Statelets: high dimensional predominant shapes in dynamic functional network connectivity

Md Abdur Rahaman[1, 2], Eswar Damaraju[2], Debbrata Kumar Saha[1, 2], Sergey M. Plis[2], Vince D. Calhoun[1,2]

[1]*Georgia Institute of Technology, Atlanta, GA, USA*
[2]*Tri-Institutional Center for Translational Research in Neuroimaging and Data Science (TReNDS), Georgia State University, Georgia Institute of Technology, Emory University, Atlanta, GA, USA*

***Abstract*— Dynamic functional network connectivity (dFNC) analysis has been attracting interest over the past years by elucidating crucial details of brain activation patterns in severe neurological or psychiatric disorders. Consequently, a considerable amount of work has been conducted to analyze dFNC and leverage it for an improved understanding of neuro-dynamics and cognition. However, state-of-art methods mostly focus on fixed patterns of connectivity that reoccur. Such approaches do not capture the more transient feature space of the brain's neural system and its dynamic properties. Likewise, dynamical system approaches model the system continuously and do not well capture unmodelled heterogeneity across subjects and time. Here, we seek an improved understanding of these connectivity states and the mechanism through which their dynamics vary across individuals. Relying on the concept of shapelets to model the temporal dynamics of functional associations between intrinsic brain networks, we develop the 'statelet' - a high dimensional state-shape representation of dynamics. We handle scale differences using the earth mover distance as our similarity metric and utilizing kernel density estimation to build a probability density profile for each shapelet. Unlike clustering the time series, we approximate short, variable length connectivity patterns using the probability density profile of the shapelets from lower dimensions. We apply the proposed approach to study patients with schizophrenia (SZ) exhibit and identify reduced modularity increased statelet recurrence in their brain network organization relative to healthy controls (HC). An analysis of the connections' consistency across time reveals significant differences within visual, sensorimotor, and default mode regions where HC subjects show higher consistency than SZ. The introduced statelet-approach enables the handling of dynamic information in cross-modal applications to study healthy and disordered brains and multi-modal fusion within a single dataset.**

Keywords—Earthmover distance (EMD), dFNC, shapelets, kernel density estimator, dFNC states, Schizophrenia.

I. Introduction

Dynamic functional network connectivity (dFNC) enables the investigation of temporal changes in connections among different regions of the brain. It has been established as a powerful technique for analyzing time-varying activity coordination across brain regions [1, 2]. The Pearson correlation typically characterizes the coordination, computed using a sliding window (e.g., 44s in length) technique for measuring the interdependence among brain regions or independent components of interest extracted from group ICA [1, 3]. Recent empirical evidence has strongly suggested that this dynamic spatiotemporal configuration better models brain activities and network organization [2, 4-7]. The analysis can track down the time-varying transitions in connectivity strengths, thus moving beyond the firm and incorrect assumption of a single static connectivity pattern in dFNC state analysis [3, 6, 8]. Additionally, such an approach is informative about the dynamic variations in brain network organizations, which are known to be highly dynamic [9, 10]. Yet, many analysis paradigms hypothesize a fixed discrete connectivity pattern with occupancy varying over time and do not fully capture recurring transient information. The obtained dFNC states essentially fuse all homogenous connectivity patterns ignoring their order and temporal index. State-of-the-art methods miss the dynamic properties of dFNC, such as network topology, synchronizability, and intermittent connectivity, which is also different in neuropsychiatric disorders like schizophrenia [11, 12]. These transformations tend to be transient, over a slower timescale, the dynamics can modulate functional network topology [9, 13].

Modern neuroimaging studies which focus on data-driven symptom investigation and bio-marker discovery, focus on identifying interpretable subsets of data that characterize diseases and controls [14-16]. For time-series data, researchers have focused on discovering shapelets - subsequences of a time series with high predictive power [17]. Shapelets have been applied to various fields of study, including genetics and medicine [18, 19]. However, to our best knowledge, shapelets have not been employed in the context of dynamic functional connectivity. Extracting characteristic features from a dFNC time series irrespective of their class label is not well studied. Since analyzing the entire dFNC time course and mining all timepoints quickly leads to an intractable feature space, in this paper, we address the tradeoff between the desire to model transient dynamics and the need to constrain the set of extracted features to a reasonable size. Researchers have already established the approach of ignoring some data to better cluster/subgrouping the samples [20]. Furthermore, considerable dFNC patterns can be transient enough to be contained in short subsequences, so shapelet-based analysis has great potential. As such, rather than feeding every window of a time-course, our method captures highly frequent and the most predictive connectivity patterns. The hypothesis of obtaining descriptive statelets between patients and controls helps interpret the group differences, and explainable features form both trajectories.

Neuroimaging features are inherently high dimensional and heterogeneous across individuals. The analysis becomes complex, following the diverse interpretations of a neural signature in distinct scenarios. Therefore, pattern mining in brain imaging data is complicated, even within a one-dimensional time series. The high dimensionality induces from all-to-all connectivity among a set of components in dFNC analysis. For each subject, dFNC generates an $O(n^2*t)$ connectivity space where n is the number of components, and t is the number of windows (length of time series). For a reasonably sized dataset, space grows exponentially, thus exploring the dynamics becomes more convoluted.

XXX-X-XXXX-XXXX-X/XX/$XX.00 ©20XX IEEE

Researchers have already implemented conventional clustering (e.g., *k*-means) on the dFNC time series, also known as dFNC state analysis [1, 3, 6, 21, 22]. Typically, these methods cluster the entire time series. Moreover, a recent study of tri-clustering dFNC data suggested three-dimensional states of conenctivity and reveals more significant group differences than previous studies [23]. So, our idea is to reduce the data down to a subset of patterns that potentially better explain the dynamics. The study suggests tracking down the representative patterns from the dFNC time series for enhancing the sensitivity of the outcomes and a more accurate assessment of the dynamic [24]. So, our method searches for the most prominent trends in an unsupervised manner, which can be used as a prototype for the shape of connectivity through which we can evaluate the dynamics more closely. Shapelets of a time series are defined as the small patterns in the sequence highly predictive of a class and introduced in this study [17]. In our study, we define a shapelet as the most repetitive pattern in a dFNC time course. Then, we generalize the local shapelets for discovering robust and high dimensional state shapes of dFNC, which we call 'statelets'. The method uses a kernel density estimator (KDE) to compute the probability density of the shapelets, which are later used to check the eligibility of shapelets to be interpreted as statelets [25]. We aim to approximate a corset of shapes for the dynamics rather than subgrouping the time series using the shapelets similarity. Our experiments show that investigating connectivity through a subset of characteristic shapes (statelets) can explain the dynamics better.

Functional magnetic resonance imaging (fMRI) of the human brain has been used broadly for studying brain disorders [26-28]. Outcomes from fMRI analysis reinforce the intuition of dysconnectivity in disorders such as schizophrenia, i.e., unusual connections among distinct brain networks [3, 29, 30]. The decomposition of neural features and generating biomarkers helps distinguish schizophrenia to a greater extent [31-34]. The majority of the system remains unexplored a task that is even more challenging because of its heterogeneous nature [35, 36]. Thus, the researchers started using subgrouping/clustering of the subjects to minimize the dissimilarity and making a more sensible and fair comparison [37-39]. We propose statelets for evaluating the prototypes of connectivity patterns by extracting the predictive signatures. These prototypes represent state 'movies' of brain connectivity patterns. Additionally, statelets enable investigation of dynamic properties of brain kinetics such as recurrence, modularity, and synchronizability. We explore the dynamic behavior of a healthy and schizophrenic brain and their group differences. Results show reduced modularity in patient's (SZ) dynamics relative to healthy controls (HC) (Fig. 12). Passage coding - convolving a pattern with the time course of both dynamics by their own statelet establishes patterns in HC that are more frequently recurring than in SZ, reflected in the time decay graph (Fig. 13). The probability density of group statelets reveals unique sequences of connections among functional brain networks that provide new insights into the functional connectivity changes associated with the disorder.

II. Definitions and Background

A. Definition 1: Dynamic functional connectivity(dFNC)

dFNC is a time-varying correlation typically computed using a sliding window (block of time points, e.g., 30-60 second windows) technique among brain regions or independent components (ICs) of interest. The correlation value at each window is quantified over the functional connectivity strength between a pair of brain networks within the time frame. In our problem, each object in the dataset is a dFNC time course that characterizes the connections between two ICs.

