Multi-Kernel Capsule Network for Schizophrenia Identification

Tian Wang, Anastasios Bezerianos, *Senior Member, IEEE*, Andrzej Cichocki, *Fellow, IEEE*, and Junhua Li*, *Senior Member, IEEE*

***Abstract*— Schizophrenia seriously affects the quality of life. To date, both simple (e.g., linear discriminant analysis) and complex (e.g., deep neural network) machine learning methods have been utilized to identify schizophrenia based on functional connectivity features. The existing simple methods need two separate steps (i.e., feature extraction and classification) to achieve the identification, which disables simultaneous tuning for the best feature extraction and classifier training. The complex methods integrate two steps and can be simultaneously tuned to achieve optimal performance, but these methods require a much larger amount of data for model training. To overcome the aforementioned drawbacks, we proposed a multi-kernel capsule network (MKCapsnet), which was developed by considering the brain anatomical structure. Kernels were set to match with partition sizes of brain anatomical structure in order to capture interregional connectivities at the varying scales. With the inspiration of widely-used dropout strategy in deep learning, we developed capsule dropout in the capsule layer to prevent overfitting of the model. The comparison results showed that the proposed method outperformed the state-of-the-art methods. Besides, we compared performances using different parameters and illustrated the routing process to reveal characteristics of the proposed method. MKCapsnet is promising for schizophrenia identification. Our study first utilized capsule neural network for analyzing functional connectivity of magnetic resonance imaging (MRI) and proposed a novel multi-kernel capsule structure with consideration of brain anatomical parcellation, which could be a new way to reveal brain mechanisms. In addition, we provided useful information in the parameter setting, which is informative for further studies using a capsule network for other neurophysiological signal classification.**

***Index Terms*—Multi-Kernel Capsule Network (MKCapsnet), Schizophrenia Diagnosis, Brain Connectivity, Deep Learning (DL), Functional Magnetic Resonance Imaging (fMRI)**

I. INTRODUCTION

SCHIZOPHERNIA is among the most universal psychiatric disorders, affecting about 1% of the population worldwide [1], [2]. Patients with schizophrenia may have deficits in attention, memory, and behavior [3]. At present, schizophrenia diagnosis relies on the qualitative examination of obvious mental symptoms and patients' self-reported experience, which is not feasible to detect disease at the early stage. Machine learning technique could help the diagnosis by disease detection based on neurophysiological signals [4]-[9], even for the prediction of disease development [10]. As shown in the published papers [11]-[16], machine learning technique might be able to detect subtle abnormality at the early stage of schizophrenia. Diverse features were fed into machine learning models to distinguish patients with schizophrenia from healthy controls. Volumetric features and tissue density features extracted from gray matter, white matter, and cerebrospinal fluid areas were used in the early studies of schizophrenia detection [17]-[19]. The investigation concentration was shifted to functional connectivity. Based on the observations of functional connecitivty in patients with schizophrenia, the phenomenon of functional dysconnectivity among disparate brain regions exists [20]-[24]. This functional dysconnectivity exhibited connectivity strength abnormalities between brain regions, which can be used to distinguish patients with schizophrenia from healthy people by machine learning methods [25]-[33]. Up to now, both simple and complex methods have been employed for schizophrenia identification and achieved good performance based on the functional magnetic resonance imaging data. For instance, Li et al. assessed all connectivity features to select top discriminative features and employed simple methods, such as linear

This work was supported by the National Natural Science Foundation of China (No. 61806149) and the Guangdong Basic and Applied Basic Research Foundation (No. 2020A1515010991), and also supported by the Ministry of Education and Science of the Russian Federation (No. 14.756.31.0001) and Polish National Science Center (No. UMO-2016/20/W/NZ4/00354).

Wang Tian is with the Laboratory for Brain-Bionic Intelligence and Computational Neuroscience, Wuyi University, Jiangmen, 529020, China, and also with the Singapore Institute for Neurotechnology, National University of Singapore, Singapore, 117456, Singapore.

Anastasios Bezerianos is with the Hellenic Institute of Transport (HIT), Centre for Research and Technology Hellas (CERTH), Thermi Thessaloniki, 57001, Greece.

