A Kernel-based Nonlinear Manifold Learning for EEG Channel Selection with Application to Alzheimer's Disease

S. Rajintha. A. S. Gunawardena,[1] Ptolemaios G. Sarrigiannis,[2] Daniel J. Blackburn,[3] and Fei He[1,*]

[1] *Centre for Computational Science and Mathematical Modelling, Coventry University, Coventry CV1 2JH, UK*

[2] *Department of Neurophysiology, Royal Devon and Exeter NHS Foundation Trust, Exeter, EX2 5DW, UK*

[3] *Department of Neuroscience, University of Sheffield, Sheffield, S10 2HQ, UK*

This paper introduces a novel EEG channel selection method to determine which channel inter-relationships provide the best classification accuracy between a group of patients with Alzheimer's disease (AD) and a cohort of age matched healthy controls (HC). Thereby, determine which inter-relationships are more important for the in-depth dynamical analysis to further understand how neurodegenerative diseases such as AD affects global and local brain dynamics. The channel selection methodology uses kernel-based nonlinear manifold learning via Isomap and Gaussian Process Latent Variable Model (Isomap-GPLVM). The Isomap-GPLVM method is employed to learn both the spatial and temporal local similarities and global dissimilarities present within the EEG data structures. The resulting kernel (dis)similarity matrix is used as a measure of synchrony between EEG channels. Based on this information, channel-specific linear Support Vector Machine (SVM) classification is then used to determine which spatio-temporal channel inter-relationships are more important for in-depth dynamical analysis. In this work, the analysis of EEG data from HCs and AD patients is presented as a case study. Our analysis shows that inter-relationships between channels in the fronto-parietal region and the rest are better at differentiating between AD and HC groups.

***Key words*: Alzheimer's disease, kernel matrix, EEG, channel selection, manifold learning, machine learning**

I. INTRODUCTION

The electroencephalogram (EEG), even though recorded at the scalp level, reflects the grossly summed currents of the electrical fields generated by the neural activity in the cortical neural circuits. Through the EEG, the behaviour and integrity of underlying neural circuits can thus be indirectly studied [1]. The analysis of hidden structures within EEG data is therefore important and has gained considerable attention [2].

Numerous EEG studies have previously revealed the importance of nonlinear methods for the diagnosis of neurological diseases [2]. Among these, in-depth dynamical analysis, such as the analysis of linear and nonlinear dynamic inter-relationships between EEG channels, causality and cross-frequency coupling analysis takes special precedence [2–4]. However, these methods incur high computational costs and or accuracy issues because of dimensionality. Consequently, in practice, the selection of important EEG channels is vital [5] before conducting an in-depth dynamical analysis of high dimensional EEG data.

In EEG channel selection techniques, usually features from the channels are extracted and important channels are selected via feature selection. These feature selection methods can be categorised into the following three groups [5]. a) *Filtering methods*: Independent evaluation criteria, such as distance measures, entropy, mutual information, are used for the purpose of channel selection. Filtering methods are good at eliminating irrelevant and redundant features. b) *Wrapper methods*: Subsets of features are generated based on a method of choice. Each subset is evaluated using a classification algorithm to select a subset of channels. These are based on greedy search algorithms aiming to find the best possible combination of features. c) *Embedded Methods*: These techniques do the feature selection and the classification simultaneously. For example LASSO based feature selection, logistic regression and decision-tree are techniques that come under embedded methods. Ranking of features can be easily done using embedded methods.

The brain is a complex nonlinear system, and the nonlinear nature in brain dynamics can be readily detected considering the inter-relationships between channels [2]. On the other hand, the analysis of nonlinear structures locally and globally within EEG data can also be used for channel selection purposes [6], which is often implicitly considered in many EEG channel selection approaches discussed above. In this paper, an EEG channel selection method based on spatio-temporal linear and nonlinear EEG inter-relationships is presented. This is achieved by using pairwise representation [7] via kernel-based manifold learning (dimensionality reduction). The features based on spatio-temporal pairwise distance measures are assessed using linear SVM classification to evaluate which features are better at differentiating between Alzheimer's patients and healthy controls. Furthermore, the linear SVM classification weights are used to rank the pairwise features. Therefore, the method presented is of a hybrid form [5] of the aforementioned three categories of feature selection methods.

* Correspondence to: fei.he@coventry.ac.uk

Pairwise representation has gained attention in various fields such as bioinformatics, neuroscience, cognitive psychology, social sciences, text and web mining [8–13], as it can be used to model highly nonlinear data and reveals many hidden structures within the data [7]. In pairwise representation, various distance measures relating to the concepts of (dis)similarity within the data are considered. Similarity or dissimilarity between two objects, in general, express the degree to which the two objects are respectively alike/related or different/distinct. Euclidean and Manhattan distances are examples of distance measures. These can be used for the comparison of (dis)similarities of objects in the space of measured variables [14]. Hence (dis)similarity can be used as a measure of synchrony and linear/nonlinear inter-relationships [13, 15]. In neuroscience, brain functional connectivity is characterised using different measures of (dis)similarity [13]. Thus, regardless of structural connectivity, brain regions functional connected under one measure does not necessarily mean the same or could be even disconnected in another [13]. Dauwels *et al.* [6] showed that various (dis)similarity measures can be correlated to each other, such as in the application of different (dis)similarity measures for the early diagnosis of AD. Hence, these correlated measures can often be grouped, and one measure from each group is sufficient to analyse the structures within data [6].

In this study, a spatio-temporal pairwise representation is applied using kernel-based nonlinear manifold learning and the focus is on EEG channel selection. The motivation behind utilising a kernel-based method is based on the fact that positive semi-definite (PSD) kernels are a generalisation of distance measures [16]. In this work, the kernel matrix is evaluated using nonlinear manifold learning via Gaussian Process Latent Variable Model (GPLVM) [17]. Robust kernel Isomap [18] is used as an initialisation method for GPLVM (Isomap-GPLVM). This enables the learning of both local similarities and global dissimilarities within the EEG data and embedding it to the reduced-dimension manifold (latent space) [19]. Furthermore, since dimensionality reduction is used to reduce the temporal dimension, temporal structures within the data are taken into account. This is not possible in the direct use of popular distance measures, as these methods only consider the spatial aspect of the data and not the temporal aspect unless a delay embedding is used.