B. Definition 2: Connections/pair

Functional connectivity is defined as the pairwise correlation between time courses of independent neural components extracted from group ICA of fMRI data. We used 'connection' and 'pair' interchangeably in our writing. Since we are considering 47 components, we have 1081 different connections for each subject. The strength of these connections is given by the dFNC time courses.

C. Definition 3: dFNC Time Course

A dFNC time course, $T = W_1, W_2, W_3, \ldots\ldots, W_n$ is a sequential set of real values (correlation) where n is the length of the time course, and W stands for the window.

D. Definitions 4: Shapelet

Shapelet is a subsequence of the dFNC time course that is considered representative of a type of connectivity pattern. A shapelet $H_{i,m}$, is a subset of m contiguous real values from the time course started at the window i. Shapelets can also be treated as a connectivity pattern in the dFNC time course.

E. Definition 5: Statelet

Statelet is the representative of a group of shapelets extracted from multiple dimensions of the dataset. Therefore, statelets produce a high dimensional state shape representation of the time series using the probability density profile of shapelets observed in lower dimensions. These patterns can also be considered as the prototype of dFNC characteristic shapes. We will discuss statelets more in the following paragraphs.

Studies described the dynamic functional connectivity patterns as a state of the brain [1, 40]. Thus, state analysis has become prominent and yields significant insight into the human brain's cognitive system. Here, window stands for a vector of k correlation values where k is the number of connections or the number of pairs of independent components we are considering. For example, if we are considering n independent components, the length of the vector is ${}^{n}C2 = n(n-1)/2$. Each value in these vectors indicates the functional connectivity strength between two brain components within a given time frame, which can be treated as a transient neural connection. Brain connectivity is a complex system, gazing at the whole brain at a time is not a

Fig. 1. Comparison between two arbitrary signatures A and B in a conventional clustering method. It considers the whole sequence for measuring the distance (e.g., Euclidean). In dFNC time course, it considers a connectivity pattern for a given window W_i across all the connections. For each window, the length of the signature is n, where n is the number of pairs of components.

very efficient way to exploit the relations and explore the system. As an example, we can show the following scenario where we take two random dFNC time courses. Fig. 1 shows the conventional way of comparing the time series in the previously reported dFNC state analysis studies. Time courses A & B look different if we consider the whole sequence. Considering all the data at once makes it harder to observe the similarities in these signals. It also fails to provide a meaningful outcome for several reasons. Firstly, the most relevant (and repeated) patterns might not have spread across the brain. Secondly, these patterns might be transient. Also, the presence of noise in the data can hugely limit the performance of the comparison. So, we require a better approach to capture the dynamics, track down transient and length invariant patterns. To do this, we can focus on the most frequent patterns of the time series. The intuition behind this is to make sure that we are investigating the more frequent and recurring patterns. Hence, we estimate the probability density of the shapelets rather than using information gain or other discriminative measures. The motifs that are random and unpredictable is less helpful to the learning of the model. So, instead of generating some fixed-length signatures with limited temporal information,

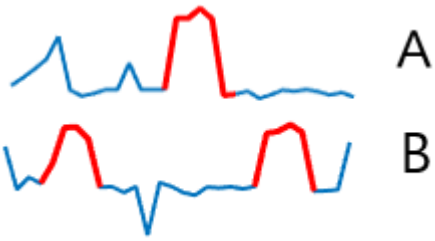

Fig. 2. Comparing only the shapelets (bold red) of dFNC time courses.

statelets account for the dynamic changes in functional brain connectivity and precisely quantifies the patterns leading the dynamics, including their temporal and spatial adjustments. These dynamic perturbations of connections between discrete brain components also extract information about the brain connectivity graph. The connectivity graph is a multigraph formulated from the brain. The components as the nodes and the connections are the edges; corresponding correlation values indicate the edges' weight.

We can see that the comparison has improved drastically in fig. 2. The shapelets from the time series look very similar to each other using the dominant patterns only. The distance between them is smaller, as well. In the earlier comparison, the differences create redundant states that potentially complicate the outcomes. Creating more states may help with that issue, but introduces others since researchers often use *k*-means clustering [13, 40] for estimating dFNC states. Adding more states can result in assigning a lot of loosely connected windows to each of the *k* states. Therefore, the outcomes that have been reported from those states have less concrete meaning in the context. Furthermore, model order selection is a free choice in state-of-the-art methods. We tried to address this issue in our framework by capturing state shapes representation of all intrinsic subgroups. To our best knowledge, this is the first endeavor to evaluate the state shapes from local shapelets based on their probability density throughout the datasets. Discovering local patterns to interpret as statelets, the journey of a shapelet is controlled by its consistency across the dynamics. All the local patterns are propagated along the dimensions; consequently, the statelets represent a significantly extensive collection of connectivity patterns that portray several states of the brain.

III. Data Collection and Preprocessing

We used the fBIRN dataset for generating the dFNC time courses in this project [41]. The data repository has resting-state functional magnetic resonance imaging data collected from 163 healthy controls (117 males, 46 females; mean age 36.9), and 151 age-and-gender matched patients with SZ (114 males, 37 females; mean age 37.8) during the eyes-closed condition. Collected data pass-through data quality control

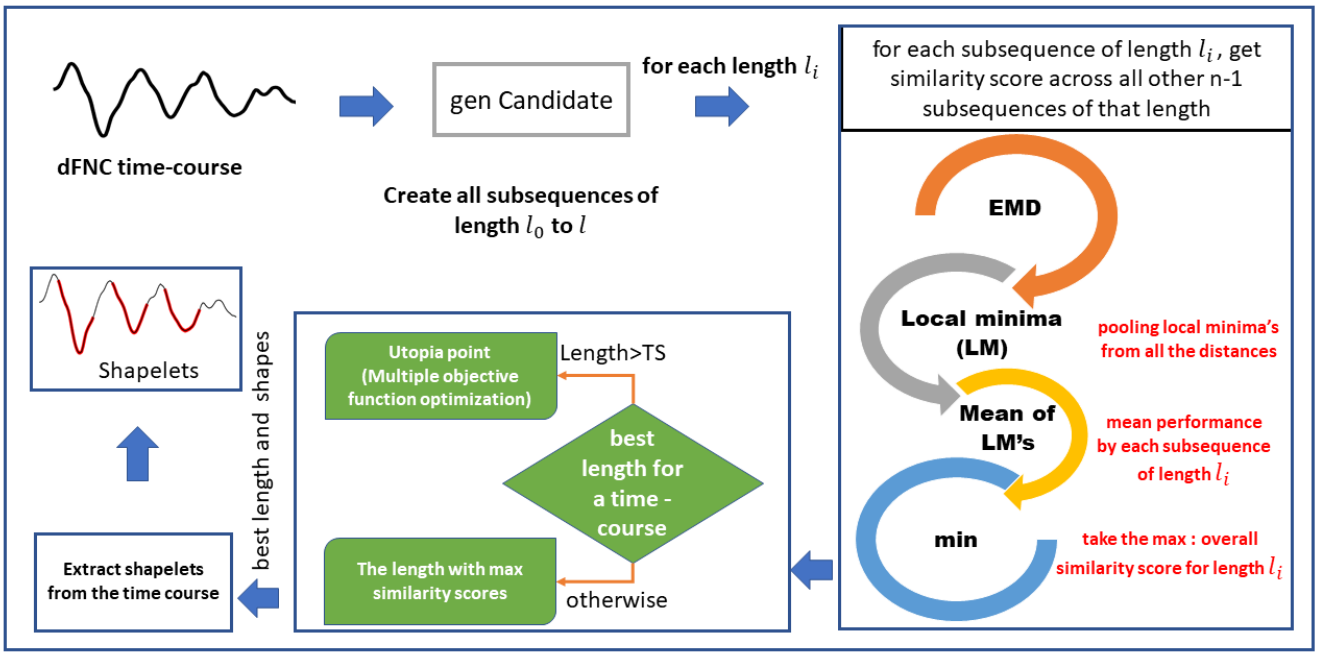

Fig. 3. The framework for extracting shapelets from the dFNC time series. Earthmover distance (EMD) is used as a similarity measure. The method takes one-dimensional dFNC time courses and returns the most persistent patterns from them. It also provides information about the multiple occurrences of a shapelet in a time series.

