82.00	90.00

TABLE VI. EXAMINING THE STATISTICAL SIGNIFICANCE OF THE PROPOSED METHOD IN COMPARISON WITH ITS COMPETITORS REPORTED IN TABLE.V, *: P < 0.05, **: P < 0.0125 AFTER BONFERRONI CORRECTION.

	MI-ADAPTIVE / MDM	MI-ADAPTIVE /TSLDA	MI-ADAPTIVE / FGMDM	MI-ADAPTIVE / RKSVM
P-VALUE	0.006467 **	0.005431 **	0.005137 **	0.039743 *

V. CONCLUSION

In this paper, we proposed a new representation for describing EEG signals. In this representation, we apply the MI

framework to consider the non-stationarity of the EEG signal. The concepts in this representation are described by the sample covariance matrix of EEG segments. These segments are the result of the adaptive segmentation of the signal and all the analysis are performed in Riemannian framework. Using MI kernel for describing EEG signal is a suitable choice which magnifies the discriminative information by considering the similarity between each pair of concepts (including positive and negative concepts). In addition, automatically attenuating the extra-physiologic noise contribution and exploiting the discriminative information of physiological artifacts, where there exists any dependency between the artifact and the class

label, lead to superior result of the proposed representation in comparison with representation in single instance framework.

Experimental evidences confirm the significant superiority of this representation, especially in the cases the data are not linearly separable in single instance framework. The superiority of the proposed representation in comparison with competitors comes from considering the non-stationarity of the EEG signal using MI framework without the need for explicit induction of EEG segments, using covariance matrix for describing segments, determining the boundaries of the segments adaptively and using the benefits of analysis of covariance matrices in Riemannian framework.

Describing a subject using a bag of patterns and the existence of similar patterns within a class and dissimilar patterns between the classes lead to high within class similarity and low between class similarity in describing the subjects using a similarity matrix.

This representation also confirms the existence of some discriminative patterns between Iranian/ Swiss normal subjects, depression/normal subjects, ADHD/BMD subjects, and Schizophrenia/ normal subjects.

Using MI framework for EEG representation in outdoor problems would be investigated as a future work and the efficiency of MI kernel would be investigated for representing the dataset in this category of problems.

REFERENCES

[1] American Psychiatric Association. *Diagnostic and statistical manual of mental disorders (DSM-5®)*. American Psychiatric Pub, 2013.

[2] Acharya, U. Rajendra, Vidya K. Sudarshan, Hojjat Adeli, Jayasree Santhosh, Joel EW Koh, and Amir Adeli. "Computer-aided diagnosis of depression using EEG signals." *European neurology* 73, no. 5-6 (2015): 329-336.

[3] Iosifescu, Dan V., Scott Greenwald, Philip Devlin, David Mischoulon, John W. Denninger, Jonathan E. Alpert, and Maurizio Fava. "Frontal EEG predictors of treatment outcome in major depressive disorder." *European Neuropsychopharmacology* 19, no. 11 (2009): 772-777.

[4] Salustri, Carlo, Franca Tecchio, Filippo Zappasodi, Gabriela Bevacqua, Mauro Fontana, Matilde Ercolani, Daniele Milazzo, Rosanna Squitti, and Paolo Maria Rossini. "Cortical excitability and rest activity properties in patients with depression." *Journal of psychiatry & neuroscience* 32, no. 4 (2007): 259.

[5] Fingelkurts, Alexander A., Andrew A. Fingelkurts, Heikki Rytsälä, Kirsi Suominen, Erkki Isometsä, and Seppo Kähkönen. "Composition of brain oscillations in ongoing EEG during major depression disorder." *Neuroscience research* 56, no. 2 (2006): 133-144.

[6] Adeli, Hojjat, Ziqin Zhou, and Nahid Dadmehr. "Analysis of EEG records in an epileptic patient using wavelet transform." *Journal of neuroscience methods* 123, no. 1 (2003): 69-87.

[7] Acharya, U. Rajendra, Chua Kuang Chua, Teik-Cheng Lim, Dorithy, and Jasjit S. Suri. "Automatic identification of epileptic EEG signals using nonlinear parameters." *Journal of Mechanics in Medicine and Biology* 9, no. 04 (2009): 539-553.

[8] Chua, Chua Kuang, Vinod Chandran, Rajendra U. Acharya, and Lim Choo Min. "Cardiac health diagnosis using higher order spectra and support vector machine." *The open medical informatics journal* 3 (2009): 1.