In summary, the kernel matrix evaluated, using Isomap-GPLVM can be regarded as a more objective (dis)similarity measure containing information on both linear and nonlinear spatio-temporal EEG inter-relationships. Hence, it can be a better alternative to using various (dis)similarity measures each from the respective groups as mentioned earlier.

This paper is organised as follows; section III A briefly introduces GPLVM. Section II provides specifics of the EEG data used and the pre-processing steps. Section III discusses the manifold learning methodology via Isomap and GPLVM to evaluate the kernel-based (dis)similarity matrix for the classification of EEG data of a group of AD patients from an age-matched healthy control cohort. Section IV discusses the results obtained followed by the concluding remarks and the future work in section V.

II. DATA

Most EEG data from AD and Healthy Controls (HC) have been previously used in a recent study by Blackburn et al. in [20]. Task-free EEG recordings that require minimal cooperation of AD patients were used; typically this group of patients can have difficulty engaging with and following cognitive tasks. The details of data acquisition and other data information are provided in [20]. In this work, we include a total of 20 AD cases and 20 age and gender matched healthy controls (less than 70 years) all selected based on previously described clinical and radiological diagnostic criteria [20].

A. EEG Data

A modified 10/10 overlapping a 10/20 international system of electrode placement method was utilised to acquire the EEG recordings. Subsequently, the following 23 bipolar channel pairs were produced for the analysis; F8–F4, F7–F3, F4–C4, F3–C3, F4–FZ, FZ–CZ, F3–FZ, T4–C4, T3–C3, C4–CZ, C3–CZ, CZ–PZ, C4–P4, C3–P3, T4–T6, T3–T5, P4–PZ, P3–PZ, T6–O2, T5–O1, P4–O2, P3–O1 and O2–O1. During the EEG recording, the participants were encouraged not to think about anything specific. As mentioned in [20] all EEGs were reviewed by an experienced neurophysiologist, thus from resting-state EEG recordings, 12-second artefact-free epochs under Eye-Open (EO) and Eye-Closed (EC) conditions were isolated. In this study, three of these 12-seconds epochs of the EO EEG data are used.

B. Pre-processing Tasks

In this study, since the high-dimensional temporal structures of the multi-channel EEG are examined, the use of filters would pose a major issue due to the distortions induced [21]. Hence, the Fast Fourier Transform (FFT) was implemented to remove unwanted frequency components. Thereafter, the inverse-FFT was used to reconstruct the time-domain signal without the unwanted frequency components. The analysis in this work was conducted using EEG frequencies between 2 to 100Hz, avoiding eye-blinks and slow movement artefacts and EMG related activity. After removing the unwanted frequency components, the time-domain signal was then downsampled to 200Hz.

III. MEATHODS

The main purpose of using the kernel (dis)similarity matrix, $\mathbf{K}$ in eq. (1), is to select channels that can well distinguish patients and control groups for more in-depth dynamical analysis. Thus, in this study, channels that can differentiate well between AD and HC participants are revealed by examining the linear and nonlinear channel inter-relationships.

A. Gaussian Process Latent Variable Model (GPLVM)

A GPLVM [17] assumes that the observed dataset $\mathbf{Y}$ is generated from a lower dimensional data $\mathbf{X}$. Therefore, the GPLVM framework learns the mapping of a high-dimensional dataset $\mathbf{Y} \in \mathbb{R}^{N\times D}$ and the corresponding low dimensional latent space $\mathbf{X} \in \mathbb{R}^{N\times Q}$, $Q < D$, using a Gaussian Process (GP) mapping [22] from $\mathbf{X} \rightarrow \mathbf{Y}$. Where $\mathbf{Y} = [\mathbf{y}_1, \cdots, \mathbf{y}_N]^T$, $\mathbf{y}_i \in \mathbb{R}^{1\times D}$ and $\mathbf{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_N]^T$, $\mathbf{x}_i \in \mathbb{R}^{1\times Q}$. The likelihood of the data given the latent positions is $P(\mathbf{Y}|\mathbf{X},\theta) = \exp\left(\frac{1}{2}\operatorname{tr}\left(\mathbf{K}^{-1}\mathbf{Y}\mathbf{Y}^T\right)\right) \sqrt[-2]{2\pi^{D\times N}|\mathbf{K}|}$, where $\mathbf{K} \in \mathbb{R}^{N\times N}$,

$$\mathbf{K}(\mathbf{X},\mathbf{X}) = \begin{bmatrix} k(\mathbf{x}_1,\mathbf{x}_1) & \cdots & k(\mathbf{x}_1,\mathbf{x}_N) \\ \vdots & \ddots & \vdots \\ k(\mathbf{x}_N,\mathbf{x}_1) & \cdots & k(\mathbf{x}_N,\mathbf{x}_N) \end{bmatrix} \tag{1}$$

is a positive semi-definite matrix, where $k(\mathbf{x}_i,\mathbf{x}_j)$ is the kernel/covariance function with a set of hyper-parameters θ. The use of a kernel function that allows for nonlinear functional mappings from $\mathbf{X}$ to $\mathbf{Y}$ provides a probabilistic nonlinear latent variable model. Therefore GPLVM is a nonlinear manifold learning technique. In GPLVM, the maximising of the marginal log-likelihood, $\log P(\mathbf{Y}|\mathbf{X},\theta)$, is done with respect to both $\mathbf{X}$ and θ, hence optimal estimates $\hat{\mathbf{X}}$ and $\hat{\theta}$ are obtained jointly. It should be noted that when using the RBF kernel with GPLVM, the length-scale hyper-parameter influence how much similarity or dissimilarity information from the data space is embedded within the latent space [19].

B. Kernel-Based nonlinear Manifold Learning of High-dimensional EEG Data Using Isomap-GPLVM

The GPLVM is a global dissimilarity preserving mapping [19]. To also preserve the local similarities, we employ another manifold learning technique the robust kernel Isomap [18], which is used as an initial condition for the estimated latent space $\hat{\mathbf{X}}$ in the GPLVM optimisations (see section III A). In such a way, both local similarities and global dissimilarities are preserved within the reduced-dimensional data. To demonstrate this, FIG. 1 illustrates Isomap-GPLVM operating on the swiss-roll dataset.