(explained in Damaraju et al., 2014 [1]). The participant's consent was obtained before scanning following the Internal Review Boards of corresponding institutions. Data were collected with a TR of 2 seconds on 3T scanners. Imaging data for six of the seven sites were collected on a 3T Siemens Tim Trio System and a 3T General Electric Discovery MR750 scanner at one site. Resting-state fMRI scans were acquired using a standard gradient-echo echo-planar imaging paradigm: FOV of 220 × 220 mm (64 × 64 matrices), TR = 2 s, TE = 30 ms, FA = 770, 162 volumes, 32 sequential ascending axial slices of 4 mm thickness and 1 mm skip. Subjects had their eyes closed during the resting state scan. Data pre-processing, quality control, and dFNC collection follow the standard pipeline described in [3]. Firstly, rigid body motion correction has been done using the INRIAlign toolbox in SPM to correct for subject head motion followed by a slice-timing correction to account for timing differences in slice acquisition [42]. Then the scans went through a 3dDespike algorithm to regress out the outlier effect and warped to a Montreal Neurological Institute (MNI) template and resampled to 3 mm^3 isotropic voxels. Instead of Gaussian smoothing, we smoothed the data to 6 mm full width at half maximum (FWHM) using the BlurToFWHM algorithm, which performs smoothing by a conservative finite difference approximation to the diffusion equation. Additionally, The voxel time course was variance normalized before perform independent component analysis (ICA) [43] as this has shown better to decompose subcortical sources in addition to cortical networks. Then, we performed group ICA on the preprocessed data and identified 47 intrinsic connectivity networks (ICNs) from the decomposition of 100 components and dFNC between two ICA time courses was evaluated using a sliding window approach with a window size of 22 TR (44 s) in steps of 1 TR [1, 44]. The window constituted a rectangular window of 22 time points convolved with Gaussian of sigma 3 TRs to obtain tapering along the edges [1]. We estimated covariance from the regularized inverse covariance matrix (ICOV) [45, 46] using the graphical LASSO framework [47]. The regularization parameter was optimized for each subject by evaluating the log-likelihood of the subject's unseen data in a cross-validation framework—all post hoc steps for extracting and validating. The subject wise dFNC values were Fisher-Z transformed and residualized with respect to age, gender, and site.

IV. Our Proposed Framework

A. Local Shapelets discovery

We implement the Earth mover distance (EMD) for computing distances between shapelets. The EMD is a measure of dissimilarity between two probability distributions [48, 49]. It corresponds to the minimum amount of work required to make two distributions look-alike [50]. For *p* and *q* two given distributions, *F (p, q)* represents the set of all possible flow between *p* and *q*. Then, the work is defined as follows,

$$Work\ (F, p, q) = \sum_{I=1}^{m} \sum_{j=1}^{n} f_{ij}\, d_{ij} \tag{1}$$

here d_{ij} is the distance between p_i and q_j, m and n denote the number of clusters in distribution *p* and *q* respectively. The EMD between *p* and *q* is the minimum amount of work required to match these two distributions, which are given by the following equation [50, 51].

$$EMD\ (p,\ q) = \frac{\min_{F=(f_{ij}) \in F(p,q)} Work\ (F, p, q)}{Total\ flow} \tag{2}$$

EMD implementation

We use a particular case of EMD for one-dimensional time series data, which parses through the vectors and keeps track of how much flow must be done between consecutive time points. Here, flow is the difference between the amplitudes of two shapes at a given time point. The method recursively accumulates the absolute work done at each time point, and finally, the summation over the time points corresponds to the distance (see equation 5 and 6). Equation 3 and 4 normalize shapes P and Q, respectively.

$$P_i = P_i - min\ (P),\ \ where\ P_i = \frac{P_i}{\sum_{i=1}^{l} P_i} \tag{3}$$

$$Q_i = Q_i - min\ (Q)\ ,\ where\ \ Q_i = \frac{Q_i}{\sum_{i=1}^{l} Q_i} \tag{4}$$

$$EMD_{i+1} = EMD_i + (P_i - Q_i),\ where\ EMD_0 = 0 \tag{5}$$

$$Distance\ (P, Q) = \sum |EMD_i| \tag{6}$$

We divide the dataset into two groups including 163 healthy control subjects (HC) and 151 schizophrenia patients (SZ). Each subject has ${}^{47}C_2 = 1081$ pairs of components: thus, 1081 dFNC time courses. For each pair, we take time courses from all the subjects within that group (HC/SZ). We have 151 and 163 dFNC time courses per pair of components for SZ and HC, respectively. Fig. 3 demonstrates the method for performing local shapelet discovery. The input is the dFNC time course in our study. At first, the process creates all the candidate shapelets of different lengths and passes them through the following mini layers.

1) EMD layer

For a given candidate of length l_i, the algorithm searches through the timepoints by taking all other candidates of that length and computing the EMD between them. We can consider the selected candidate as a kernel here, and the idea is to apply that kernel on the input signal to compute distances from that candidate to others.

2) Pooling Local Minima

The intuition is to pool a subset of top matches (minimum distance) instead of having just one across the signal to get an overall performance by that candidate across the signal. It divides the input signal into several vicinities and pool one minimum for each vicinity.

3) Mean of LM's

This layer computes the mean of those local minima, representing the overall performance by that specific candidate of length l_i.

4) Global Minimum

After finishing computation for all the candidates of l_i, the method selects the global minimum, which corresponds to the best acquirable matching score (compliment of distance) if the plan goes with the length l_i and also store candidate (leader) that shows the best performance. These scores are the recommender unit's input, where it decides the best length from available options.

Using the leader shapelet as the benchmark, the method locates similar non-overlapping occurrences across the time course. Finally, store those shapelets for creating global dominants across the subjects for a given pair.

B. *Statelets*

For each connection, we have a group of shapes from all the subjects. Now, we run a module called "FindDominants" consist of the four following steps (a, b, c, and d) mentioned in fig. 4 to identify pairwise dominant shapelets. Then, we gather those pairwise, highly frequent shapes into a bag. These shapes cover all the subjects and pairs of the group (SZ/HC). Now, we estimate the dominants across the dataset using the probability density of the shapelets in our collection. To do this we extract the representative of each subgroup, which we call statelets. Each statelet has a persistence level, which is given by their probability density across group dynamics. We follow the same strategy for both subject groups (SZ/HC).

a) *Distance measure*

We use EMD as the distance measure after normalizing the time courses to make our distance invariant to scale and offset. Then, compute cross EMD across the shapes, which creates a square matrix of n by n dimension where n is the number of shapes in the given collection.

b) *Kernel density estimator (KDE)*

Here, we use kernel density estimation to evaluate the probability density function of a random variable, which gives a better approximation of the studied probability distribution compared to a traditional histogram [52, 53]. Our method applies KDE to calculate probability density by each of the shapelets in the subset. We incorporate KDE with a Gaussian kernel and optimal bandwidth [54, 55]. The kernel density at a point x is given by equation (7),

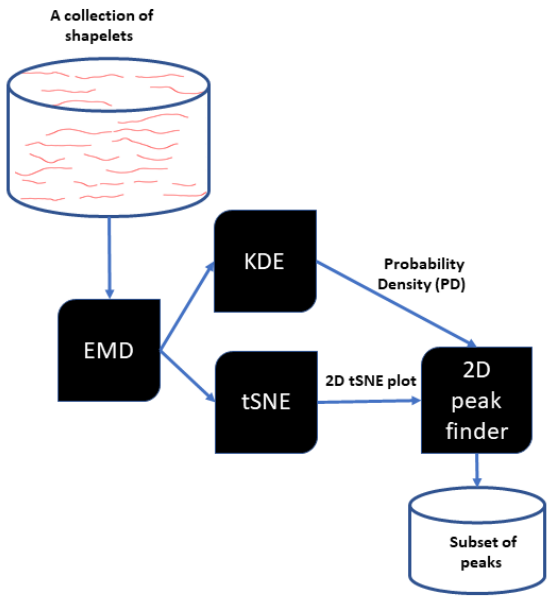

Fig. 1. Fig. 4. Our proposed technique for selecting representative shapes from a collection of local shapelets. By definition, statelets represent a subset of shapelets, so, this model handles the probability density of shapelets first and then applies them to approximate statelets. The method evaluates a subset of peaks that cover all the subgroups of a given pile of length-varying shapes. The distance matrix computed using EMD is used to plot tSNE and for computing probability density (PD). Subsequently, we weight the tSNE points by the PD's, and a 2D peak finder finds the peaks from each subgroup.