[9] Martis, Roshan Joy, U. Rajendra Acharya, Jen Hong Tan, Andrea Petznick, Ratna Yanti, Chua Kuang Chua, EY Kwee Ng, and Louis Tong. "Application of empirical mode decomposition (EMD) for automated detection of epilepsy using EEG signals." *International journal of neural systems* 22, no. 06 (2012): 1250027.

[10] Yuan, Qi, Weidong Zhou, Shasha Yuan, Xueli Li, Jiwen Wang, and Guijuan Jia. "Epileptic EEG classification based on kernel sparse representation." *International journal of neural systems* 24, no. 04 (2014): 1450015.

[11] Adeli, Hojjat, Samanwoy Ghosh-Dastidar, and Nahid Dadmehr. "A spatio-temporal wavelet-chaos methodology for EEG-based diagnosis of Alzheimer's disease." *Neuroscience letters* 444, no. 2 (2008): 190-194.

[12] Boostani, Reza, Khadijeh Sadatnezhad, and Malihe Sabeti. "An efficient classifier to diagnose of schizophrenia based on the EEG signals." Expert Systems with Applications 36, no. 3 (2009): 6492-6499.

[13] Puthankattil, Subha D., and Paul K. Joseph. "Classification of EEG signals in normal and depression conditions by ANN using RWE and signal entropy." *Journal of Mechanics in Medicine and Biology* 12, no. 04 (2012): 1240019.

[14] Faust, Oliver, Peng Chuan Alvin Ang, Subha D. Puthankattil, and Paul K. Joseph. "Depression diagnosis support system based on EEG signal entropies." *Journal of mechanics in medicine and biology* 14, no. 03 (2014): 1450035.

[15] Hosseinifard, Behshad, Mohammad Hassan Moradi, and Reza Rostami. "Classifying depression patients and normal subjects using machine learning techniques and nonlinear features from EEG signal." *Computer methods and programs in biomedicine* 109, no. 3 (2013): 339-345.

[16] Congedo, Marco, Alexandre Barachant, and Anton Andreev. "A new generation of brain-computer interface based on riemannian geometry." *arXiv preprint arXiv*:1310.8115 (2013).

[17]   Pennec, Xavier, Pierre Fillard, and Nicholas Ayache. "A Riemannian framework for tensor computing." *International Journal of computer vision* 66, no. 1 (2006): 41-66.

[18]   Kaplan, Alexander Ya, Andrew A. Fingelkurts, Alexander A. Fingelkurts, Sergei V. Borisov, and Boris S. Darkhovsky. "Nonstationary nature of the brain activity as revealed by EEG/MEG: methodological, practical and conceptual challenges." *Signal processing* 85, no. 11 (2005): 2190-2212.

[19]   Sanei, Saeid, and Jonathon A. Chambers. "EEG signal processing." (2007).

[20]   Dietterich, Thomas G., Richard H. Lathrop, and Tomás Lozano-Pérez. "Solving the multiple instance problem with axis-parallel rectangles." *Artificial intelligence* 89, no. 1-2 (1997): 31-71.

[21]   Lee, John M. Riemannian manifolds: an introduction to curvature. Vol. 176. *Springer Science & Business Media*, 2006.

[22]   Jost, Jürgen, and Jèurgen Jost. Riemannian geometry and geometric analysis. Vol. 42005. *Berlin: Springer*, 2008.

[23]   Lin, Tong, and Hongbin Zha. "Riemannian manifold learning." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30, no. 5 (2008): 796-809.

[24]   Arsigny, Vincent, Pierre Fillard, Xavier Pennec, and Nicholas Ayache. "Log-Euclidean metrics for fast and simple calculus on diffusion tensors." *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine* 56, no. 2 (2006): 411-421.

[25]   Dryden, Ian L., Alexey Koloydenko, and Diwei Zhou. "Non-Euclidean statistics for covariance matrices, with applications to diffusion tensor imaging." *The Annals of Applied Statistics* 3, no. 3 (2009): 1102-1123.

[26]   Andrews, Stuart, Ioannis Tsochantaridis, and Thomas Hofmann. "Support vector machines for multiple-instance learning." *In Advances in neural information processing systems*, pp. 577-584. 2003.

[27]   Ramon, Jan, and Luc De Raedt. "Multi instance neural networks." In Proceedings of the *ICML-2000 workshop on attribute-value and relational learning*, pp. 53-60. 2000.

[28]   Blockeel, Hendrik, David Page, and Ashwin Srinivasan. "Multi-instance tree learning." *In Proceedings of the 22nd international conference on Machine learning*, pp. 57-64. ACM, 2005.