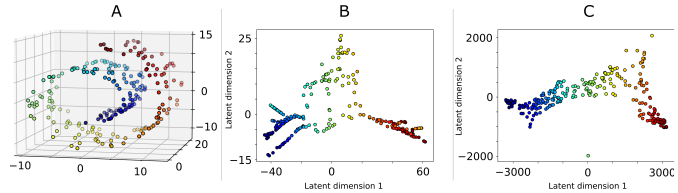

FIG. 1. **Isomap, as an initialisation method for GPLVM, enables preservation of both local similarities and global dissimilarities within the data. A).** Original Swiss Roll data. Data sampled from the original high dimensional manifold. **B).** Isomap projection of a lower-dimensional manifold. Local similarities in the data structure from *Data space* to *Latent space* mapping. **C).** GPLVM optimisation of initial Isomap lower-dimensional manifold. Global dissimilarities in the data structure from *Latent space* to *Data space* mapping.

Robust kernel Isomap approximates the geodesic distance (along the surface of the high dimensional manifold) to project the data into the latent space (i.e. lower-dimensional manifold) while preserving topological stability and providing a method for eliminating critical outliers [18]. The data points are projected according to how close the points are along the high dimensional manifold surface (i.e. local similarities), as shown in FIG. 1B. GPLVM uses the Isomap projection of the data as an initial condition to the latent space. GPLVM then operates on this latent space and optimises it according to how far away the data points are, along the high dimensional manifold surface (i.e. global dissimilarities), in the original data space (FIG. 1C). The resulting kernel matrix $\mathbf{K}(\mathbf{X},\mathbf{X})$, from the optimised GPLVM, governs the latent space to data space mapping (see section III A). This kernel matrix will thus contain the (dis)similarity information within the original data, and named a kernel (dis)similarity matrix.

Isomap-GPLVM is used to analyse the EEG data by reducing the temporal dimension. Thus, the manifold learning method will look into the local and global (dis)similarities along the spatio-temporal manifold of the high dimensional EEG data. Hence the resulting kernel matrix, $\mathbf{K}(\mathbf{X},\mathbf{X}) \in \mathbb{R}^{23\times 23}$, will therefore contain the spatio-temporal (dis)similarity information between the 23 respective EEG channels.

Isomap-GPLVM is applied individually to each participant EEG data, thus obtaining a kernel (dis)similarity matrix for each AD and HC participant. Therefore, following the definition of the data space $\mathbf{Y} \in \mathbb{R}^{N\times D}$ in section III A, $N = 23$ (23 EEG channels, section II A) and D is the temporal dimension that is to be reduced. The respective latent space of each AD and HC participant will be $\mathbf{X} \in \mathbb{R}^{N\times Q}$, where $Q = 7$. The Radial Basis Function (RBF) was used as the kernel function. There are two reasons for setting the latent dimension $Q = 7$: i) to achieve a nearly 100% recovery accuracy in the GPLVM mapping to the original data space $\mathbf{X}$ to

$\mathbf{Y}$; and ii) to achieve the best results in distinguishing between AD and HC. Before applying nonlinear manifold learning, each EEG channel of all participants was de-meaned and normalised such that the absolute maximum value attained by the respective channels is 1.

C. Kernel (dis)similarity matrix based channel selection

Using kernel (dis)similarity measures, brain regions (EEG channels) that have higher *synchrony changes* between patient and control groups are determined. A kernel (dis)similarity matrix for each HC and AD participant is evaluated using Isomap-GPLVM. Each row in the matrix contains a measure of (dis)similarity against other respective channels. Thus, channel-specific rows from all the participant-specific kernel (dis)similarity matrices are grouped into 23 different feature spaces (23 EEG channels, section II A). Each channel-specific feature space will contain 22 features, i.e. (dis)similarity measures against other respective channels. Hence, 23 different linear SVM classifiers are used to assess how well the differentiation between HC and AD groups can be made and to determine which channels are more important. Therefore, the selection of channels is made based on the *changes* in the linear and nonlinear spatio-temporal EEG channel inter-relationships between AD and HC groups. FIG. 2 illustrates this methodology.

Linear SVM adds weights to the features according to importance. Thus, the normalised absolute linear SVM coefficients (weights on the 22 features) can be regarded as how much the inter-relationships between an individual channel and another affect the differentiation between HC and AD groups. Thus, a measure of relative importance of the changes in synchrony among two channels between AD and HC groups.

Kernel (dis)similarity matrices of the EEG data are produced for each AD and HC participant using Isomap-GPLVM. EEG data from the *first* 12-second epoch (out of the total three, see section II A) of each case was used for training and testing the linear SVM classifiers. Out of the 20 AD patients and 20 HC subjects, the kernel (dis)similarity matrices of 10 AD and 10 HC subjects were used to train the 23 linear SVM classifiers. The kernel (dis)similarity matrices of the remaining 10 AD and 10 HC subjects, were used to test the prediction accuracies of the trained classifiers.

It should be re-emphasised that the focus of this work is on channel selection, linear SVM classification is only used to determine which channel inter-relationships exhibit the highest *synchrony changes* between patient and control groups. This is achieved using spatio-temporal (dis)similarities between channels as a measure of synchrony. The proposed methodology can also be further developed as a diagnostic tool for the classification between HC and AD patients provided a sufficiently large number of participants.