$$p(x)=\frac{1}{n}\sum_{i=1}^{n} K_h(x-x_i) \tag{7}$$

For D dimensions, the equation is simplified to following,

$$p(x) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{h^D}K\left(\frac{x-x_i}{h}\right) \tag{8}$$

where n is the number of samples, k_h is the kernel with a smoothing parameter h called the bandwidth. The kernel is a non-negative function and $h > 0$. For a smoother density model, the common practice uses the Gaussian kernel, simplifying the following kernel density model.

$$p(x)= \frac{1}{n}\sum_{i=1}^{n}\frac{1}{(2\pi h^2)^{\frac{D}{2}}}\, e^{-\frac{|x-x_i|^2}{2h^2}} \tag{9}$$

here h is the standard deviation of the Gaussian components. Another crucial part of the process is to select an appropriate bandwidth (h) for the density, and there are several strategies for selecting h [56, 57]. Our approach incorporates the rule-of-thumb bandwidth estimator for Gaussian [58]. The optimal choice for h is following,

$$h = \left(\frac{4\hat{\sigma}^5}{3n}\right)^{\frac{1}{5}} \approx 1.06\,\hat{\sigma}\, n^{\frac{-1}{5}} \tag{10}$$

here $\hat{\sigma}$ is the standard deviation of the samples, and n is the number of total samples. The term $(x - x_i)$ in the equation corresponds to the distance between the sample x and $x_{i \in n}$. So, we assign EMD distances (previously computed) between the shapes as a value of $(x - x_i)$ in the equation. Using equation (9) KDE generates probability density for all shapelets in the collection.

c) *tSNE using EMD score*

t-distributed stochastic neighbor embedding (tSNE) leads to a powerful and flexible visualization of high-dimensional data by giving each datapoint a location in a two or three-dimensional map [59, 60]. Decentralized stochastic neighbor embedding (dSNE) is used to separate the subgroups in the data using their distance metrics [61, 62]. tSNE considers each shape a data point in the subspace and uses EMD distances to select the neighbors for the embeddings. Then, the method weights all the points using their corresponding probability density. Fig. 5 shows all possible subgroups revealed by tSNE. The color indicates the probability density.

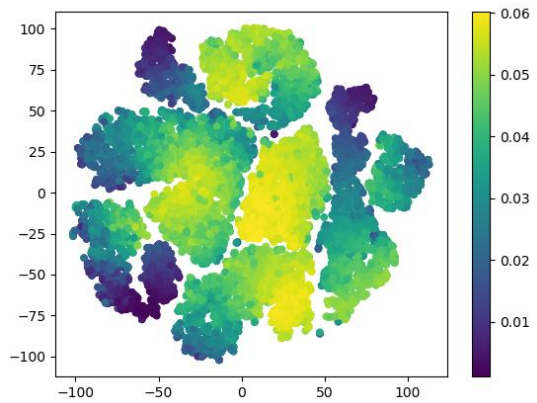

Fig. 5. The figure represents tSNE on a collection of shapelets, where each data point indicates a shapelet, and the weight (color) corresponds to the probability density computed using KDE.

After that, the method maps all tSNE points to a two-dimensional grid to accumulate the probability density of all closely neighboring shapes onto a single cell. It creates an erratic increment of frequency in the corresponding vicinity, making the peak points more divergent. The process discretizes the range of coordinates into a set of integer intervals. As a result, multiple points from the tSNE plot are stacked together into one cell.

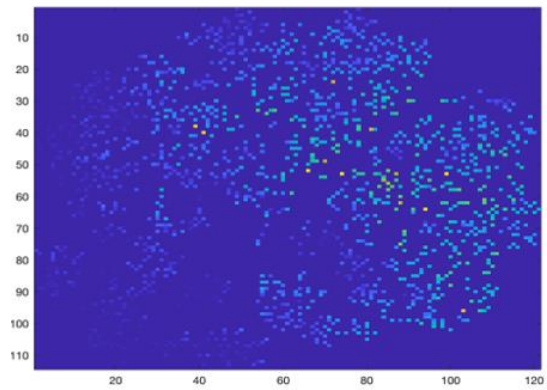

Fig. 6. A 2D grid view of the tSNE points after mapping them to a two-dimensional matrix. The probability density of close neighbors are accumulated in the process. Therefore, we observe the higher density data with a brighter color in the figure.

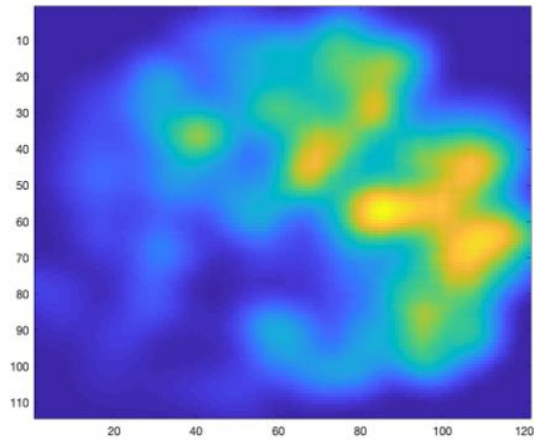

Fig. 7. After applying the Gaussian blur referred to in Fig. 6. The idea is to defocus less dense data points.

d) *Gaussian Blur and 2D peak finder*

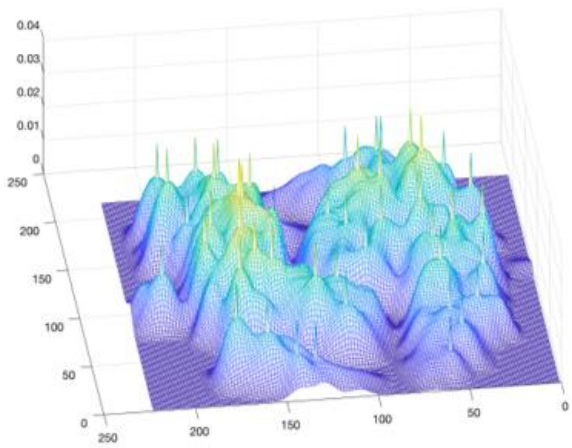

Fig. 8. A three-dimensional view of the tSNE plot after marking the peaks (extracted by 2D peak finder) from each subgroup. The peak finder selects at least one peak from each high-density region. Later, we map these peaks tSNE coordinates back to get real shapelets the points representing.

Defocusing noise helps extract significant features/data points representing the subgroup peaks. Subsequently, we blur the tSNE image and use a two-dimensional peak finder to compute the high-density data points. This extracts a characteristic shape from each subgroup with a higher probability density (PD). For more clarity, the method restricts the peak finding to a certain level to not end with a subset of similar shapes. The objective is to find a diversified set of shapes which well approximates the overall dynamics. These peaks are considered as statelets for that collection.

V. Experimental Results

Our method evaluates a subset of statelets from both groups of subjects (SZ/HC). The probability density of these high dimensional state shapes denote the replicability of those statelets across connections among multiple brain regions. In fig. 9, we show a subset of most frequent statelets from both groups. In Fig. 10, each bar indicates how often a pair of components has appeared in the group dynamics. They are also sorted based on their relative rank computed on their frequency in both patient and control groups. We can see that the pair (connections) rank is very different in each group, suggesting a diverse influence of connections driving the dynamics of patients vs. controls. The frequency of statelets in each connection in HC is higher than the SZ. To test the statistical significance of this difference between the two distributions, we computed a Kendall tau rank correlation coefficient, which is traditionally used to measure the ordinal association between two entities [63, 64]. The hypothesis test evaluates τ = 0.045373 with a p-value = 0.025 indicating a statistically significant difference.

A. *Dynamic features analysis*

We focus more on the dynamic features of the neural system, which have been mostly ignored in previous investigations of dFNC.

