

A model for classification based on the functional connectivity pattern dynamics of the brain

Regina Mészlányi^{1,3}, Ladislav Peska², Viktor Gál³, Zoltán Vidnyánszky³, Krisztian Buza³

¹Department of Cognitive Science,
Budapest University of Technology and
Economics,
Budapest, Hungary

²Faculty of Mathematics
and Physics
Charles University in Prague
Czech Republic

³Brain Imaging Centre
Research Centre for Natural Sciences
Hungarian Academy of Sciences
Budapest, Hungary

Abstract—Synchronized spontaneous low frequency fluctuations of the so called BOLD signal, as measured by functional Magnetic Resonance Imaging (fMRI), are known to represent the functional connections of different brain areas. Dynamic Time Warping (DTW) distance can be used as a similarity measure between BOLD signals of brain regions as an alternative of the traditionally used correlation coefficient and the usage of the DTW algorithm has further advantages: beside the DTW distance, the algorithm generates the warping path, i.e. the time-delay function between the compared two time-series. In this paper, we propose to use the relative length of the warping path as classification feature and demonstrate that the warping path itself carries important information when classifying patients according to cannabis addiction. We discuss biomedical relevance of our findings as well.

Keywords—fMRI; functional connectivity patterns; dynamic time warping; warping path; classification

I. INTRODUCTION

The increasing interest in methods that can be used to analyse brain networks may be attributed to various factors such as the enormous costs related to the treatment of brain disorders, large-scale research projects (such as the BRAIN initiative and the European Human Brain Project), and the increasing computational power that allows to process large amount of data describing brain activity.

Despite the recent advances in brain research, in case of many drugs, its effect on the organisation of the brain is largely unknown. One of such drugs is cannabis. The importance of understanding its effect on brain networks is underlined by the fact that its potential clinical use is being debated in several countries. Therefore, in this study, we focus on machine learning techniques that are able to identify differences between brain connectivity patterns of cannabis users and non-users, and to reveal insights about the effect of cannabis usage on brain function.

One of the state-of-the-art methods to capture brain activity is functional magnetic resonance imaging (fMRI), which captures the change of blood-oxygen-level, the so called BOLD signal. Synchronised spontaneous low frequency fluctuations of the BOLD signal during rest has been shown to represent the functional networks of brain areas [1]–[3]. Traditionally,

resting-state brain networks are analysed with techniques that imply static zero-lag linear dependence between brain regions, i.e. the strength of functional connectivity between any pair of voxels is usually characterized by the Pearson correlation coefficient of the two measured signals [1]. Other methods for revealing functional networks like independent component analysis [4] are similarly popular in the neuroimaging community, yet they still rely on measures of linear dependence.

Recently, growing number of neuroimaging studies have suggested that functional networks display dynamic changes in connectivity strength [5]–[7], as well as varying phase difference (nonzero time-lag) between regions [8]. To address these problems, Dynamic Time Warping (DTW) distance [9] can be used as an alternative measure of similarity between BOLD signals [10], since DTW is able to handle non-stationary processes and it results in more stable functional connectivity patterns than correlation. Beside DTW distance, the DTW algorithm computes the warping path as well, which encodes the dynamic time-delay structure between any two compared time-series. In case of BOLD signals, the warping path itself can contain relevant information about the relationship of the brain areas.

Recent studies [11]–[13] show that resting-state functional connectivity has great potential as a biomarker of several type of drug addictions. We hypothesise that descriptive scalar features of the warping path, in particular, the length of the warping path may hold potentially relevant information about the dynamics of the relationship.

In this study, we empirically validate the above hypothesis and propose to use classification features based on the length of warping path for the task of distinguishing cannabis users from non-users. Furthermore, we interpret our results from the biomedical point of view as well.

The rest of the paper is organized as follows: Section II reviews the background that is necessary to understand our work, Section III describes our approach, while in Section IV we present the results of the experimental evaluation. Section V discusses the relevance of our findings from the biomedical point of view. Finally, our conclusions are presented in Section VI.

II. BACKGROUND

In this section, we review major techniques on which our proposed approach relies.

A. Dynamic Time Warping

Dynamic Time Warping is a distance measure that takes potential shifting and elongations into account when comparing two time series. DTW was originally designed for speech recognition [9]. In the last decades it was shown to work exceedingly well for time series classification [14], [15], thus it became one of the most prominent time series distance measures in the machine learning community, see e.g. [16] and the references therein.