Andrzej Cichocki is with the Skolkovo Institute of Science and Technology (Skoltech), Moscow, 121205, Russia, and also with the Nicolaus Copernicus University (UMK), Torun, 87-100, Poland.

Junhua Li is with the School of Computer Science and Electronic Engineering, University of Essex, Colchester, CO4 3SQ, UK, also with the Laboratory for Brain-Bionic Intelligence and Computational Neuroscience, Wuyi University, Jiangmen, 529020, China, also with the Singapore Institute for Neurotechnology, National University of Singapore, Singapore, 117456, Singapore, and also with the Centre for Multidisciplinary Convergence Computing, School of Computer Science and Engineering, Northwestern Polytechnical University, Xi'an, 710072, China.

* indicates the corresponding author (e-mail: juhalee.bcmi@gmail.com, junhua.li@essex.ac.uk).

discriminative analysis, to perform schizophrenia classification [29]. Other studies using complex methods (e.g., Deep Neural Network, DNN) achieved a better performance in schizophrenia classification according to the comparison results [26]. However, these complex methods require a large amount of data for model training in order to reach such better performance. In practice, the scale of available data is usually not enough to meet the requirement due to a variety of factors including a limited number of participants and expensive cost in data collection.

Very recently, a new type of network called capsule neural network was proposed by Sabour et al., which does not require huge data for model training and could achieve good performance [34]. Capsule neural network was proposed to initially aim for classifying handwritten digits of postcodes and has now been extended to image recognition and text mining [35]-[41]. All these studies demonstrated that capsule neural network has advantages over other methods. For instance, capsule neural network outperformed convolutional neural network (CNN) in the recognition of brain tumour types based on the data of magnetic resonance imaging (MRI) [42]. The CNN is good at capturing elements (e.g., objects) that are even spatially variable, but it is not capable of the learning of relationships between elements. This shortcoming is overcome in the model of capsule neural network. However, capsule neural network is still not perfect. It was proposed for image classification and was not designed for brain disease detection. As we know, the brain can be structurally divided into small brain regions. These small brain regions constitute larger areas that act brain functionalities. The numbers of small brain regions constituting the larger areas are not consistent. Some larger areas (e.g., frontal area) comprise a greater number of small brain regions (hereinafter, size refers to the number of small brain regions) while others may have less number of small brain regions (e.g., subcortical area). Different sizes of the larger areas and the hierarchy in the brain parcellation represent brain anatomical structure. Without considering the brain anatomical structure, the capsule neural network could not achieve a good performance in schizophrenia identification.

To this end, we proposed a multiple-kernel capsule network, in which the multi-kernel was designed in line with the varying sizes of the larger areas. Each kernel was intended to capture the information of a particular size of the larger area and the relationships between the larger areas at the same scale. To prevent the overfitting of learning and improve the training efficiency of the model, dropout strategy is frequently utilized. In our model, we investigated this strategy but proposed capsule dropout to maximise the benefit, which appeared to be more suitable compared to the scalar dropout and vector dropout in the context of our study. In this paper, we also explored parameter optimisation and gave informative results and discussions, which provide insights into how the performance was changed with different settings and could inform other researchers of parameter tuning and parameter determination. Moreover, we visualized the routing process to demonstrate the model learning and reveal the engagements of brain parcellation-corresponding kernels.

In this study, the proposed model was compared to not only the methods that had been used in the functional connectivity-based schizophrenia identification (i.e., k-Nearest Neighbours, k-NN; Linear Discriminant Analysis, LDA; Linear Support Vector Machine, L-SVM; Support Vector Machine with Radial Basis Function kernel, RBF-SVM; and Deep Neural Network, DNN), but also the methods that achieved excellent performance in other classification problems (i.e., Random Forest, RF; Gradient Boosting Machine, GBM; Graph Convolutional Network, GCN; Long Short-Term Memory, LSTM; and Generative Adversarial Network, GAN). All methods were assessed using the same publicly available dataset[1]. The performance comparison was done in terms of average classification accuracy obtained by the 10-fold cross-validation.