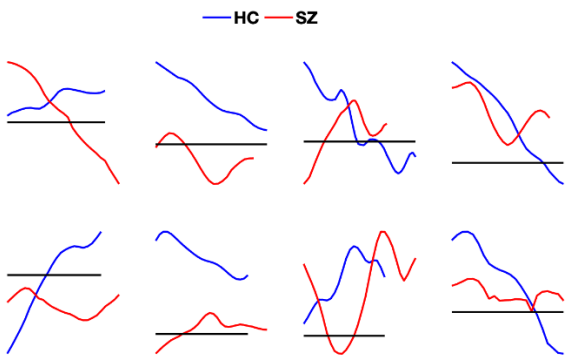

Fig. 9. A subset of dominant statelets from SZ and HC dynamics. Each subplot shows two shapes from both phenotypes sorted according to their overall probability density. The red and blue correspond to patient and control groups, respectively.

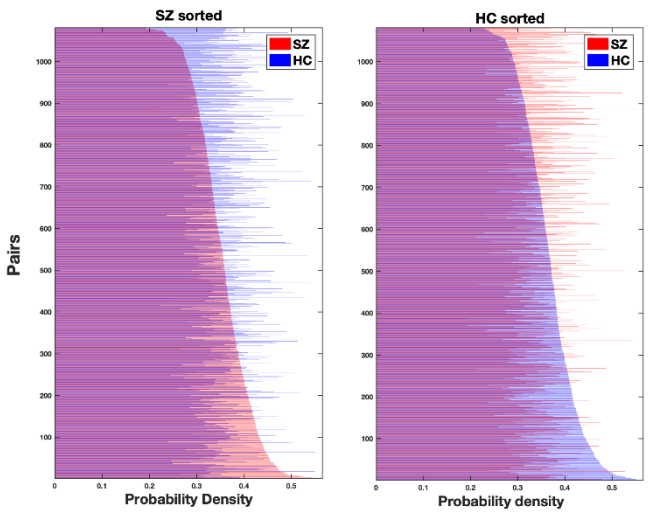

Fig. 10. Probability density (PD) of the connections demonstrates how frequent the statelets extracted from a connection appear in the entire group dynamic. Each connection has two density values – one per subject group. Blue demonstrates a probability density in controls (HC) and red in patients (SZ). In the left subplot, we sort the connections low to high according to their probability density in the SZ group, and HC sorts the right subplot. Y-axis represents the order of the links after sorting, and X-axis depicts their corresponding probability density. We observe that the rank of connections differs in both dynamics in terms of their PD, and we show this difference is statistically significant by hypothesis testing.

a) *Recurrence*

In fig. 11, we show the occurrences of the group statelets in individual dynamics (randomly selected three subjects from each group). We investigated the recurrence of statelets for all the subjects and got consistent results, as shown in the figure. As we can see, the group statelet repetition in HCs is significantly higher than SZs. Also, a smaller number of connections are present in the SZ dynamics than HC since the SZ group statelets can characterize a lower number of pairs since the links are more disrupted.

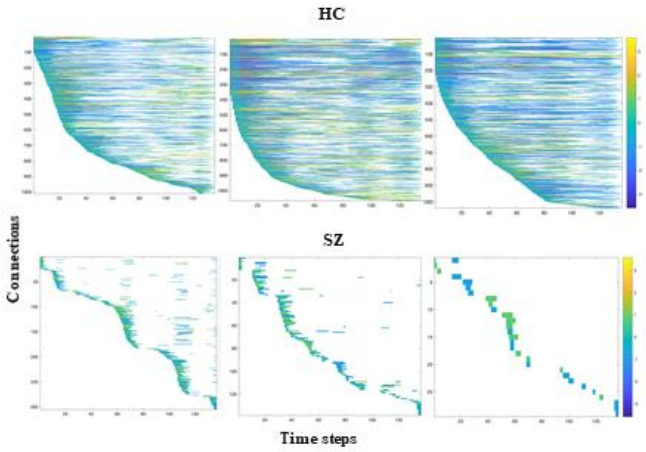

Fig. 2. Fig. 11: Each graph represents the occurrences of the most recurring statelets over each subject's time course. These are three randomly selected subjects from the HC (top) and SZ (bottom) group. We convolve the group statelets with the subject dynamics (all the dFNC time courses, 1081) to investigate the recurrence of the statelets over time. Then, we sorted the pairs based on the dominant shape's first occurrence in their time series. Consequently, the early the statelets appear in a pair's time course, the higher the pair/connection in that subject's dynamics—we threshold these convolutions matching scores at 0.8 for both groups to track down the strong appearances only. The color intensity corresponds to how strongly/weakly the shape appears in that part of the course. The color bar is identical for all the reference subjects in the figure.

Statelets are less recurring in patients and appear for once or twice during the course of the scan. On the other hand, HC statelets show up all over the time series. This graph also demonstrates the temporal modifications in statelets of a connection. The higher convolution value (brighter color) indicates the steady occurrence of the statelets, i.e., with similar shapes, but the lesser values stand for the statelets' distorted appearance. So, it explains the gradual modification in the state shapes with time and calrify the rebounding to the earlier form patterns.

b) *Time decay (TD): a consistency measure*

TD is a measure of how consistently a pair's connectivity exhibits the dominant shapelet early in the time course within each phenotypic group dynamics. The following equation depicts the idea of *TD*.

$$TD = \frac{1}{|S|}\sum_{i=1}^{|S|}\frac{1}{T_i} \tag{11}$$

Here, |S| = The total number of subjects in the group SZ/HC and T_i the index at which the pair has appeared for the first time in a subject (T_i = 1 to 136). If the pair is not present in a subject's dynamic, then $T_i = \infty$ thus, time decay for this pair in that subject is zero. TD can characterize a connection as consistently late (low TD value), consistently early (high TD value), pseudo-random positioning (medium TD value e.g.,

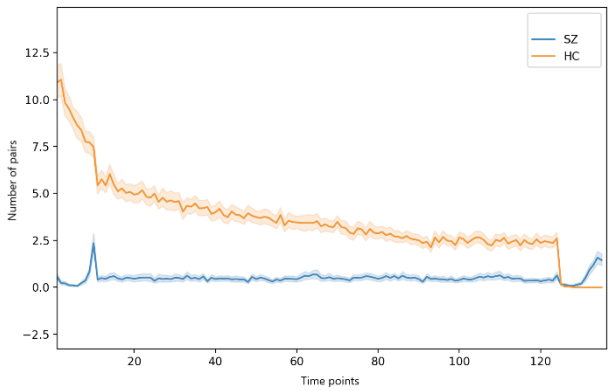

Fig. 12. From Fig. 10, we computed an aggregate appearance of the connections across all the subjects. The x-axis shows the time steps, and the Y-axis corresponds to the number of pairs /connections that show the first statelet at that step, which indicates the activation of that connection.

0.10 to 0.15), consistent but sparse in the group dynamics (0.5 – 0.09), etc.

c) *Modularity*

The line graphs in fig. 12 illustrates the ensemble behavior of the connections between a pair of networks in both dynamics (SZ/HC). At each time point, the value represents the number of connections with similar connectivity patterns. Throughout the scan session, comparatively more HC connections are activated in identical time steps than patient connections. This highlights the higher modularity in the HC network/graph (connection among the functional components of the brain). Modularity is imperative because it quantifies the strength of the network's division into modules, and each module consists of a subset of nodes, which are independent brain components in our case. These modules help data transmission by performing specific tasks independently. In addition, we

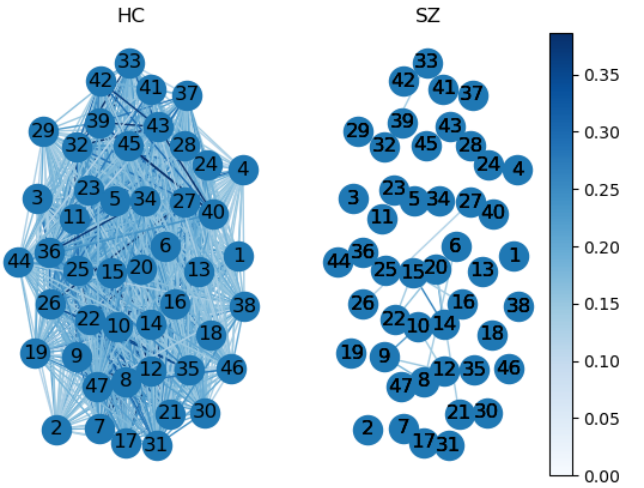

Fig. 13: Time decay graphs from both groups. The nodes are functional networks, edges correspond to the connection between them (maximum 1081 possible), and the weight represents mean time decay (TD) of a connection within a group dynamic. We compute color scaling from 95 percentile of the total values. After thresholding at average group mean (SZ group mean + HC group mean)/2, 1061 edges survive in the HC group and only 16 edges in the SZ group.

observe the SZ connections become modularized at the end of the time series whereas the HC ensembles are reduced by their size (fig. 12). This ensemble activation also demonstrates a distinct synchronization pattern between SZ and HC brain and suggests an interchanging control by various neural components.