DTW is an edit distance, i.e., when comparing two time series, it calculates the “cost” of transforming one of the time series into the other one. Calculation of the DTW distance of two time series of length l_1 and l_2 , can be implemented by filling-in the entries of an $l_1 \times l_2$ matrix. This approach is illustrated in Fig. 1. Each entry of the matrix corresponds to the distance between two prefixes of the time-series. The entries can be filled-in column-by-column and row-by-row, in order to fill an entry $D(i,j)$ of the matrix, we use the following rule:

$$D(i,j) = \|t_1(i) - t_2(j)\| + \min(D(i-1,j-1), D(i-1,j), D(i,j-1)) \quad (1)$$

where $t_1(i)$ denotes the i -th value in time series t_1 and $t_2(j)$ denotes the j -th value in time series t_2 . Once the matrix is filled, the value in the entry $D(l_1, l_2)$ is the DTW-distance of the two time series.

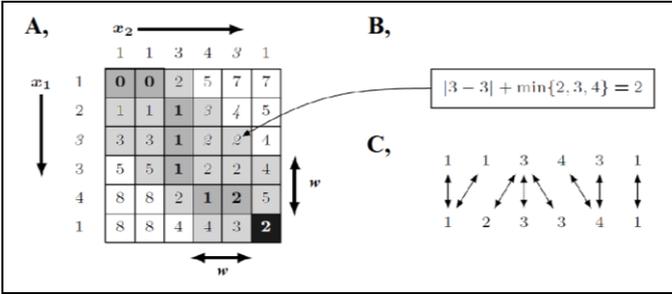


Fig. 1. A, Calculation of DTW distance by filling in the DTW matrix. Elements of x_1 correspond to rows, while elements of x_2 correspond to columns of the matrix; w denotes the size of the warping window, the maximal allowed time-lag between two matched time series element. The optimal warping path is highlighted with dark grey. B, Formula to calculate entry (i,j) : distance of $x_1(i)$ and $x_2(j)$ plus the minimum of the matrix entries $(i-1,j)$, $(i-1,j-1)$, $(i,j-1)$. C, Optimal matching of the elements of x_1 and x_2 revealed by the DTW matrix. Adapted with permission from [16].

In order to speed-up DTW-calculations, it is enough to calculate the cells close to the main diagonal of the matrix [9]. This corresponds to limiting the shifting that is allowed between matched positions of the two time series by applying a warping window. For example, when calculating only the marked entries in Fig. 1A, the size of the warping window is $w = 2$.

Once the matrix is filled, starting from $D(l_1, l_2)$, by considering which of the neighboring entries has led to the minimum in Eq. (1), we can construct the warping path, or, equivalently, the matching between the positions of the time series, see Fig. 1C for an example.

B. K-nearest neighbors (KNN)

We treat the classification according to cannabis addiction as a binary classification task with class labels “0” and “1”. As we will describe in Section III, we map brains into a vector space which allows to use any conventional vector classifier to solve the classification task.

Due to its simplicity and theoretical performance guarantees, K-nearest neighbors (KNN) is one of the most popular classifiers [17]. KNN classifies unknown objects based on its K nearest neighbors which are determined using the Euclidian distance in our case.

We used adaptive threshold KNN classification. Let us denote the instance to be classified by x . Its predicted class label $\bar{y}_N(x)$ is calculated as follows:

$$\bar{y}_N(x) = \begin{cases} 0, & \text{IF } \|\{x_i: x_i \in N_k(x) \wedge y_i = 0\}\| \geq t \\ 1 & \text{OTHERWISE} \end{cases} \quad (2)$$

where $N_k(x)$ is the set of nearest neighbors of x , y_i denotes the class label of x_i and t is a hyper parameter.

The number of nearest neighbors, K , can be seen as a regularization term: larger values of K reduce the effect of noise on the classification, but they could result in high bias, whereas low values of K may lead to overfitting.

We applied hyper parameter learning to determine appropriate values of K and t , see also Section IV.C. When evaluating the accuracy of classification, in order to calculate ROC curves, we need the continuous output of the classifier, by which we mean the ratio of nearest neighbors from the class labeled “1”.

C. LASSO regression

The Least Absolute Shrinkage and Selection Operator (LASSO) is a regularized regression method that performs feature selection, which enables it to deal with high dimensional datasets [18].