II. Methods

A. Evaluation Dataset

All methods were evaluated using a publicly available dataset consisting of 148 participants, which was collected by the Center for Biomedical Research Excellence [43]. High-resolution T1-weighted MRI and resting-state functional magnetic resonance imaging (fMRI) scans were collected by a 3-Tesla Siemens Trio scanner. The High resolution T1-weighted MRI was collected with the utilization of a multi-echo magnetization-prepared rapid gradient echo sequence (repetition time (TR) = 2.53 s, echo time (TE) = [1.64, 3.5, 5.36, 7.22, 9.08] ms, flip angle = 7°, slab thickness = 176 mm, field of view (FOV) = 256×256 mm, acquisition matrix = 256×256, voxel resolution = 1×1×1 mm^3). The resting-state fMRI data were obtained by single-shot full k-space echo-planar imaging (EPI) with the inter-commissural line as a reference (TR = 2 s, TE = 29 ms, matrix size = 64×64, slices = 33, voxel resolution = 3×3×4 mm^3).

B. Data Preprocessing

The MRI data were preprocessed using three toolboxes: (1) statistical parametric mapping (SPM12), (2) resting-state fMRI data analysis toolkit [44], and (3) data processing assistant for resting-state fMRI advanced edition [45] in the environment of MATLAB (Mathworks, Inc., Natick, Massachusetts, USA). Three participants were excluded from the preprocessing procedure because of unavailable category information (there is no label to recognize whether the data were from a patient) or a short length of volume scanning, resulting in 145 participants. Additional 14 participants were removed due to excessive head movements (i.e., the maximal inter-scan motion exceeded 2.5 mm translation or 2.5 degrees rotation in any direction). This exclusion resulted in 60 patients with schizophrenia and 71 healthy controls. After the preprocessing procedure comprising

[1] The dataset used in this study can be obtained at http://fcon_1000.projects.nitrc.org/indi/retro/cobre.html (released by the Center for Biomedical Research Excellence).