[29]   Auer, Peter, and Ronald Ortner. "A boosting approach to multiple instance learning." *In European Conference on Machine Learning*, pp. 63-74. Springer, Berlin, Heidelberg, 2004.

[30]   Barachant, Alexandre, Anton Andreev, and Marco Congedo. "The Riemannian Potato: an automatic and adaptive artifact detection method for online experiments using Riemannian geometry." *In TOBI Workshop lV*, pp. 19-20. 2013.

[31]   Jayasumana, Sadeep, Richard Hartley, Mathieu Salzmann, Hongdong Li, and Mehrtash Harandi. "Kernel methods on Riemannian manifolds with Gaussian RBF kernels." *arXiv preprint arXiv*:1412.0265 (2014).

[32]   Cox, Trevor F., and Michael AA Cox. Multidimensional scaling. *Chapman and hall/CRC*, 2000.

[33]   Ayd, Frank J., ed. Lexicon of psychiatry, neurology, and the neurosciences. *Lippincott Williams & Wilkins*, 2000.

[34]   Pekalska, Elzbieta, Pavel Paclik, and Robert PW Duin. "A generalized kernel approach to dissimilarity-based classification." *Journal of machine learning research* 2, no. Dec (2001): 175-211.

[35]   Pękalska, Elżbieta, and Robert PW Duin. "Dissimilarity representations allow for building good classifiers." *Pattern Recognition Letters* 23, no. 8 (2002): 943-956.

[36]   Sadatnezhad, Khadijeh, Reza Boostani, and Ahmad Ghanizadeh. "Proposing an adaptive mutation to improve XCSF performance to classify ADHD and BMD patients." *Journal of neural engineering* 7, no. 6 (2010): 066006.

[37]   Kazemi, Reza, Reza Rostami, Sanaz Khomami, Jiri Horacek, Martin Brunovsky, Tomas Novak, and Paul B. Fitzgerald. "Electrophysiological correlates of bilateral and unilateral repetitive transcranial magnetic stimulation in patients with bipolar depression." *Psychiatry research* 240 (2016): 364-375.

[38]   WHO, "Official WHO health days," in World Health Organization, World Health Organization, 2016. [Online]. Available: http://www.who.int/mediacentre/events/annual/world_suicide_prevention_day/en. Accessed: Nov. 12, 2016.

[39]   Higuchi, Tomoyuki. "Approach to an irregular time series on the basis of the fractal theory." Physica D: Nonlinear Phenomena 31, no. 2 (1988): 277-283.

[40]   Stoica, Petre, and Randolph L. Moses. *Introduction to spectral analysis*. Vol. 1. Upper Saddle River, NJ: Prentice hall, 1997.

[41]   Sabeti, Malihe, Serajeddin Katebi, and Reza Boostani. "Entropy and complexity measures for EEG signal classification of schizophrenic and control participants." *Artificial intelligence in medicine* 47, no. 3 (2009): 263-274.

[42]   Deivanayagi, S., M. Manivannan, and Peter Fernandez. "Spectral analysis of EEG signals during hypnosis." *International Journal of Systemics, Cybernetics and Informatics* 4 (2007): 75-80.

[43]   Adeli, Hojjat, Samanwoy Ghosh-Dastidar, and Nahid Dadmehr. "A wavelet-chaos methodology for analysis of EEGs and EEG subbands to detect seizure and epilepsy." *IEEE Transactions on Biomedical Engineering* 54, no. 2 (2007): 205-211.

[44]   Suykens, Johan AK, and Joos Vandewalle. "Least squares support vector machine classifiers." *Neural processing letters* 9, no. 3 (1999): 293-300.

[45]   Chang, Chih-Chung, and Chih-Jen Lin. "LIBSVM: a library for support vector machines." *ACM transactions on intelligent systems and technology* (TIST) 2, no. 3 (2011): 27.

[46]   Bhatia, Rajendra, and John Holbrook. "Riemannian geometry and matrix geometric means." *Linear algebra and its applications* 413, no. 2-3 (2006): 594-618.

[47]   Foulds, James, and Eibe Frank. "A review of multi-instance learning assumptions." *The Knowledge Engineering Review* 25, no. 1 (2010): 1-25.

[48]   Feragen, Aasa, and Søren Hauberg. "Open Problem: Kernel methods on manifolds and metric spaces. What is the probability of a positive definite geodesic exponential kernel?." *In Conference on Learning Theory*, pp. 1647-1650. 2016.

[49]   Wu, Gang, Edward Y. Chang, and Zhihua Zhang. "An analysis of transformation on non-positive semidefinite similarity matrix for kernel machines." *In Proceedings of the 22nd International Conference on Machine Learning*, vol. 8. 2005.