The class label predicted by LASSO, denoted as $\bar{y}(\bar{x})$ can be calculated as follows:

$$\bar{y}(x) = \begin{cases} 0, & \text{IF } \bar{x}^T \bar{\theta} \leq 0 \\ 1 & \text{OTHERWISE} \end{cases} \quad (3)$$

where $\bar{x} \in \mathbb{R}^d$ is an instance and vector $\bar{\theta} \in \mathbb{R}^d$ is the parameter of the model, the values of which are determined during training. In particular, LASSO’s objective is to find the parameter vector $\bar{\theta}$ that minimizes the sum of squared errors and the regularization term:

$$\bar{\theta} = \arg \min_{\bar{\theta}} \frac{1}{N} \|\bar{y} - X\bar{\theta}\|_2^2 + \lambda \|\bar{\theta}\|_1 \quad (4)$$

where N is the number of examples, $X \in \mathbb{R}^{N \times d}$ matrix contains the cases, d is the number of features (for the sake of simplicity, x_0 and θ_0 represents the bias term and its weight), $\bar{y} \in \mathbb{R}^N$ contains the desired output values, and $\lambda \in \mathbb{R}$ is a hyper parameter controlling the regularization, or equivalently, the sparsity of the resulting model. Before regression we normalized the input matrix X by setting each column’s standard deviation to 1 and mean to 0.

With continuous output of LASSO in case of a test instance \tilde{x} we mean the product $\tilde{x}^T \hat{\theta}$.

III. OUR APPROACH

We assume that the brain (or that part of it on which the domain expert’s analysis focuses) is segmented into regions of interest (ROIs) and the activity of the brain is given for each of the ROIs as function of time. We propose to calculate DTW between the time series associated with the ROIs and to characterize the phase shift between each pair of ROIs by the relative length of the warping path. With relative length of the warping path, we mean the difference between the actual length of the warping path and the main diagonal. This allows us to describe the brain’s dynamic behavior as a vector containing a component for each pair of ROIs. In other words: we map fMRI data into a vector space. We note that we map fMRI data of various subjects into the same vector space. Thus, conventional classifiers may be trained. Subsequently, these classifiers may be used to classify new patients.

Due to the complexity of the data, such as the relatively high amount of features compared to the usual number of subjects, we propose to use the two classifiers described in Section II. We choose the LASSO regression based classifier, since it is able to perform well in high-dimensional feature spaces and yields biologically interpretable results in terms of selected features, and KNN with adaptive threshold for to its simplicity and good performance, even though kNN is known to be affected by the curse of dimensionality. To address this issue in future work, application of some extensions of kNN classifiers, e.g. hubness-aware kNN [16] should be considered.

IV. EXPERIMENTAL EVALUATION

A. Data and preprocessing

In order to assist reproducibility and to perform classification of fMRI data according to a standard protocol, we downloaded a preprocessed public resting-state fMRI database from the 1000 Functional Connectomes Project, Addiction Connectome Preprocessed Initiative. In our study we used the MTA 1 dataset with the ANTS registered, no scrubbing, no global signal regression preprocessing pipeline. Detailed description of the preprocessing strategy is available at the homepage of the dataset (http://fcon_1000.projects.nitrc.org/indi/ACPI/html/). The downloaded dataset contains 126 subjects’ resting-state data as well as phenotypic information including cannabis addiction (62 addicted, 64 non-addicted subject).

We used a functional atlas of FIND Lab consisting of 90 functional regions of interest (ROI) [19] to obtain 90 functionally meaningful averaged BOLD signals per subject. In the downloaded data, one ROI (Basal Ganglia 4) included no meaningful measurement for any of the 126 subjects, therefore we used the remaining 89 ROIs, resulting in 3912 individual connectivity features for classification.

The aforementioned 89 ROIs correspond to 89 time series that capture the activities of the brain in various regions. From these 89 time series we calculated full connectivity matrices with DTW, i.e., we calculated the DTW matrix between each pairs

of time series. When calculating DTW, we set the size of the warping window to 100 s corresponding to 50 time-points. This is in accordance with the fact that the time-series in the publicly available version of the data are bandpass filtered with 0.01 Hz lower cut-off frequency.

B. Baselines

As baseline we used classification based on DTW distance (without considering the warping path) and Pearson correlation. For the later baseline, we mapped resting-state-fMRI (rs-fMRI) data into a vector space in the same way as in case of our approach, with the only difference that we calculated the Pearson correlation between each pair of time series and used the Pearson correlations as features describing the connectivity between ROIs. In case of the other baseline, we simply used the standard DTW distance instead of the relative path length.

C. Experimental Protocol

We performed experiments according to the leave-one-subject-out cross-validation protocol with both KNN and LASSO classifiers.