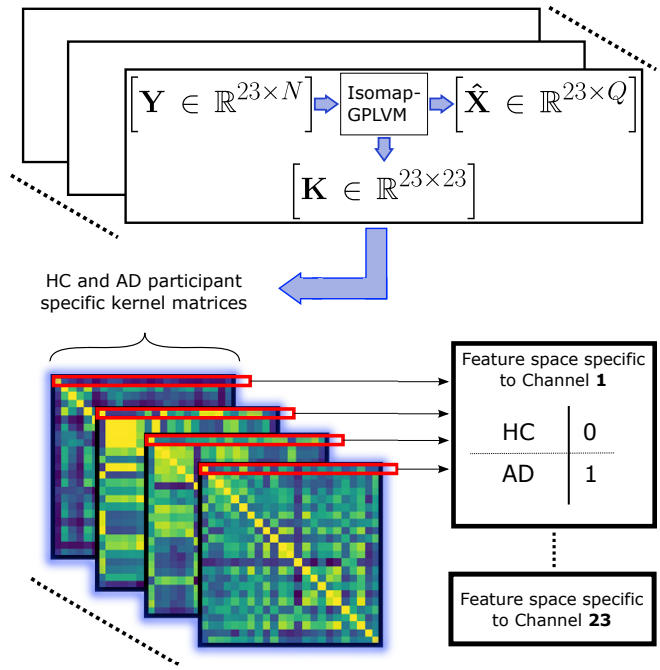

FIG. 2. **Kernel based (dis)similarity analysis methodology of AD and HC EEG data**. Participant specific kernel (dis)similarity matrices are evaluated using Isomap-GPLVM. All channel-specific rows from these matrices are grouped into channel-specific feature spaces. Each feature space has two classes, AD and HC. The classification is binary thus AD is denoted as 1 and HC as 0. Individual linear SVM classifiers are used on each feature space to determine which channels and their respective inter-relationships with other channels are better at distinguishing between AD and HC groups.

D. GPLVM Kernel hyper-parameter initial condition determination

The (dis)similarity measures from the kernel matrix depend considerably on the initial conditions of the RBF kernel hyper-parameters θ. The initial conditions are determined using a grid search method. The best set of initial conditions for θ are chosen based on how well all the classifiers perform on the training feature set, the prediction accuracy on the testing feature set (section III C) and the performance on an additional validation set. The validation set is composed of the kernel (dis)similarity matrices from the 2nd and 3rd epochs of all the participants (section II A). This was done for several latent dimensions, i.e. $Q = [5, 8]$ and $Q = 7$ was chosen, producing the most favourable classification accuracies. Thus, the initial conditions determined in this manner leads the manifold learning method to produce a kernel matrix. Its (dis)similarity measures are relatively more specific for the differentiation of HC and AD EEG data.

Ideally, a new set of EEG data from both HC and AD groups should be used for validation in order to attest the repeatability. However, due to the limited number

of participants, the 2nd and 3rd epochs from the same participants are used for validation (section II A). This can be deemed appropriate to check for repeatability as the brain is an extremely complex and stochastic system. Furthermore, considering the time interval between epochs and the length of Eye-Open and Eye-Close time periods as mentioned in [20], the use of the other epochs for validation can be justified.

IV. RESULTS AND DISCUSSION

The Isomap-GPLVM method introduced in section III was applied to the EO EEG data (section II A). The best set of initial conditions for the hyper-parameters, θ, was determined using the three 12-second epochs in the manner explained in section III D. From the kernel (dis)similarity matrices evaluated the channel inter-relationships that were able to differentiate between AD and HC groups well are presented in this section. This study is conducted using 20 HC and 20 AD participants. Each participant has three sets of 12-second epochs except for one AD participant. Only one epoch from this participant is available.

FIG. 3 illustrates the kernel (dis)similarity matrices across all the epochs of all the 20 HC (FIG. 3 HC) and 20 AD (FIG. 3 AD) participants used in this study. In FIG. 3, the red dotted lines separate the respective epochs of each case. Furthermore, the black dotted line across the 1st epoch of both groups (HC and AD) separates the participants used for training the classifiers and the remaining that are used for testing (section III C). As mentioned in section III D, the 2nd and 3rd epochs are used to validate the initial conditions for the hyper-parameters. From FIG. 3, it is evident that the HC group has varying and complex patterns; while the AD group (FIG. 3) more or less exhibits a common pattern. Furthermore, from the pattern observed, it can be noticed that the AD group has strong local inter-relationships but weak global inter-relationships. Note that inter-relationships in this context is quantified using local similarities and global dissimilarities (sections III B and III C). This leads to the conclusion that this is a reflection of the specific patterns of dysfunction that has been mentioned in the literature, in which AD EEG data exhibits less dynamic complexity compared to HC controls [2].

TABLE I shows the classification results of all the kernel (dis)similarity matrix based channel-specific feature spaces (section III C). The bold and grey highlighted channels are considered the best corresponding feature spaces (will be referred to as the best channels), because of the classification performance on the training set as well as the testing and validation sets. Therefore, the spatio-temporal inter-relationships associated with these channels have considerable differences between HC and AD groups. FIG. 4B shows these best channels on the electrode placement map used in this study. It is interesting to see that all the best channels are within the fronto-central and parietal regions of the headspace.