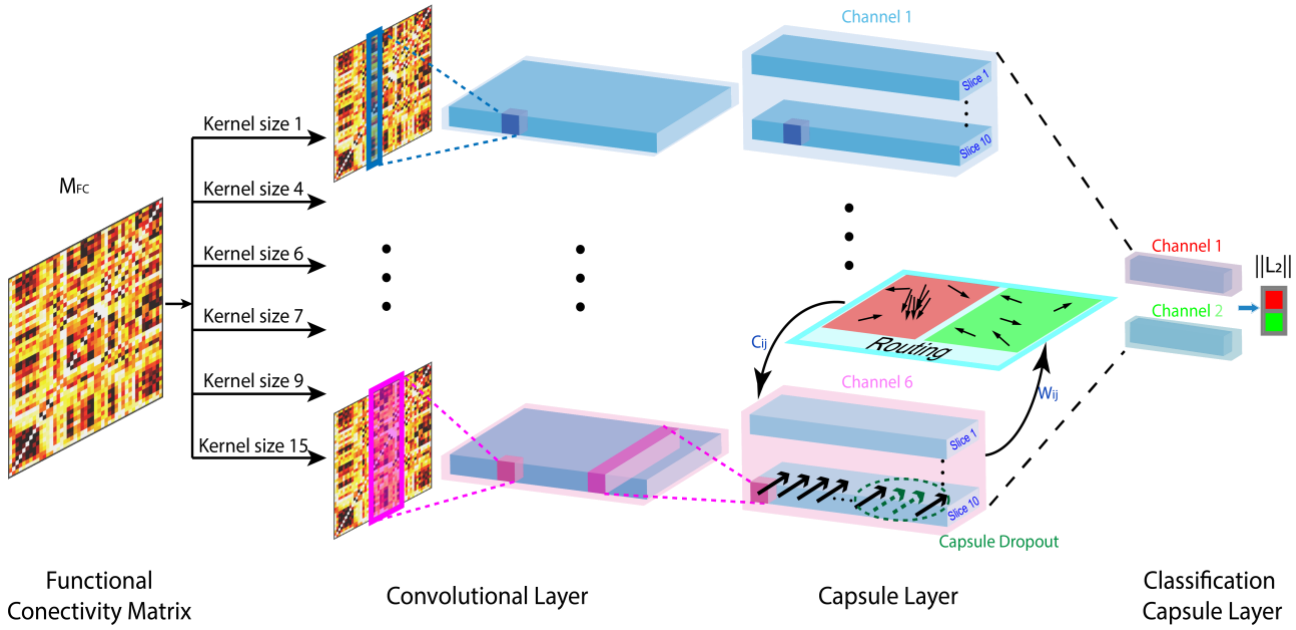

Fig. 1. Model structure of multi-kernel capsule network. The model consists of convolutional layer, capsule layer, and classification capsule layer. In the capsule layer, capsule dropout strategy was embedded inside each channel. The dropout was separately set for each channel and the dropout rate (50 %) was identical for all channels. Routing agreement algorithm was used to learn based on capsules. The illustration of the routing is shown in Fig.3.

volume removal, motion correction, slice timing correction, spatial normalization, signal regression with the regressors of 24 head motion parameters, cerebrospinal fluid, and white matter, temporal band-pass filtering with cut-off frequencies of 0.01Hz and 0.08Hz and spatial smoothing, a parcellation atlas named automated anatomical labelling (AAL) was applied to parcellate the brain into 116 regions of interest (ROIs) [46]. Pearson correlation was subsequently utilized to estimate connectivity strengths for all pairs of ROIs. Fisher's r-to-z transformation was then applied to improve the normality of connectivity strength values. All values were assembled to form a functional connectivity matrix, representing as M_{FC}, which is the input of all compared models.

C. *Multi-Kernel Capsule Network*

Fig. 1 depicts the structure of the proposed model, namely the multi-kernel capsule network (MKCapsnet). It consists of three layers: convolutional layer, capsule layer, and classification capsule layer. We set six convolutional kernels (i.e., kernel sizes: 1, 4, 6, 7, 9, and 15 columns) with diverse sizes in the first layer to match with varying region sizes of anatomical parcellation of the brain (e.g., a kernel size of 4 corresponds to the subcortical area that comprise 4 small brain regions and a kernel size of 9 corresponds to the cerebellum area that comprise 9 small brain regions. The illustration of examples of brain areas to which each kernel corresponds can be found in the Fig. S1). In the following layer (i.e., capsule layer), the extracted connectivity information from the first layer is represented as vectors (known as capsules), whose directions stand for attributes and whose lengths indicate the probabilities of being each attribute. These vectors are assigned to six channels corresponding to six kernels we set. Inspired by the dropout strategy in deep learning, we designed a capsule dropout strategy in the capsule layer, where the routing agreement algorithm is employed to learn based on capsules. Finally, the margin loss is utilized in the classification capsule layer to update weights by backpropagation process.

Once the functional connectivity matrix M_{FC} was obtained from the pre-processing procedure, it was fed into the developed deep learning model. The columns in the M_{FC} were convoluted using kernels with different sizes. For each kernel, the output of convolutional layer is fed into capsule layer as vectors (i.e., u_i for the ith capsule in Capsule Layer) which is considered as the input of capsule layer. The vector u_i is transformed (by the weighting matrix W_{ij}) into a predicted vector $\hat{u}_{j|i}$ to predict the output of the capsule i corresponding to higher level capsule j (i.e., u_j for capsule j in Classification Capsule Layer). The calculation process is as follows:

$$\hat{u}_{j|i} = W_{ij}\, u_i \qquad (1)$$

where the weighting matrix W_{ij} is updated when the whole routing process was finished (several routing iterations), by the backpropagation process with the support of $L2$-norm margin loss (as shown in Function (6)) if u_i is not dropped by the dropout strategy. The input of capsule j in the Classification Capsule Layer is a weighted summation of all the predicted vectors from the capsules in Capsule layer, obtaining by

$$s_j = \sum_{i \in I} c_{ij}\, \hat{u}_{j|i} \qquad (2)$$

where c_{ij} is a coupling coefficient, representing the routing coefficient from the lower level capsule i to the higher level capsule j and is updated every routing step as shown in Fig. 3. $I = \{1,2,3,\cdots,i,\cdots\}$, which is the set of all capsules in the Capsule Layer. The coupling coefficient c_{ij} is determined by a softmax function as follows,

$$c_{ij} = \exp(b_{ij}) / \sum_{j \in J} \exp(b_{ij}) \quad (3)$$

where b_{ij} is a logarithmic prior probability that capsule i is coupled to capsule j, which is iteratively updated during the routing process (b_{ij} are initialized to 0 at the first step). $J = \{1,2,3,\cdots,j,\cdots\}$, which is the set of all capsules in the Classificaion Capsule Layer.