In both cases, in each round of the cross-validation, the values of the hyper parameters λ , K and t were determined using the training data only. In particular, we performed an *internal* cross-validation on the training data in order to select values of hyper parameters that maximize macro-averaged F-measure. We note that the macro-averaged F-measure converges to the accuracy, when the two classes are balanced (like in our case), and the number of positive and negative labels predicted by the classifier are also equalized. Therefore, searching for hyperparameters that maximize macro-averaged F-measure essentially corresponds to maximizing accuracy.

For KNN, K was learned from the set of odd numbers between 1 and 10, and we checked thresholds $t \in \{1, \dots, K\}$. For LASSO, λ was learned from the interval between 0.0005 and 0.5.

D. Performance metrics

To assess the performance of the proposed classification approach as well as the baselines, we calculated accuracy, the area under the receiver operator characteristic curve (AUC) and F-measure.

We define accuracy as the ratio of correctly classified subjects. When calculating AUC, we considered the class of cannabis addicts as the “positive” class and used the continuous output of the classifiers (see also Section II for the description of how the continuous output was calculated).

We calculated F-measure, i.e., the harmonic mean of precision and recall, for both classes and averaged the F-measures of the two classes. This led to a macro-averaged F-measure which we used to assess the quality of the classifiers.

In order to test if the differences between our approach and the baselines are statistically significant, we used the binomial test suggested by Salzberg [20].

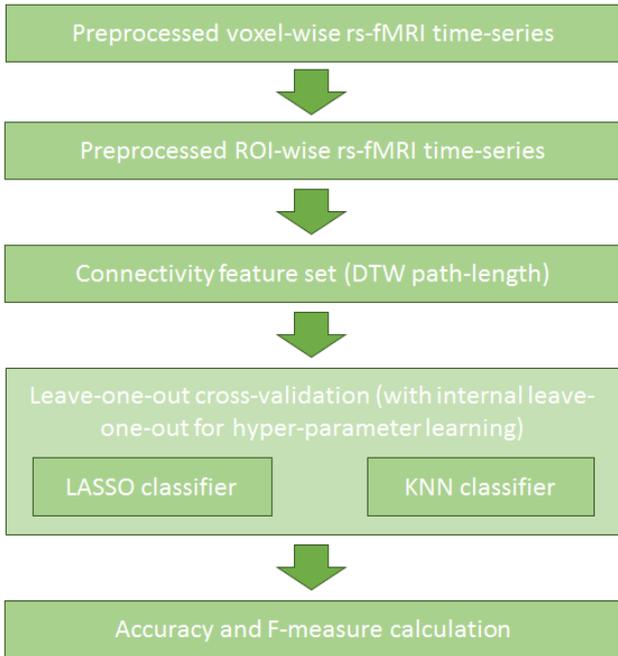


Fig. 2. The pipeline of our compound model: data preprocessing, classification and evaluation.

E. Classification results

For simplicity, we only report results for macro-averaged F-measures obtained with KNN and LASSO classification in Table I. However, we observed similar trends when we used accuracy and ROC as evaluation metrics.

TABLE I. CLASSIFICATION RESULTS OF CANNABIS ADDICTION

Average F-measure	Classification methods	
	<i>LASSO</i>	<i>KNN</i>
Correlation coeff.	0.27	0.50
DTW distance	0.47	0.42
Relative-path-length	0.62	0.64

For both classification methods, the best performance was achieved with the relative-path-length features. Classifiers using the proposed relative-path-length features significantly ($p < 0.05$) outperform both the classifiers based on correlation and the ones based on DTW distance.

The ROC curves of the classifiers using the proposed relative path length features are plotted on Fig. 3.

F. Stability of selected features

By assigning zero weights to a subset of features, the LASSO classifier applies feature selection: we consider a feature as “selected” if it has a non-zero weight. Next, we examine which