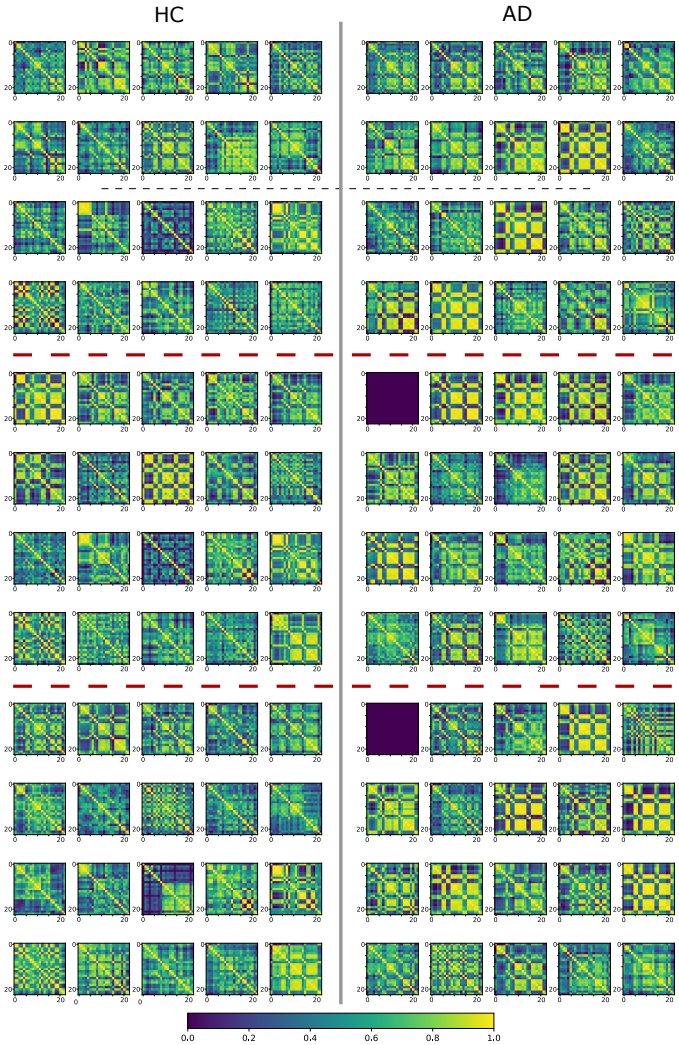

FIG. 3. **Participant specific RBF kernel (dis)similarity matrices of all epochs from all HC and AD participants for the best initial condition, eye-open data.** The red dotted lines separate each set of respective epochs. The black dotted line in the 1st epoch separates participants used for training and testing of the classifiers (section III C). 2nd and 3rd epochs are used to validate the set of initial conditions (section III D). It is evident from the validation set that in a general sense the AD group does exhibit a certain pattern of spatio-temporal (dis)similarity or synchrony. This leads to the conclusion that this is a reflection of the specific patterns of dysfunction the literature mentions about AD patients – less dynamic complexity. The EEG data for epochs 2 and 3 of one participant in the AD group was not available. This is indicated by the all-zero kernel (dis)similarity matrices.

Taking channel F3-C3 as an example and considering all epochs, FIG. 5 illustrates what channel specific spatio-temporal inter-relationships between F3-C3 and other channels are important in differentiating between HC and AD groups. FIG. 5A visualises the channel-specific feature space of F3-C3 in relation to HC and AD

TABLE I. [HC cases + AD cases] that are correctly classified

Channel specific feature space		Training set (out of 20)	Testing and Validation sets (out of 98)	Total (out of 118)
0	O1-O2	18	53	71
1	P4-O2	17	64	81
2	P3-O1	16	61	77
3	T5-O1	19	55	74
4	T6-O2	18	58	76
5	P3-PZ	18	69	87
6	P4-PZ	17	66	83
7	T3-T5	16	60	76
8	T4-T6	17	52	69
9	**C3-P3**	**20**	**75**	**95**
10	C4-P4	19	62	81
11	CZ-PZ	18	66	84
12	**C3-CZ**	**19**	**70**	**89**
13	C4-CZ	19	63	82
14	T3-C3	15	69	84
15	T4-C4	18	50	68
16	**F3-C3**	**19**	**75**	**94**
17	**FZ-CZ**	**18**	**73**	**91**
18	**F3-FZ**	**19**	**72**	**91**
19	F4-FZ	18	51	69
20	**F4-C4**	**18**	**74**	**92**
21	F7-F3	17	57	74
22	F8-F4	17	55	72

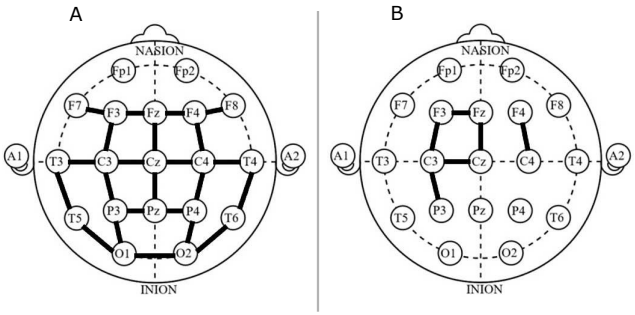

FIG. 4. **A). All the 23 channels, bipolar montage (section II A). B). The channels whose feature spaces are able to effectively differentiate between HV and AD cases.** The spatio-temporal inter-relationships specific to these channels have considerable differences between HC and AD groups.

groups. It illustrates, for all participants in the respective groups, the variations in the normalised (dis)similarity values (y-axes of FIG. 5A) between F3-C3 and the rest of the 22 channels (x-axes of FIG. 5A indicates the respective channel indexes, TABLE I). This shows, in general, how the (dis)similarities related to F3-C3 varies between HC and AD. FIG. 5B shows the normalised absolute linear SVM coefficients for the feature space of F3-C3. Each coefficient indicates the weight of each particular feature in making the classification decision; and more specifically, how much the (dis)similarity between the channel in question and the other channels influence the differentiation between HC and AD groups. The colours in the boxplots of FIG. 5A is a reflection of the relative channel importance as shown by the legend in FIG. 5C. The relative channel importance can be easily mapped into the electrode layout as shown in FIG. 5C to visualise and better comprehend the spatial layout of the important (dis)similarity measures (inter-relationships) in regard to a specific channel, e.g. F3-C3 at this instance.