$$b_{ij} \leftarrow b_{ij} + \hat{u}_{j|i} \cdot v_j \quad (4)$$

where v_j is the output vector of capsule j and obtained by a non-linear 'squashing' function as follows,

$$v_j = \frac{\|s_j\|^2}{1+\|s_j\|^2} \frac{s_j}{\|s_j\|} \quad (5)$$

This step normalizes the length of the output vector to be within the range [0, 1]. The above coefficient update procedure is called routing. After that, the outputs of the capsule layer are inputted into the subsequent classification capsule layer. The number of capsules in this layer is the same as the number of classes. $L2$-norm margin loss is utilized to update the model parameters. The total $L2$-norm margin loss is the summation of the losses of all capsules, calculating by

$$\begin{aligned} L2 &= \sum_{j \in J} L2_j \\ &= \sum_{j \in J} T_j \ \max\left(0, m^+ - \|v_j\|\right)^2 \\ &\quad + \sum_{j \in J} \lambda\left(1 - T_j\right) \ \max\left(0, \|v_j\| - m^-\right)^2 \end{aligned} \quad (6)$$

where T_j is 1 if and only if the class j is present. m^+ and m^- were set as 0.9 and 0.1 respectively. The down-weighting factor λ was set as 0.5.

During the training phase, the weighted matrix is updated every training trails according to the backpropagation process which is used to minimize the total loss. The update based on the routing process is taking within a single trail. During the testing phase, the lengths of the capsules were calculated to have probabilities of being each class. The class with the largest probability is the class of the sample.

TABLE I
PARAMETERS SETTINGS IN THE TRAINING

Parameter	Setting
Epoch Number	500
Learning Rate	0.01
Batch Size	3
Dropout Rate	0.5
Activation Function	Softplus
Early Stop Criterion	0.008

III. RESULTS

A. Model Architecture Comparison

In general, deep learning models have many parameters, which influence the performance of the models strongly. These parameters need fine tuning to obtain a desirable performance. In our proposed model, the parameters can be grouped into two categories: model training parameters and model architecture parameters. The model training parameters were tuned by the means of grid search, and the model architecture parameters were optimized by maximizing classification performance. The final settings of model training parameters are listed in Table I. The optimal model architecture settings are shown in Table II.

TABLE II
PARAMETERS OF LAYER SETTING USED IN THE PROPOSED MODEL

Layer	Type	Kernel Size	Stride	Filter/Slice Size	Channel	Vector Length
1	Convolution	[{1, 4, 6, 7, 9, 15}*,108]	1	64 filters	-	-
2	Capsule	[1, 1, 64]	1	10 slices	6	20
3	Capsule	-	-	-	2	20

* Kernel sizes {1, 4, 6, 7, 9, 15} in the convolution layer correspond to the brain anatomical parcellation.

TABLE III
PERFORMANCE COMPARISON AMONG VARIOUS MODEL ARCHITECTURE SETTINGS

Row	Dropout Strategy	Kernel Type	Multi-Slice Channel	Loss Norm	Accuracy	Sensitivity	Specificity
1	No	Column(size 1)	X	L2	77.14%	80.36%	73.33%
2	No	Column(size 15)	X	L2	76.32%	83.03%	68.33%
3	No	Square(size 15)	X	L2	42.75%	44.64%	40.00%
4	Scalar•	Column(size 1)	X	L2	77.14%	78.93%	**75.00%**
5	Vector^	Column(size 1)	X	L2	77.97%	84.82%	70.00%
6	Capsule*	Column(size 1)	X	L2	78.57%	81.43%	**75.00%**
7	Capsule	Column(size 4)	X	L2	77.09%	83.04%	70.00%
8	Capsule	Column(size 6)	X	L2	77.86%	83.04%	71.67%
9	Capsule	Column(size 7)	X	L2	78.63%	81.61%	**75.00%**
10	Capsule	Column(size 9)	X	L2	77.09%	83.04%	70.00%
11	Capsule	Column(size 15)	X	L2	78.63%	83.04%	73.33%
12	Capsule	Square(size 15)	X	L2	63.19%	71.61%	53.33%
13	Capsule	Multiple	X	L2	80.88%	**88.57%**	71.67%
14	Capsule	Multiple	√	L1	69.34%	75.89%	61.67%
15	Capsule	Multiple	√	L2	**82.42%**	**88.57%**	**75.00%**