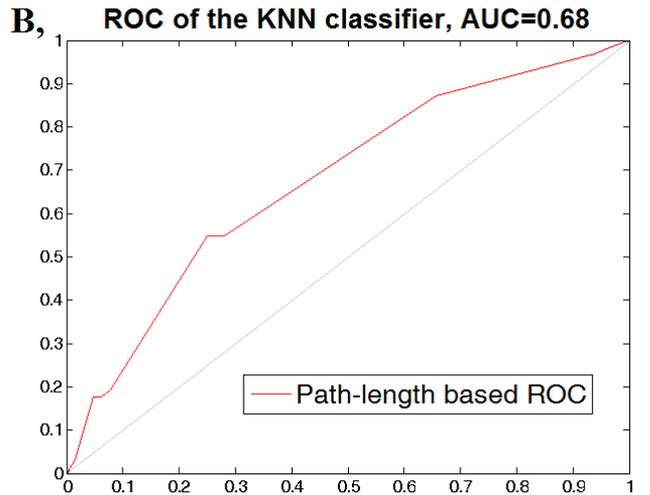
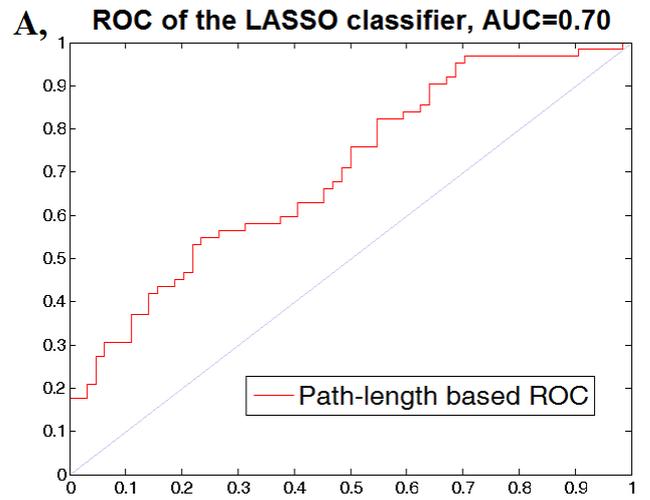


Fig. 3. ROC curves of the path-length based classifiers. A, ROC curve and AUC value of the LASSO classifier based on path-length features B, ROC curve and AUC value of the KNN classifier based on path-length features.

features are selected and how stable is the set of selected features. We considered a feature stable, if it was selected at least 100 times out of the 126 cycles (~80%) of leave-one-subject-out cross-validation. While the proposed relative path length based classifier has 61 stable features, DTW distance yields 13 and correlation only 4. It is important to note that the number of stable features does not correspond to the total number of selected features in the cross-validation cycles, e.g. the number of selected features range between 26 and 159 with relative-path-length-based classifiers, 9 and 97 with DTW distance based classifiers and 1 and 182 in case of correlation based classifiers. Considering the results presented in Table I, the number of stable features is closely linked to the classification performance of LASSO.

Fig. 4 shows stability of selected feature sets in more detail.

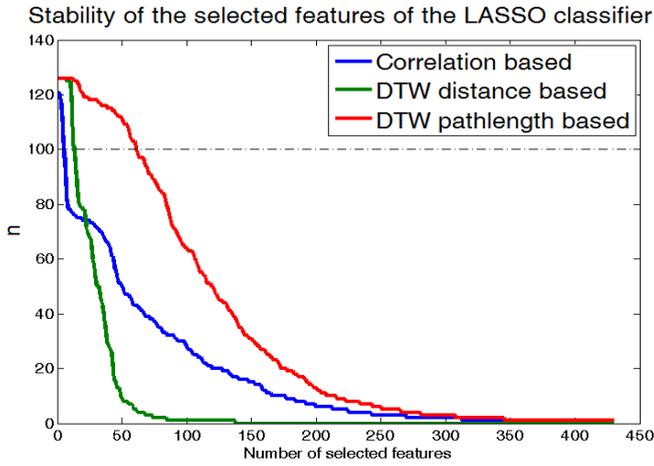


Fig. 4. The stability of the selected features based on correlation, DTW distance and warping path-length. Considering the 126 cycles of leave-one-out crossvalidation, we count how many features appear at least n -times among the selected features. The vertical axis shows n , while the horizontal axis displays the number of features that appear at least n -times among the selected features. The threshold of 100 selection is represented with the black dashed line.

The 61 features stably selected by relative-path-length-based LASSO classifier represents a brain network that contains ROI connections with the most different dynamics between addicted and non-addicted subjects. This connectivity pattern is visualized in Fig. 5.