[50]   Gärtner, Thomas, Peter A. Flach, Adam Kowalczyk, and Alexander J. Smola. "Multi-instance kernels." *In ICML*, vol. 2, pp. 179-186. 2002.

[51]   Zhang, Tao, and Wanzhong Chen. "LMD based features for the automatic seizure detection of EEG signals using SVM." *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 25, no. 8 (2017): 1100-1108.

[52]   Garrett, Deon, David A. Peterson, Charles W. Anderson, and Michael H. Thaut. "Comparison of linear, nonlinear, and feature selection methods for EEG signal classification." *IEEE Transactions on neural systems and rehabilitation engineering* 11, no. 2 (2003): 141-144.

[53]   Lin, Yuan-Pin, Chi-Hong Wang, Tzyy-Ping Jung, Tien-Lin Wu, Shyh-Kang Jeng, Jeng-Ren Duann, and Jyh-Horng Chen. "EEG-based emotion recognition in music listening." *IEEE Transactions on Biomedical Engineering* 57, no. 7 (2010): 1798-1806.

[54]   Makeig, Scott, Christian Kothe, Tim Mullen, Nima Bigdely-Shamlo, Zhilin Zhang, and Kenneth Kreutz-Delgado. "Evolving signal processing for brain–computer interfaces." *Proceedings of the IEEE* 100, no. Special Centennial Issue (2012): 1567-1584.

[55]   Vaid, Swati, Preeti Singh, and Chamandeep Kaur. "EEG signal analysis for BCI interface: A review." In Advanced Computing & Communication Technologies (ACCT), 2015 *Fifth International Conference* on, pp. 143-147. IEEE, 2015.

[56]   Subha, D. Puthankattil, Paul K. Joseph, Rajendra Acharya, and Choo Min Lim. "EEG signal analysis: a survey." *Journal of medical systems* 34, no. 2 (2010): 195-212.

[57]   Chong, Derek J., and Lawrence J. Hirsch. "Which EEG patterns warrant treatment in the critically ill? Reviewing the evidence for treatment of periodic epileptiform discharges and related patterns." *Journal of Clinical Neurophysiology* 22, no. 2 (2005): 79-91.

[58]   Ames, Daniel L., and Susan T. Fiske. "Cultural neuroscience." *Asian journal of social psychology* 13, no. 2 (2010): 72-82.

[59]   Chiao, Joan Y., and Nalini Ambady. "Cultural neuroscience: Parsing universality and diversity across levels of analysis." (2007).

[60]   Adeli, Hojjat, and Samanwoy Ghosh-Dastidar. Automated EEG-based diagnosis of neurological disorders: *Inventing the future of neurology*. CRC Press, 2010.

[61]   Jayaram, Vinay, Morteza Alamgir, Yasemin Altun, Bernhard Scholkopf, and Moritz Grosse-Wentrup. "Transfer learning in brain-computer interfaces." *IEEE Computational Intelligence Magazine* 11, no. 1 (2016): 20-31.

[62]   Stevens, Janice R. "Disturbances of ocular movements and blinking in schizophrenia." *Journal of Neurology, Neurosurgery & Psychiatry* 41, no. 11 (1978): 1024-1030.

[63]   Vapnik, Vladimir. The nature of statistical learning theory. *Springer science & business media*, 2013.

[64]   Barachant, Alexandre, Stéphane Bonnet, Marco Congedo, and Christian Jutten. "Riemannian geometry applied to BCI classification." In *International Conference on Latent Variable Analysis and Signal Separation*, pp. 629-636. Springer, Berlin, Heidelberg, 2010.

[65]   Kamida, Akira, Kenta Shimabayashi, Masayoshi Oguri, Toshihiro Takamori, Naoyuki Ueda, Yuki Koyanagi, Naoko Sannomiya et al. "EEG Power Spectrum Analysis in children with ADHD.*" Yonago acta med*ica 59, no. 2 (2016): 169.

[66]   Sadatnejad, Khadijeh, and Saeed Shiry Ghidary. "Kernel learning over the manifold of symmetric positive definite matrices for dimensionality reduction in a BCI application." Neurocomputing 179 (2016): 152-160.

[67]   Vapnik, Vladimir. The nature of statistical learning theory. Springer science & business media, 2013.

[68]   Barachant, Alexandre, Stéphane Bonnet, Marco Congedo, and Christian Jutten. "Classification of covariance matrices using a Riemannian-based kernel for BCI applications." Neurocomputing 112 (2013): 172-178.