• Scalar Dropout: the dropout was performed on the elements of vectors.
^ Vector Dropout: the dropout was not separately set for each channel and a dropout rate of 50% was applied to the channels together
* Capsule Dropout: the dropout was separately set for each channel and the dropout rate of 50% was identical for all channels.

In order to provide insights how the performance was changed with different settings of parameters in the model architecture, we compared performances in different architecture settings (i.e., different dropout strategies, different kernel types, with or without multi-slice channel, and applying different loss norms). The comparison results were listed in Table III (all accuracies were obtained through 10-fold cross-validation). According to the comparison results, the model architecture with the settings of multi-kernel, multi-slice channel, and capsule dropout strategy achieved the best performance (i.e., Accuracy, 82.42%; Sensitivity, 88.57%; and Specificity, 75.00%). With the benefit from the multi-kernel setting, the accuracy was elevated by 2.25% compared to that of the best column kernel setting (see the rows 11 and 13 in Table III). It was dramatically improved by 17.69% compared to that of the square kernel setting (see the rows 12 and 13 in Table III), which is frequently utilized in image and video processing when deep learning model is employed. Multi-slice channel was better than single slice channel regarding the performance. Moreover, the capsule dropout strategy performed 1.43% better than the scalar dropout (see the rows 4 and 6 in Table III) and 0.60% better than the vector dropout (see the rows 5 and 6 in Table III). In the performance comparison of loss norms, the $L2$-norm loss showed obvious superiority compared to the $L1$-norm loss.

B. *Models Comparison*

In order to demonstrate the advantage of the proposed model (i.e., multi-kernel capsule network, MKCapsnet), we compared the proposed method to not only the methods that had been used in the functional connectivity-based schizophrenia identification, but also the methods that achieved relatively good performance in other classification problems. The conventional methods, k-Nearest Neighbours (k-NN), Linear Discriminant Analysis (LDA), Linear Support Vector Machine (L-SVM), Support Vector Machine with Radial Basis Function kernel (RBF-SVM), were used in the schizophrenia identification [29]. Besides, a Deep Neural Network (DNN) with 3-hidden-layer and the settings of pre-training and L1-norm was applied to the schizophrenia identification problem [26]. In addition, Random Forest (RF) [47], [48], Gradient Boosting Machine (GBM) [49], [50], Graph Convolutional Network (GCN) [51], [52], Long Short-Term Memory (LSTM) [53], [54], and Generative Adversarial Network (GAN) [55] were included for the method comparison in this study because they achieved excellent performance in other classification problems. With the inclusion of these methods, the method comparison in this study is comprehensive and the proposed method can be thoroughly assessed. For all the above compared methods, we followed respective descriptions in the papers to establish the methods and to set the same values in the parameters exactly when available in the papers. If a parameter