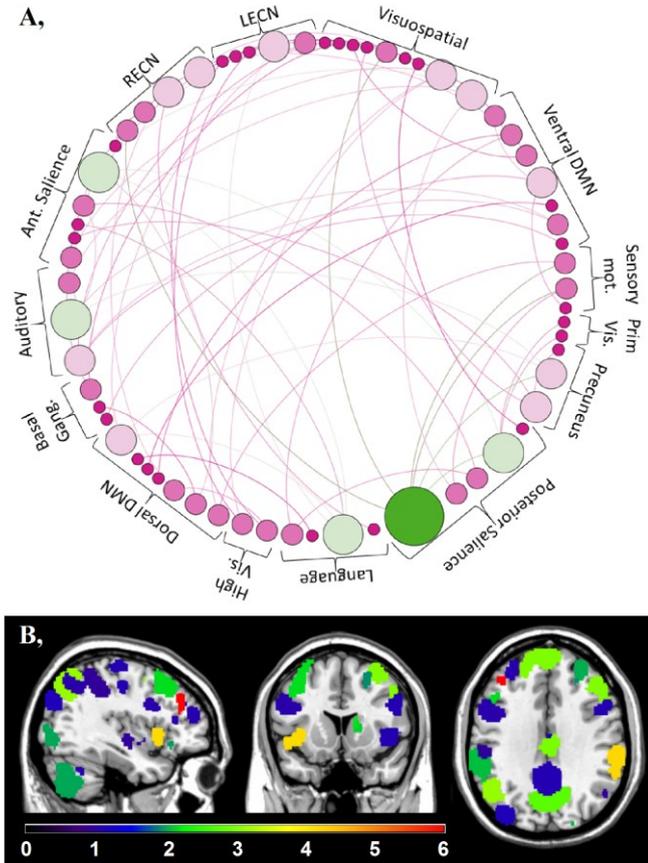


Fig. 5. A, Graph representation of the Cannabis network based on pathlength. The coloring and sizing of nodes corresponds to the degree of the given node (pink – low degree, green – high degree). B, The same network mapped back to functional ROIs, the coloring corresponds to the degree of the given node.

V. DISCUSSION

KNN and LASSO classification results suggest that in cannabis classification, information about the dynamics of the relationship of the ROIs is more valuable than in case of other connectivity measures such as correlation.

We note that in the general case, DTW distance and the length of the warping path are not independent. Consider the case of random time-series: longer path means more editing steps, and as the two series cannot be meaningfully matched this will result in higher DTW distance. On the other hand, in case of connections where the two time-series are highly linearly correlated, the warping path will almost exactly follow the main diagonal, resulting in very short path-length and in case of a good match, a relatively low DTW distance too. Consequently, DTW distance and the length of warping path may be highly correlated in the general case.

However, the length of the warping path itself contains additional information about the relationship in those connections, where a good match is achievable, but the time-delay structure between the signals is dynamically changing. In such cases, the DTW distance is rather small, whereas the warping path is relatively long. This hypothesis is supported by the finding, that while the overall correlation of DTW distances and warping path-lengths within subjects is $0.99 \pm 3e-5$, the same correlation calculated from only those 61 features that were invariably selected by LASSO in warping path based cannabis classification is only 0.57 ± 0.11 , implying that LASSO de facto chooses connections where warping path-length contains information beyond DTW distance value.

The cannabis network identified by the LASSO algorithm can be also interesting from the clinical point of view. The pattern of selected functional ROIs include several cognitive control and sensory processing related areas that have been reported to be associated with cannabis use [12], [21], [22], including connections which were found to be stronger in addicted subjects, resulting in both lower DTW distance and shorter path-length values due to some compensatory mechanisms [22], and connections that were shown to be weaker due to the general neural desynchronization reported in cannabis use [23] increasing the length of the warping path.

VI. CONCLUSIONS

Classification of brain activity patterns according to regular usage of cannabis is not only important because of the potential social impact of the explanations derived from the models learned by the classifiers, but also a challenging task from the machine learning point of view, because cannabis usage was reported to have only a minor effect in terms of conventional connectivity measures [24]. Therefore, we proposed a compound model using features based on the relative length of the DTW warping path for the classification of cannabis addiction. As shown by our experimental results, such features may be considered as a feasible descriptor of connectivity dynamics that may distinguish cannabis users from subjects who do not use cannabis regularly. Furthermore, we demonstrated that the model may contribute to reveal

biomedically relevant knowledge about brain functionality. We envision that features based on the length of the warping path may be used for other classification tasks related to brain functionality as well, such as classification tasks related to other types of addiction.

In future multi-task learning experiments, where other type of mental disorders, like ADHD, autism or Alzheimer's disease are examined beside addiction, intelligent combination [25], [26] of connectivity strength features like correlation and DTW distance, and features representing the connectivity dynamics like the length of the warping path should be considered. As classifiers based on different feature sets have different domains of competence, application of these combined models have great potential.

ACKNOWLEDGEMENTS

This work was supported by a grant from the Hungarian Brain Research Program (KTIA 13 NAP-A-1/18) to Z. Vidnyánszky. L. Peska was supported by project SVV-2015-260222. K. Buza was supported by the National Research, Development and Innovation Office - NKFIH PD 111710 and the János Bolyai Research Scholarship of the Hungarian Academy of Sciences.