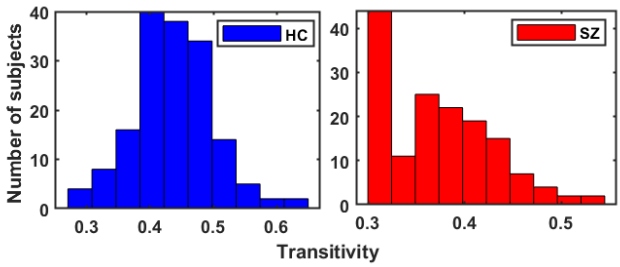

Fig.14. Histogram of transitivity from subject-level time decay graphs. It refers to the extent to which the relation between two nodes in a network connected by an edge is transitive. A significant portion of SZ subjects shows 0 transitivity, which means the connections are less consistent across different subjects. We show the differences are statistically significant suing a two-sample t-test on both distributions. Transitivity is also related to the clustering coefficient.

d) *Consistency graph*

Fig.13 demonstrates the average time decay graphs computed on both dynamics. Here, the vertices are independent components, and the edge represents the connection between two functional networks. The weight of the edge stands for the time decay, the darker, the higher. As we can see, HC functional networks are strongly connected, and most of the connections (out of 1081) pass the time decay benchmark. On the other hand, only a small subset of SZ connections converge to their group statelets. In fig.14, we show the histograms of transitivity of subjectwise time decay (TD) graphs in both groups. Transitivity symbolizes the global probability of the network (connectivity graph) to have adjacent nodes interconnected, revealing the existence of tightly connected communities, e.g., clusters and subgroups. The majority of the SZ graphs have zero or very low transitivity compared to the HC. We run a two sample t-test on the transitivity scores from both groups and the pvalue, p = 9.23e-88 indicates the difference is statistically significant at a 1 % significance level. The analysis indicates that HC networks intercommunicate more and consistently keep the channels active compared to SZ. This result is consistent with previous work, which suggested that the patients exhibit more erratic and less efficient communication among brain regions than controls. Fig. 15 shows the average time decay of each connection in both SZ and HC groups. The rightmost map shows HC-SZ group differences. All the pairs in HC show a more substantial time decay that SZ, especially in visual, sensorimotor, and default mode regions. All HC subjects show higher time decay than patients with schizophrenia.

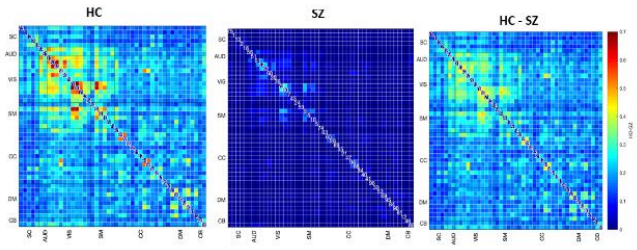

Fig. 15: Pairwise time decays in both patient and control groups and the HC - SZ group differences.

since the time decay metric shows group differences almost all over the brain, we intuitively make an effort to use it for classification. Therefore, we feed subject wise time decay data to multiple classifiers and observe significant improvement in the accuracy. We compare the performance between running the models directly on the dFNC matrix and time decay data. Fig. 16 shows how the models perform in both cases. We can see a simple SVM or logistic regression model using time decay outperforms all other methods, including LSTM with attention.

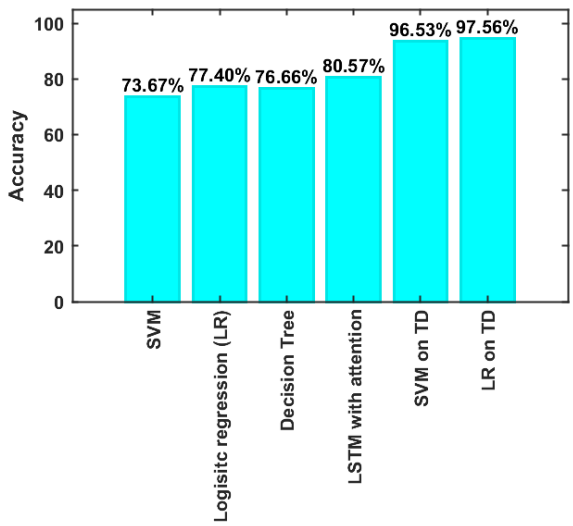

Fig. 16: Classification accuracy for different methods. First, three methods were applied to the dFNC matrix and LSTM with the attention model applied to the dFNC time course. The last two methods use time decay (TD) for classification. We run the models on time decay information of all the subjects for 100 repeated iterations, and the accuracies are mean across the iterations. We train the model on 200 random samples in each iteration and cross-validate them using the remaining 114 subjects.

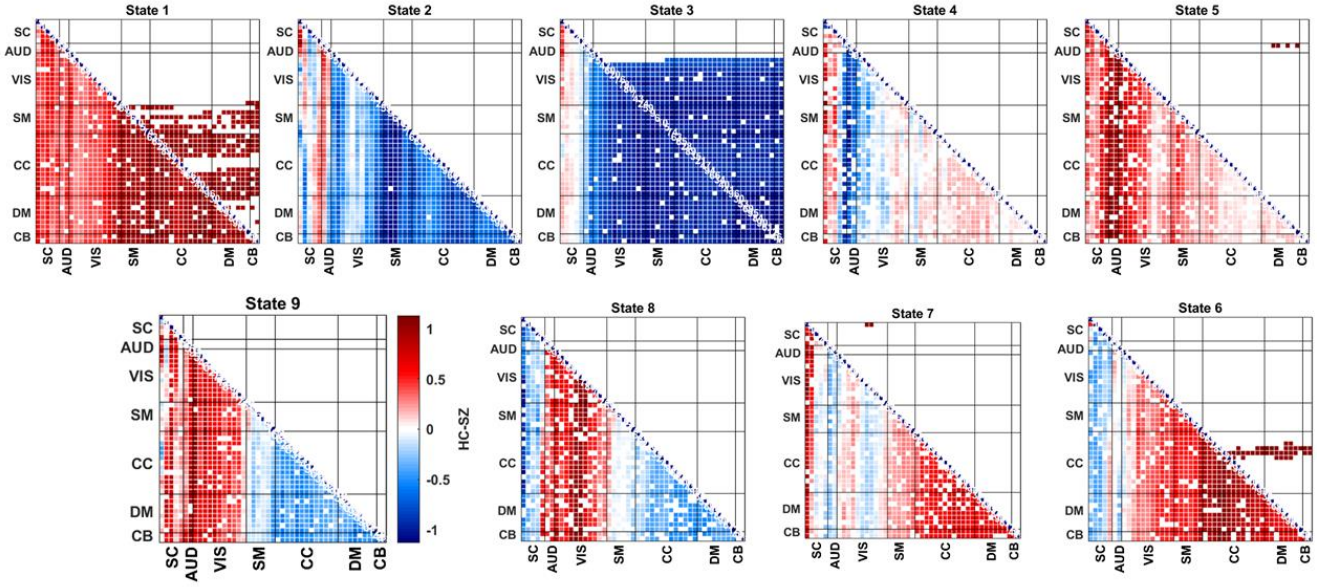

Fig. 17. HC - SZ group differences in terms of max connectivity strength. State-wise group differences in functional connectivity (FC). We have both SZ and HC subjects' groups at each state. For each pair of components, we have a subset of statelets from HC subjects and a subset from SZ subjects. Then, we compute the maximum connectivity strength of those statelets from both subgroups. A two-sample t-test using a null hypothesis of "No group difference" compares the max connectivity of patients vs. controls. A higher t-value indicates the rejection of the null hypothesis irrespective of their sign. However, the sign of t-values represents the directionality of the group difference. The pair matrix (47 x 47) is labeled into seven different domains subcortical (SB), auditory (AUD), visual (VIS), sensorimotor (SM), cognitive control (CC), default-mode (DM), cerebellar (CB) respectively. White cells in the matrix indicate either the absence of that pair or non-significant group differences for this pair within a state. The upper triangle represents the FDR corrected differences.