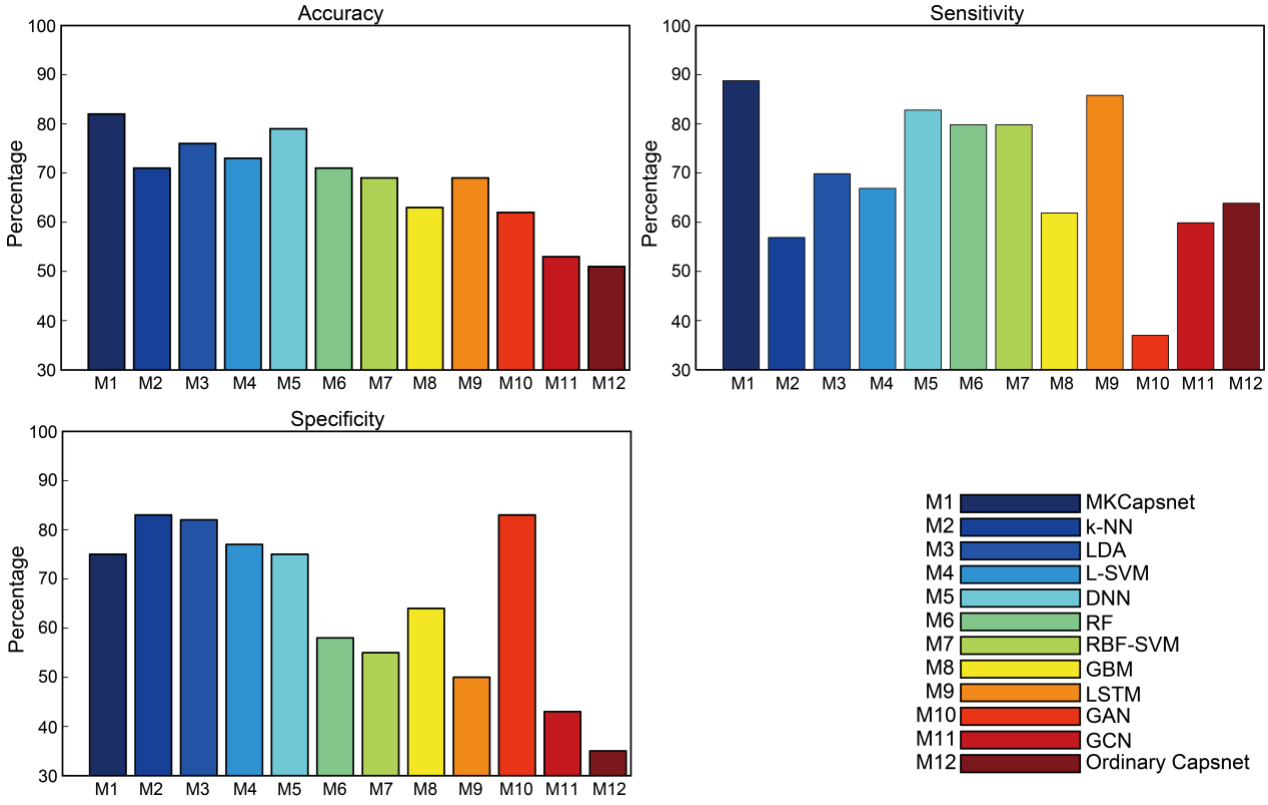

Fig. 2. Performance comparisons between the proposed model (M1) and k-NN (M2), LDA (M3), L-SVM (M4), DNN (M5), RF (M6), RBF-SVM (M7), GBM (M8), LSTM (M9), GAN (M10), GCN (M11) and Ordinary Capsnet (M12) models in schizophrenia identification. For k-NN, L-SVM, LDA, RBF-SVM and DNN methods, a feature selection procedure was utilized before the classification to boost performance as used in Li et al.'s paper [29]. The parameters used in DNN, RF, GBM, LSTM, GAN, GCN, and ordinary Capsnet models complied with the papers [26], [47], [50], [53], [55], [51], and [34] respectively

was not described in the reference papers, we utilized the grid search to find the optimal value for the parameter in order to maximise the classification performance. The details of the parameter settings for all the compared models can be found in the supplementary Table SI, Table SII, Table SIII, Table SIV, Table SV, Table SVI, Table SVII, Table SVIII, Table SIX, and Table SX. We also equally treated the data and fed the same functional connectivity matrices into the above compared methods in order to make fair comparison. MKCapsnet outperformed all the other methods in terms of classification accuracy and sensitivity (see Fig. 2). MKCapsnet achieved the highest accuracy of 82% whereas k-NN, LDA, L-SVM, DNN, RF, RBF-SVM, GBM, LSTM, GAN, GCN, and Ordinary Capsnet had the accuracies of 71%, 76%, 73%, 79%, 71%, 69%, 63%, 69%, 62%, 53%, and 51% respectively. In terms of sensitivity, MKCapsnet was at least 3% better than the others (MKCapsnet: 89% vs. k-NN: 57%, LDA: 70%, L-SVM: 67%, DNN: 83%, RF: 80%, RBF-SVM: 80%, GBM: 62%, LSTM: 86%, GAN: 37%, GCN: 60%, Ordinary Capsnet: 64%). In terms of specificity, MKCapsnet did not exhibit advantage and was even slightly worse than some compared methods (MKCapsnet: 75% vs. k-NN: 83%, LDA: 82%, L-SVM: 77%, DNN: 75%, RF: 58%, RBF-SVM: 55%, GBM: 64%, LSTM: 50%, GAN: 83%, GCN: 43%, Ordinary Capsnet: 35%).