REFERENCES

- [1] B. Biswal, F. Z. Yetkin, V. M. Haughton, and J. S. Hyde, 'Functional connectivity in the motor cortex of resting human brain using echo-planar MRI', *Magn. Reson. Med.*, vol. 34, no. 4, pp. 537–541, Oct. 1995.
- [2] B. B. Biswal, M. Mennes, X.-N. Zuo, S. Gohel, C. Kelly, S. M. Smith, C. F. Beckmann, J. S. Adelstein, R. L. Buckner, S. Colcombe, A.-M. Dagonowski, M. Ernst, D. Fair, M. Hampson, M. J. Hoptman, J. S. Hyde, V. J. Kiviniemi, R. Kötter, S.-J. Li, C.-P. Lin, M. J. Lowe, C. Mackay, D. J. Madden, K. H. Madsen, D. S. Margulies, H. S. Mayberg, K. McMahon, C. S. Monk, S. H. Mostofsky, B. J. Nagel, J. J. Pekar, S. J. Peltier, S. E. Petersen, V. Riedl, S. A. R. B. Rombouts, B. Rypma, B. L. Schlaggar, S. Schmidt, R. D. Seidler, G. J. Siegle, C. Sorg, G.-J. Teng, J. Veijola, A. Villringer, M. Walter, L. Wang, X.-C. Weng, S. Whitfield-Gabrieli, P. Williamson, C. Windischberger, Y.-F. Zang, H.-Y. Zhang, F. X. Castellanos, and M. P. Milham, 'Toward discovery science of human brain function', *Proc. Natl. Acad. Sci.*, vol. 107, no. 10, pp. 4734–4739, Mar. 2010.
- [3] M. D. Fox and M. E. Raichle, 'Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging', *Nat. Rev. Neurosci.*, vol. 8, no. 9, pp. 700–711, Sep. 2007.
- [4] M. J. McKeown, S. Makeig, G. G. Brown, T. P. Jung, S. S. Kindermann, A. J. Bell, and T. J. Sejnowski, 'Analysis of fMRI data by blind separation into independent spatial components', *Hum. Brain Mapp.*, vol. 6, no. 3, pp. 160–188, 1998.
- [5] R. M. Hutchison, T. Womelsdorf, E. A. Allen, P. A. Bandettini, V. D. Calhoun, M. Corbetta, S. Della Penna, J. H. Duyn, G. H. Glover, J. Gonzalez-Castillo, D. A. Handwerker, S. Keilholz, V. Kiviniemi, D. A. Leopold, F. de Pasquale, O. Sporns, M. Walter, and C. Chang, 'Dynamic functional connectivity: Promise, issues, and interpretations', *NeuroImage*, vol. 80, pp. 360–378, Oct. 2013.
- [6] D. A. Handwerker, V. Roopchansingh, J. Gonzalez-Castillo, and P. A. Bandettini, 'Periodic changes in fMRI connectivity', *NeuroImage*, vol. 63, no. 3, pp. 1712–1719, Nov. 2012.
- [7] E. A. Allen, E. Damaraju, S. M. Plis, E. B. Erhardt, T. Eichele, and V. D. Calhoun, 'Tracking whole-brain connectivity dynamics in the resting state', *Cereb. Cortex N. Y. N 1991*, vol. 24, no. 3, pp. 663–676, Mar. 2014.
- [8] C. Chang and G. H. Glover, 'Time-frequency dynamics of resting-state brain connectivity measured with fMRI', *NeuroImage*, vol. 50, no. 1, pp. 81–98, Mar. 2010.
- [9] H. Sakoe and S. Chiba, 'Dynamic programming algorithm optimization for spoken word recognition', *IEEE Trans. Acoust. Speech Signal Process.*, vol. 26, no. 1, pp. 43–49, Feb. 1978.
- [10] Regina Mészlyei, L. Peska, V. Gál, Z. Vidnyánszky, and K. Buza, 'Classification of fMRI data using Dynamic Time Warping based functional connectivity analysis', *Proc. 24th Signal Process. Conf. EUSIPCO*, 2016.
- [11] M. D. Rosenberg, E. S. Finn, D. Scheinost, X. Papademetris, X. Shen, R. T. Constable, and M. M. Chun, 'A neuromarker of sustained attention from whole-brain functional connectivity', *Nat. Neurosci.*, vol. 19, no. 1, pp. 165–171, Nov. 2015.
- [12] H. Cheng, P. Skosnik, B. Puce, M. Brumbaugh, J. Vollmer, D. Fridberg, B. O'Donnell, W. Hetrick, and S. Newman, 'Resting state functional magnetic resonance imaging reveals distinct brain activity in heavy cannabis users - a multi-voxel pattern analysis', *J. Psychopharmacol. (Oxf.)*, vol. 28, no. 11, pp. 1030–1040, Nov. 2014.