In another analysis, we focus on obtaining statelets across the dataset without separating SZ and HC subjects into two different bins. The idea is to extract the most frequent connectivity patterns from all the subjects, which is potentially extending the probability density of the statelets beyond the group boundary. In other words, this is to investigate how the shapelets represent the subject's dynamic irrespective of their group association in terms of approximating statelets for all the subgroups in the data. Here, we consider all the subjects (SZ and HC) together in one dynamic to evaluate the statelets. We estimate the probability density profile for the shapelets across all the subjects for a given connection. So, the density is no more group-specific, preferably throughout the dataset. We then run a tSNE on the collection of shapelets extracted from the whole dataset; the global tSNE yields nine different statelets. Based on those statelets, we generate nine connectivity states as shown in fig. 17. We use the EMD of extracted shapelets to determine their association towards a specific subgroup. Considering statelets as the centroid of the cluster, we assign all shapelets to the nearest (lowest EMD) subgroup. In fig. 17, we evaluate the HC-SZ group differences in connectivity strength across the states. For measuring the statistical significance of these differences, we use a two-sample t-test. The tvalues provide significant group differences in connectivity strength throughout the distinct regions of the brain. HC subjects in state 1 show stronger connections than SZ and the differences are statistically significant in sensorimotor (SM), cognitive control (CC), and default mode (DM) regions. In state 2 and 3, the differences are mostly HC < SZ; in state 3, group differences are statistically significant across the brain but indifferent strength in the connections between subcortical (SB) and other domains. State 4 shows mostly weaker differences except for a few components from auditory (AUD) and subcortical demonstrate SZ subjects are more strongly connected than HC in those regions. Two connections between AUD and DM show a statistically significant difference in state 5 and state 6 exhibits HC > SZ in CC region. Above all, states 7, 8, and 9 show significant connectivity differences in CC and DM regions. Specifically, state 7 depicts HC > SZ and 8 & 9 shows HC < SZ group distinction. In an earlier study [3], strong group differences are mostly in VIS and AUD, but here we got differences in SM, CC, and DB. We got both HC<SZ and SZ<HC in SM, CC, and DM. The most significant pattern of group differences we observe here is in CC, DM, and CB regions. This triangle shows the variation in connectivity between patients and controls across multiple states with an alternating directionality.

VI. Conclusion

We proposed a novel method for analyzing dynamic functional connectivity via extracting high-frequency texture from the connectivity space. Our main contribution is an unsupervised method of extracting the most dominant patterns from the data. Instead of clustering the shapelets, we use the local trends to generate global representations, which we term as statelets. Here, we use the probability density of shapelets from low dimensions to approximate the global frequency profile of connectivity patterns that are consistent with all dimensions. Pairwise tSNE (on the shapelets from single connections across the subjects) preserved the high-density shapes. The method continuously filters out the less prevalent patterns and leads to the analysis of a predictable subspace of the dynamic. Results show how these state movies help to investigate the dynamic properties of an inherently dynamic system (brain). The state-of-the-art methods for analyzing dFNC states assume a time-varying fixed set of connectivity without differentiating between predictable and trivial patterns. This assumption limits these methods' applicability in capturing the dynamic properties of the signal and impairs the analysis with insignificant information. Our approach ensures better interpretability in the results since the shapelets itself describes the properties of functional association between a pair of networks. The outcomes illustrate crucial differences between the healthy and schizophrenic brain in several dynamic properties like recurrence, modularity, synchronizability, etc. In a secondary analysis of dFNC states based on statelets, we observe group distinctive patterns in connectivity strengths. The difference in SM, CC, DM and, CB helps to highlight the dynamic disruptions in

schizophrenia in those regions. We observe that schizophrenia dynamics are more random, less recurring, and less commutable than healthy control dynamics. These statelets of connectivity among functional networks enrich our understanding of how networks communicate in spontaneous brain activity and the pattern of binding/impairment between them to maximize the efficiency of information transfer and minimize connection cost.

REFERENCES

1. Allen, E.A., et al., *Tracking whole-brain connectivity dynamics in the resting state.* Cerebral cortex, 2014. **24**(3): p. 663-676.
2. Sakoğlu, Ü., et al., *A method for evaluating dynamic functional network connectivity and task-modulation: application to schizophrenia.* Magnetic Resonance Materials in Physics, Biology and Medicine, 2010. **23**(5): p. 351-366.
3. Damaraju, E., et al., *Dynamic functional connectivity analysis reveals transient states of dysconnectivity in schizophrenia.* NeuroImage: Clinical, 2014. **5**: p. 298-308.
4. Liu, F., et al., *Dynamic functional network connectivity in idiopathic generalized epilepsy with generalized tonic–clonic seizure.* Human brain mapping, 2017. **38**(2): p. 957-973.
5. Ma, S., et al., *Dynamic changes of spatial functional network connectivity in healthy individuals and schizophrenia patients using independent vector analysis.* NeuroImage, 2014. **90**: p. 196-206.
6. Vergara, V.M., et al., *Dynamic functional network connectivity discriminates mild traumatic brain injury through machine learning.* NeuroImage: Clinical, 2018. **19**: p. 30-37.
7. Zhi, D., et al., *Aberrant dynamic functional network connectivity and graph properties in major depressive disorder.* Frontiers in psychiatry, 2018. **9**: p. 339.
8. Rashid, B., et al., *Dynamic connectivity states estimated from resting fMRI Identify differences among Schizophrenia, bipolar disorder, and healthy control subjects.* Frontiers in human neuroscience, 2014. **8**: p. 897.
9. Bullmore, E. and O. Sporns, *Complex brain networks: graph theoretical analysis of structural and functional systems.* Nature reviews neuroscience, 2009. **10**(3): p. 186-198.
10. McKenna, T., T. McMullen, and M. Shlesinger, *The brain as a dynamic physical system.* Neuroscience, 1994. **60**(3): p. 587-605.
11. Siebenhühner, F., et al., *Intra- and inter-frequency brain network structure in health and schizophrenia.* PloS one, 2013. **8**(8): p. e72351-e72351.
12. Yu, Q., et al., *Brain Connectivity Networks In Schizophrenia Underlying Resting State Functional Magnetic Resonance Imaging.* Current topics in medicinal chemistry, 2012.
13. Li, J., et al., *High transition frequencies of dynamic functional connectivity states in the creative brain.* Scientific reports, 2017. **7**: p. 46072.
14. Atluri, G., et al., *Complex biomarker discovery in neuroimaging data: Finding a needle in a haystack.* NeuroImage: Clinical, 2013. **3**: p. 123-131.
15. Kong, X. and P.S. Yu, *Brain network analysis: a data mining perspective.* ACM SIGKDD Explorations Newsletter, 2014. **15**(2): p. 30-38.
16. Nielsen, F.Å., L.K. Hansen, and D. Balslev, *Mining for associations between text and brain activation in a functional neuroimaging database.* Neuroinformatics, 2004. **2**(4): p. 369-379.
17. Ye, L. and E. Keogh. *Time series shapelets: a new primitive for data mining*. in *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. 2009.
18. Ghalwash, M.F. and Z. Obradovic, *Early classification of multivariate temporal observations by extraction of interpretable shapelets.* BMC bioinformatics, 2012. **13**(1): p. 195.
19. Ghalwash, M.F., V. Radosavljevic, and Z. Obradovic. *Extraction of interpretable multivariate patterns for early diagnostics*. in *2013 IEEE 13th International Conference on Data Mining*. 2013. IEEE.
20. Rakthanmanon, T., et al. *Time series epenthesis: Clustering time series streams requires ignoring some data*. in *2011 IEEE 11th International Conference on Data Mining*. 2011. IEEE.
21. Miller, R.L., et al., *Higher dimensional meta-state analysis reveals reduced resting fMRI connectivity dynamism in schizophrenia patients.* PloS one, 2016. **11**(3).
22. Saha, D.K., et al. *Classification As a Criterion to Select Model Order For Dynamic Functional Connectivity States in Rest-fMRI Data*. in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. 2019. IEEE.
23. Rahaman, M.A., et al., *A novel method for tri-clustering dynamic functional network connectivity (dFNC) identifies significant schizophrenia effects across multiple states in distinct subgroups of individuals.* bioRxiv, 2020: p. 2020.08.06.239152.
24. Morioka, H., V. Calhoun, and A. Hyvärinen, *Nonlinear ICA of fMRI reveals primitive temporal structures linked to rest, task, and behavioral traits.* NeuroImage, 2020: p. 116989.
25. Węglarczyk, S. *Kernel density estimation and its application*. in *ITM Web of Conferences*. 2018. EDP Sciences.
26. Calhoun, V.D., et al., *A method for multitask fMRI data fusion applied to schizophrenia.* Human brain mapping, 2006. **27**(7): p. 598-610.
27. Gur, R.E., et al., *An fMRI study of facial emotion processing in patients with schizophrenia.* American Journal of Psychiatry, 2002. **159**(12): p. 1992-1999.
28. Swanson, N., et al., *Lateral differences in the default mode network in healthy controls and patients with schizophrenia.* Human brain mapping, 2011. **32**(4): p. 654-664.
29. Stephan, K.E., T. Baldeweg, and K.J. Friston, *Synaptic plasticity and dysconnection in*