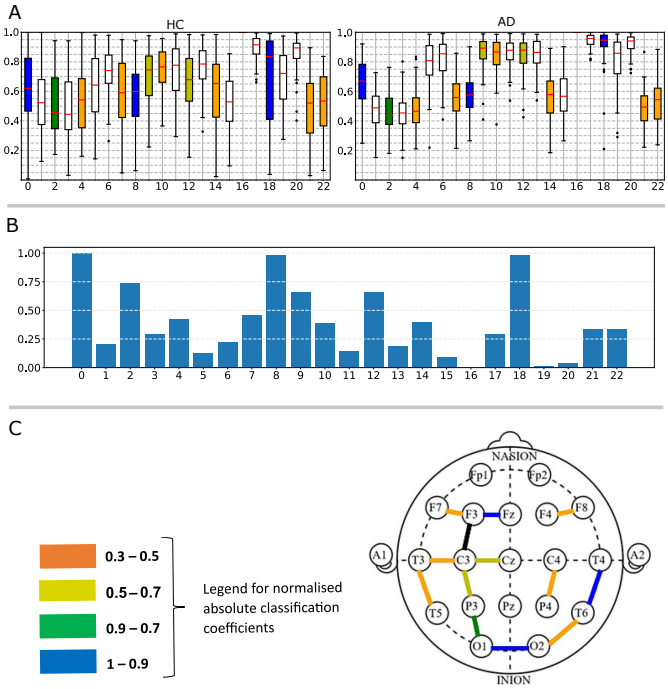

FIG. 5. **Channel specific information that can be obtained from the respective channel-specific row of all the kernel (dis)similarity matrices via Isomap-GPLVM of HC and AD groups. A).** Boxplots illustrating the variations in 'Linear + Nonlinear' inter-relationships, (dis)similarity, between channel F3-C3 and other channels for HC and AD groups respectively. Y-axes: normalised (dis)similarity values. X-axes: respective channel indexes (TABLE I). **B).** Normalised absolute linear SVM classification coefficients. Indicates how much the (dis)similarities between channel F3-C3 and other channels influence the differentiation between HC and AD. **C).** Normalised absolute linear SVM classification coefficients mapped into the electrode layout. Boxplots HC and AD, together comprise the channel-specific feature space of channel F3-C3. The colours in **C)**, signify the relative importance (normalised absolute classification coefficients) of the respective channel (dis)similarity measures against other channels.

It is clearly evident from FIG. 5A, the HC and AD group variations in the (dis)similarities between F3-C3 and other channels, the interquartile ranges and the extreme points of the AD group confined to a specific pattern. This pattern shows that the variations in channel (dis)similarities between F3-C3 and other channels have significantly reduced in the case of the AD group. Furthermore, the said pattern in the AD group variations in FIG. 5A is a reflection of the common kernel (dis)similarity matrix pattern seen in FIG. 3 AD.

FIG. 6 illustrates the feature spaces of two other channels that perform well in differentiating between HC and AD groups (see TABLE I): C3–P3 and F4–C4. As can be seen from TABLE I, channel T4–C4 is one of the

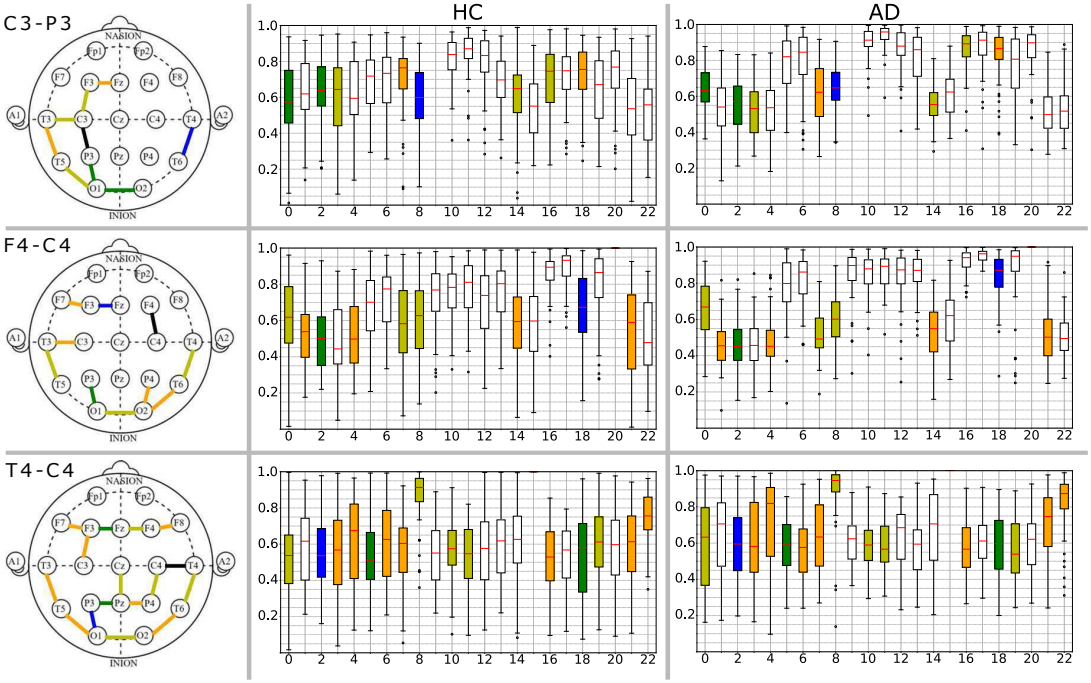

FIG. 6. **Difference between the best and the least performing channel feature spaces - An example using channel feature space of T4-C4**. The AD group variations in a poorly performing channel feature space do not exhibit significant changes in comparison to the respective HC group variations. The interquartile ranges and the extreme points of the AD group variations are not confined to form a specific pattern unlike the AD group variations in well-performing channel feature spaces. This leads to a conclusion that these channels and their respective channel inter-relationships have no significant changes in HC and AD groups.

least performing feature spaces. By comparing these feature spaces, there is a clear observable difference between the best and the least performing channel feature spaces. The AD group variations in the (dis)similarities of the least performing feature spaces do not exhibit significant differences in comparison to the respective HC group variations. Furthermore, the interquartile ranges and the extreme points of the AD group variations do not confine to a specific pattern. Thus, these channels have no significant changes in their respective channel inter-relationships. Therefore, considering the channel inter-relationships that can differentiate well between AD and HC groups, it is conclusive that the AD group exhibits a common level of complexity or a common patterns of synchrony in these respective channel inter-relationships.

Considering *only* the best performing channel-specific feature spaces (TABLE I), TABLE II and TABLE III summarises how well a participant in HC (TABLE II) and AD (TABLE III) groups on average are correctly classified using a specific epoch. Each channel-specific feature space will classify a participant as 1 for AD and 0 for HC (binary classification). Thus, for each epoch, an epoch average of these binary classifications is evaluated. A final average of all the epoch averages in also shown. The table cell colour, shading towards green indicates the value closer to 0 and towards yellow indicates the value is more close to 1. The participant identifier is matched to FIG. 3 by counting row-wise from left to right in the respective group within the specific epoch. For example, AD group participant number 13 with an identifier AD13, is the 13th kernel (dis)similarity matrix in FIG. 3 AD (when counted row-wise from left to right) within a specific epoch. In the case of AD1, the EEG data for 2nd and 3rd epochs for this participant was not available, thus denoted as N/A in TABLE II.