- [13] H. Gu, B. J. Salmeron, T. J. Ross, X. Geng, W. Zhan, E. A. Stein, and Y. Yang, 'Mesocorticolimbic circuits are impaired in chronic cocaine users as demonstrated by resting-state functional connectivity', *NeuroImage*, vol. 53, no. 2, pp. 593–601, Nov. 2010.
- [14] H. Ding, G. Trajcevski, P. Scheuermann, X. Wang, and E. Keogh, 'Querying and Mining of Time Series Data: Experimental Comparison of Representations and Distance Measures', *Proc VLDB Endow*, vol. 1, no. 2, pp. 1542–1552, Aug. 2008.
- [15] X. Xi, E. Keogh, C. Shelton, L. Wei, and C. A. Ratanamahatana, 'Fast Time Series Classification Using Numerosity Reduction', in *Proceedings of the 23rd International Conference on Machine Learning*, New York, NY, USA, 2006, pp. 1033–1040.
- [16] N. Tomasev, K. Buza, K. Marussy, and P. B. Kis, 'Hubness-aware classification, instance selection and feature construction: Survey and extensions to time-series', in *Feature selection for data and pattern recognition*, Springer-Verlag, 2015.
- [17] L. Devroye, L. Györfi, and G. Lugosi, *A Probabilistic Theory of Pattern Recognition*. Springer Science & Business Media, 2013.
- [18] R. Tibshirani, 'Regression Shrinkage and Selection via the Lasso', *J. R. Stat. Soc. Ser. B Methodol.*, vol. 58, no. 1, pp. 267–288, 1996.
- [19] W. R. Shirer, S. Ryali, E. Rykhlevskaia, V. Menon, and M. D. Greicius, 'Decoding Subject-Driven Cognitive States with Whole-Brain Connectivity Patterns', *Cereb. Cortex*, vol. 22, no. 1, pp. 158–165, Jan. 2012.
- [20] S. L. Salzberg, 'On Comparing Classifiers: Pitfalls to Avoid and a Recommended Approach', *Data Min Knowl Discov*, vol. 1, no. 3, pp. 317–328, Jan. 1997.
- [21] S. Bhattacharyya, I. Falkenberg, R. Martin-Santos, Z. Atakan, J. A. Crippa, V. Giampietro, M. Brammer, and P. McGuire, 'Cannabinoid Modulation of Functional Connectivity within Regions Processing Attentional Salience', *Neuropsychopharmacology*, vol. 40, no. 6, pp. 1343–1352, May 2015.
- [22] I. H. Harding, N. Solowij, B. J. Harrison, M. Takagi, V. Lorenzetti, D. I. Lubman, M. L. Seal, C. Pantelis, and M. Yücel, 'Functional Connectivity in Brain Networks Underlying Cognitive Control in Chronic Cannabis Users', *Neuropsychopharmacology*, vol. 37, no. 8, pp. 1923–1933, Jul. 2012.
- [23] P. D. Skosnik, J. A. Cortes-Briones, and M. Hajós, 'It's All in the Rhythm: The Role of Cannabinoids in Neural Oscillations and Psychosis', *Biol. Psychiatry*, vol. 0, no. 0, Dec. 2015.
- [24] L. Tamm, J. N. Epstein, K. M. Lisdahl, B. Molina, S. Tapert, S. P. Hinshaw, L. E. Arnold, K. Velanova, H. Abikoff, and J. M. Swanson, 'Impact of ADHD and cannabis use on executive functioning in young adults', *Drug Alcohol Depend.*, vol. 133, no. 2, pp. 607–614, Dec. 2013.
- [25] M. Woźniak and B. Krawczyk, 'Combined classifier based on feature space partitioning', *Int. J. Appl. Math. Comput. Sci.*, vol. 22, no. 4, pp. 855–866, 2012.
- [26] L. I. Kuncheva, 'Clustering-and-selection model for classifier combination', in *Fourth International Conference on Knowledge-Based Intelligent Engineering Systems and Allied Technologies, 2000. Proceedings*, 2000, vol. 1, pp. 185–188 vol.1.