schizophrenia. Biological psychiatry, 2006. **59**(10): p. 929-939.
30. Williamson, P. and J.M. Allman, *A framework for interpreting functional networks in schizophrenia.* Frontiers in human neuroscience, 2012. **6**: p. 184.
31. Calhoun, V.D., J. Liu, and T. Adalı, *A review of group ICA for fMRI data and ICA for joint inference of imaging, genetic, and ERP data.* Neuroimage, 2009. **45**(1): p. S163-S172.
32. Du, Y., et al., *NeuroMark: an adaptive independent component analysis framework for estimating reproducible and comparable fMRI biomarkers among brain disorders.* medRxiv, 2019: p. 19008631.
33. Erhardt, E.B., et al., *Comparison of multi‐subject ICA methods for analysis of fMRI data.* Human brain mapping, 2011. **32**(12): p. 2075-2095.
34. Du, Y., et al., *NeuroMark: a fully automated ICA method to identify effective fMRI markers of brain disorders.* medRxiv, 2019: p. 19008631.
35. Tsuang, M.T., M.J. Lyons, and S.V. Faraone, *Heterogeneity of Schizophrenia: Conceptual Models and Analytic Strategies.* British Journal of Psychiatry, 1990. **156**(1): p. 17-26.
36. Alnæs, D., et al., *Brain Heterogeneity in Schizophrenia and Its Association With Polygenic Risk.* JAMA Psychiatry, 2019. **76**(7): p. 739-748.
37. Luchins, D.J., D.R. Weinberger, and R.J. Wyatt, *Schizophrenia: Evidence of a Subgroup With Reversed Cerebral Asymmetry.* Archives of General Psychiatry, 1979. **36**(12): p. 1309-1311.
38. Rahaman, M.A., et al., *N-BiC: A Method for Multi-Component and Symptom Biclustering of Structural MRI Data: Application to Schizophrenia.* IEEE Transactions on Biomedical Engineering, 2020. **67**(1): p. 110-121.
39. Scarr, E., et al., *Decreased cortical muscarinic receptors define a subgroup of subjects with schizophrenia.* Molecular Psychiatry, 2009. **14**(11): p. 1017-1023.
40. Zhang, L., et al., *Tracking the Main States of Dynamic Functional Connectivity in Resting State.* Frontiers in neuroscience, 2019. **13**: p. 685.
41. Keator, D.B., et al., *The Function Biomedical Informatics Research Network Data Repository.* NeuroImage, 2016. **124**(Pt B): p. 1074-1079.
42. Freire, L. and J.-F. Mangin, *Motion correction algorithms may create spurious brain activations in the absence of subject motion.* NeuroImage, 2001. **14**(3): p. 709-722.
43. Hyvärinen, A. and E. Oja, *Independent component analysis: algorithms and applications.* Neural networks, 2000. **13**(4-5): p. 411-430.
44. Calhoun, V.D., et al., *A method for making group inferences from functional MRI data using independent component analysis.* Hum Brain Mapp, 2001. **14**(3): p. 140-51.
45. Smith, S.M., et al., *Network modelling methods for FMRI.* Neuroimage, 2011. **54**(2): p. 875-891.
46. Varoquaux, G., et al. *Brain covariance selection: better individual functional connectivity models using population prior*. in *Advances in neural information processing systems*. 2010.
47. Friedman, J., T. Hastie, and R. Tibshirani, *Sparse inverse covariance estimation with the graphical lasso.* Biostatistics, 2008. **9**(3): p. 432-441.
48. Vallender, S.S., *Calculation of the Wasserstein Distance Between Probability Distributions on the Line.* Theory of Probability & Its Applications, 1974. **18**(4): p. 784-786.
49. Levina, E. and P. Bickel. *The earth mover's distance is the mallows distance: Some insights from statistics*. in *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*. 2001. IEEE.
50. Champion, T., L. De Pascale, and P. Juutinen, *The ∞-Wasserstein Distance: Local Solutions and Existence of Optimal Transport Maps.* SIAM Journal on Mathematical Analysis, 2008. **40**(1): p. 1-20.
51. Andoni, A., P. Indyk, and R. Krauthgamer. *Earth mover distance over high-dimensional spaces*. in *Proceedings of the nineteenth annual ACM-SIAM symposium on Discrete algorithms*. 2008. Society for Industrial and Applied Mathematics.
52. Weglarczyk, S., *Kernel density estimation and its application.* ITM Web of Conferences, 2018. **23**: p. 00037.
53. Rubner, Y., C. Tomasi, and L.J. Guibas, *The Earth Mover's Distance as a Metric for Image Retrieval.* Int. J. Comput. Vision, 2000. **40**(2): p. 99-121.
54. Sheather, S. and C. Jones, *A reliable data-based bandwidth selection method for kernel density estimation.* Journal of the Royal Statistical Society: Series B (Statistical Methodology), 1991. **53**(3): p. 683-690.
55. Silverman, B.W., *Density estimation for statistics and data analysis*. 1986, London; New York: Chapman and Hall.
56. Ahmad, I.A. and M. Amezziane, *A general and fast convergent bandwidth selection method of kernel estimator.* Journal of Nonparametric Statistics, 2007. **19**(4-5): p. 165-187.
57. Turlach, B.A. *Bandwidth selection in kernel density estimation: A review*. in *CORE and Institut de Statistique*. 1993. Citeseer.
58. Scott, D.W., *Multivariate density estimation and visualization*, in *Handbook of computational statistics*. 2012, Springer. p. 549-569.
59. Gisbrecht, A., A. Schulz, and B. Hammer, *Parametric nonlinear dimensionality reduction using kernel t-SNE.* Neurocomputing, 2015. **147**: p. 71-82.
60. Maaten, L.v.d. and G. Hinton, *Visualizing data using t-SNE.* Journal of machine learning research, 2008. **9**(Nov): p. 2579-2605.
61. Saha, D.K., et al. *See without looking: joint visualization of sensitive multi-site datasets*. in *IJCAI*. 2017.
62. Saha, D.K., et al., *dSNE: a visualization approach for use with decentralized data.* BioRxiv, 2019: p. 826974.

63. Kendall, M.G., *A New Measure of Rank Correlation.* Biometrika, 1938. **30**(1/2): p. 81-93.
64. Puka, L., *Kendall's Tau*, in *International Encyclopedia of Statistical Science*, M. Lovric, Editor. 2011, Springer Berlin Heidelberg: Berlin, Heidelberg. p. 713-715.