By directly comparing TABLE II with FIG. 3 HC, it can be noted that for those HC participants that have attained a final average of 0.5 and below, i.e. HC1, HC2, HC3, HC5, HC8 and HC13, their corresponding kernel (dis)similarity matrices exhibits a similar pattern as generally observed in the AD group. This may indicate these specific participants might be at a higher risk of developing symptoms of AD in the future despite experiencing no cognitive deficits at the time their EEGs were undertaken [20]. Furthermore, for those channel inter-relationships that are considered significant in this study, further in-depth dynamical analysis [3, 4, 23–25] of these HC EEGs should be conducted alongside the clinically diagnosed AD EEGs for confirmation. Such a postulation would have to be considered in a future longitudinal study aiming to identify if our proposed approach along with the in-depth dynamical analysis can identify asymptomatic individuals that are likely to develop symptoms of dementia. By comparing TABLE III and FIG. 3 AD, all

TABLE II. HC participant classification. Considering the best performing channels

Identifier	Epoch 1 average	Epoch 2 average	Epoch 3 average	Final average
HC1	1	0	0.5	0.5
HC2	1	0.167	0.167	0.444
HC3	1	0	0.333	0.444
HC4	1	0	0.833	0.611
HC5	1	0	0.167	0.389
HC6	0.833	0.167	0.667	0.556
HC7	1	1	1	1
HC8	0.333	0	1	0.444
HC9	1	0.167	0.667	0.611
HC10	1	0.667	0.833	0.833
HC11	1	1	0.167	0.722
HC12	1	1	1	1
HC13	0.833	0.333	0.333	0.5
HC14	1	1	1	1
HC15	0.5	0.667	0.667	0.611
HC16	1	1	1	1
HC17	1	1	0.833	0.944
HC18	1	0.833	0.5	0.778
HC19	0.833	1	1	0.944
HC20	1	0	0.833	0.611
		Total average		0.697

TABLE III. AD participant classification. Considering the best performing channels

Identifier	Epoch 1 average	Epoch 2 average	Epoch 3 average	Final average
AD1	1	NA	NA	NA
AD2	1	1	0.833	0.944
AD3	0.833	1	1	0.944
AD4	1	1	1	1
AD5	0.833	1	0.5	0.778
AD6	1	1	1	1
AD7	1	1	0.333	0.778
AD8	1	1	1	1
AD9	1	1	1	1
AD10	1	0.667	1	0.889
AD11	0.667	1	1	0.889
AD12	0.333	1	1	0.778
AD13	1	0.833	1	0.944
AD14	0.833	1	0.5	0.778
AD15	0.667	1	1	0.889
AD16	1	0.667	0.667	0.778
AD17	1	1	0.833	0.944
AD18	0.667	1	1	0.889
AD19	0.833	0	1	0.611
AD20	0	0.333	1	0.444
		Total average		0.857

AD participants attain a final average well above 0.5 except AD20, which is a unique case. When comparing the kernel (dis)similarity matrices from epochs 1 and 2 (FIG. 3 AD), AD20 does not exhibit exactly the general AD pattern except in epoch 3 where a resemblance is observed.

V. CONCLUSION AND FUTURE WORKS

A novel EEG channel selection methodology using kernel-based nonlinear manifold learning is presented. This channel selection method aims to determine which spatio-temporal channel inter-relationships are better at distinguishing between healthy controls from patients with Alzheimer's disease for further dynamical analysis (e.g. causality, frequency coupling). It was shown how a kernel-based spatio-temporal (dis)similarity matrix via manifold learning can be used as a measure of synchrony between EEG channels to determine which inter-relationships can be important to characterise patients with AD. The methodology presented can not only be used to extract global features, such as the kernel (dis)similarity matrix patterns between groups, but also to determine which specific EEG channel inter-relationships are important. Furthermore, the methodology exhibits consistency in its results as shown in the discussions related to the validation set. The main advantage of the proposed method is that the channel selection process considers both global and local spatio-temporal structures within the data in order to determine which linear/nonlinear channel inter-relations are significant. Furthermore, the methodology allows for the ranking of the importance in the channel inter-relationships. Thus enabling to project this information into the EEG electrode map to visualise and better comprehend the spatial layout of all the important inter-relationships relative to a certain channel. By observing the kernel (dis)similarity matrices and the group variations in the channel inter-relationships between HC and AD, it is evident that the AD group exhibits a common pattern of synchrony. Consequently, the AD group (dis)similarity variations relating to the channel inter-relationships that can differentiate well between HC and AD groups, confined to this common pattern of synchrony.

The main purpose of this paper is to introduce this novel channel selection methodology and its computation procedure, in-depth neuro-physiological interpretations were not discussed in the current study and will be part of future work. Results of only EO data from three epochs were examined. Examination of EC data and a comparison with EO along with comparisons with different time windows (within the 12-second epochs) will be investigated in the future. Another important future work is to further investigate in more detail the forms of nonlinearity using nonlinear dynamic modelling [23, 24] and causality measures in both time and frequency domains [25–27]. These in-depth dynamical analysis methods will be applied to the channel pairs determined using the Isomap-GPLVM method. Thus, enabling the further study of the underlining dynamical processes and connectivity in patients with Alzheimer's disease. This will aid in understanding which linear and nonlinear dynamical features can be used for early-onset diagnosis.

ACKNOWLEDGEMENTS

SRASG and FH acknowledge Coventry University for the Trailblazer PhD studentship. The EEG data was funded by a grant from the Alzheimer's Research UK, grant reference number ARUK-PPG20114B-25. This is a summary of independent research carried out at the NIHR Sheffield Biomedical Research Centre (Translational Neuroscience). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR or the Department of Health.

[1] P. L. Nunez and R. Srinivasan, *Electric fields of the brain*. Oxford University Press, 2006.

[2] C. Stam, "Nonlinear dynamical analysis of EEG and MEG: Review of an emerging field," *Clinical Neurophysiology*, vol. 116, pp. 2266–2301, oct 2005.

[3] F. He and Y. Yang, "Nonlinear system identification of neural systems from neurophysiological signals," *Neuroscience*, vol. 458, pp. 213–228, 2021.

[4] F. He, P. Sarrigiannis, S. Billings, H. Wei, J. Rowe, C. Romanowski, N. Hoggard, M. Hadjivassilliou, D. Rao, R. Grünewald, A. Khan, and J. Yianni, "Nonlinear interactions in the thalamocortical loop in essential tremor: A model-based frequency domain analysis," *Neuroscience*, vol. 324, pp. 377–389, 2016.

[5] T. Alotaiby, F. E. A. El-Samie, S. A. Alshebeili, and I. Ahmad, "A review of channel selection algorithms for eeg signal processing," *EURASIP Journal on Advances in Signal Processing*, vol. 2015, p. 66, Aug 2015.

[6] J. Dauwels, F. Vialatte, T. Musha, and A. Cichocki, "A comparative study of synchrony measures for the early diagnosis of alzheimer's disease based on eeg," *NeuroImage*, vol. 49, no. 1, pp. 668–693, 2010.

[7] P. Zerzucha and B. Walczak, "Concept of (dis)similarity in data analysis," *TrAC Trends in Analytical Chemistry*, vol. 38, pp. 116–128, 2012.

[8] J. Laub, V. Roth, J. M. Buhmann, and K.-R. Müller, "On the information and representation of non-euclidean pairwise data," *Pattern Recognition*, vol. 39, no. 10, pp. 1815–1826, 2006.

[9] N. Zaki, S. Lazarova-Molnar, W. El-Hajj, and P. Campbell, "Protein-protein interaction based on pairwise similarity," *BMC Bioinformatics*, vol. 10, p. 150, May 2009.

[10] I. Gati and A. Tversky, "Representations of qualitative and quantitative dimensions.," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 8, no. 2, p. 325, 1982.

[11] R. L. Goldstone, D. L. Medin, and D. Gentner, "Relational similarity and the nonindependence of features in similarity judgments," *Cognitive psychology*, vol. 23, no. 2, pp. 222–262, 1991.

[12] K. M. Hammouda and M. S. Kamel, "Efficient phrase-based document indexing for web document clustering," *IEEE Transactions on knowledge and data engineering*, vol. 16, no. 10, pp. 1279–1296, 2004.

[13] R. Mohanty, W. A. Sethares, V. A. Nair, and V. Prabhakaran, "Rethinking measures of functional connectivity via feature extraction," *Scientific Reports*, vol. 10, p. 1298, Jan 2020.

[14] A. S. Shirkhorshidi, S. Aghabozorgi, and T. Y. Wah, "A comparison study on similarity and dissimilarity measures in clustering continuous data," *PLOS ONE*, vol. 10, pp. 1–20, 12 2015.

[15] J. Dauwels, F. Vialatte, and A. Cichocki, "Diagnosis of alzheimer's disease from eeg signals: where are we standing?," *Current Alzheimer Research*, vol. 7, no. 6, pp. 487–505, 2010.

[16] B. Schölkopf, "The kernel trick for distances," in *Advances in Neural Information Processing Systems* (T. Leen, T. Dietterich, and V. Tresp, eds.), vol. 13, MIT Press, 2001.

[17] N. Lawrence, "Gaussian process latent variable models for visualisation of high dimensional data," in *Advances in Neural Information Processing Systems* (S. Thrun, L. Saul, and B. Schölkopf, eds.), vol. 16, MIT Press, 2004.

[18] H. Choi and S. Choi, "Robust kernel isomap," *Pattern Recognition*, vol. 40, no. 3, pp. 853–862, 2007.

[19] N. D. Lawrence and J. Quiñonero-Candela, "Local distance preservation in the GP-LVM through back constraints," in *Proceedings of the 23rd international conference on Machine learning - ICML '06*, ACM Press, 2006.

[20] D. Blackburn, Y. Zhao, M. D. Marco, S. Bell, F. He, H.-L. Wei, S. Lawrence, Z. Unwin, M. Blyth, J. Angel, K. Baster, T. Farrow, I. Wilkinson, S. Billings, A. Venneri, and P. Sarrigiannis, "A pilot study investigating a novel non-linear measure of eyes open versus eyes closed EEG synchronization in people with alzheimer's disease and healthy controls," *Brain Sciences*, vol. 8, p. 134, jul 2018.

[21] S. Luck, *An Introduction to the Event-Related Potential Technique*. A Bradford Book, Bradford Books, 2014.

[22] E. Schulz, M. Speekenbrink, and A. Krause, "A tutorial on gaussian process regression: Modelling, exploring, and exploiting functions," *Journal of Mathematical Psychology*, vol. 85, pp. 1–16, aug 2018.

[23] F. He, S. A. Billings, H.-L. Wei, P. G. Sarrigiannis, and Y. Zhao, "Spectral analysis for nonstationary and nonlinear systems: A discrete-time-model-based approach," *IEEE Transactions on Biomedical Engineering*, vol. 60, pp. 2233–2241, aug 2013.

[24] F. He, H.-L. Wei, and S. A. Billings, "Identification and frequency domain analysis of non-stationary and nonlinear systems using time-varying NARMAX models," *International Journal of Systems Science*, vol. 46, pp. 2087–2100, nov 2013.

[25] F. He, S. A. Billings, H.-L. Wei, and P. G. Sarrigiannis, "A nonlinear causality measure in the frequency domain: Nonlinear partial directed coherence with applications to EEG," *Journal of Neuroscience Methods*, vol. 225, pp. 71–80, mar 2014.

[26] Y. Zhao, S. A. Billings, H. Wei, F. He, and P. G. Sarrigiannis, "A new NARX-based granger linear and nonlinear casual influence detection method with applications to EEG data," *Journal of Neuroscience Methods*, vol. 212, pp. 79–86, jan 2013.

[27] F. He, H.-L. Wei, S. A. Billings, and P. G. Sarrigiannis, "A nonlinear generalization of spectral granger causality," *IEEE Transactions on Biomedical Engineering*, vol. 61, pp. 1693–1701, jun 2014.