Given that the routing is critical for the capsule network and it is valuable to look into details, we visualized the dynamic updating process of the routing coefficient c_{ij} in Fig. 3. The subplots in the first and second rows depict the evolution of c_{ij} for the samples of patients with schizophrenia while the subplots in the third and fourth rows are for the samples of healthy controls. A value of 0 was assigned to initialize all b_{ij}. At the first iteration, c_{ij} was calculated by the formula (3) and was equal to 0.5. This value of 0.5 means that there is no preference to any class. With the evolution of c_{ij}, the paired values were gradually routed to 1 and 0, representing probabilities of being each class. As shown in Fig. 3, discriminative features (a pixel stands for one feature) exhibited high probability routing to the class of schizophrenia and low probability routing to the class of healthy control for the samples of patients with schizophrenia. In contrast, the probabilities routing to the classes were opposite for the samples of healthy control. Those features which were routed to the larger probability difference were more discriminative (showing dark yellow in the first and fourth rows and dark blue in the second and third rows in Fig. 3). We can see that all areas contributed to the schizophrenia identification but the frontal area contributed more than the other brain areas.

IV. Discussion

This study proposed a multi-kernel capsule network to identify schizophrenia disease using functional connectivity features. In this model, multiple kernels were embedded to capture intrinsic connectivity characteristics of varying sizes of anatomical brain areas. The comparison results demonstrated that the MKCapsnet overall outperformed the other methods which had been used for schizophrenia identification (i.e., k-NN, L-SVM, LDA, and DNN). In particular, the performance of MKCapsnet was 6% higher than that of the second best method in terms of sensitivity. This means that the MKCapsnet is able to more accurately identify patients with schizophrenia. In practical implication, it is the lower probability in the failure of schizophrenia detection in the case of a person with schizophrenia. The proposed method does not require an individual step of feature selection as used in methods such as k-NN and L-SVM. This reduces the number of steps for the classification procedure. The drawback of separate steps of classification procedure is that the feature extraction and classifier learning cannot be simultaneously tuned, which lowers the likelihood of the best optimization so as not to reach maximum performance [56]. We proposed multiple kernels to capture functional connectivity features related to the varying sizes of anatomical brain areas. This enables the model to have the capability to learn discriminative information existing in different scales from the local community to the global community. As shown in this study, neither the smallest kernel size of 1 nor the largest kernel size of 15 provided the best performance. This might be because none of them can capture entire information existing in both the small and large scales. This issue was tackled by the utilization of multiple kernels in the proposed model, where each kernel was intended to capture the information relevant to a particular scale and the relationships between the areas at the same scale. It is worth noting that the kernel sizes were set according to the anatomical brain parcellation (AAL), rather than random selection. With considering the brain anatomical structure, capsule neural network could achieve a better performance in schizophrenia identification. Specifically, a kernel size of 1 corresponds to the smallest area (i.e., the smallest areas are region units after parcellation according to the AAL atlas, which vary in physical size) and a kernel size of 15 (since 15 smallest areas constitute the frontal area) corresponds to the frontal area (kernel sizes of 4, 6, 7, and 9 correspond to the subcortical; pariental and temporal; insula, limbic and occipital; cerebellum, respectively) in the case of that the whole brain was parcellated according to the AAL atlas [46]. A square kernel was frequently used in the image or video processing when deep learning model was utilized for classification or segmentation. This is not suitable when applying to our case because the region sequence was rearranged when assembling all connections into a matrix, which destroyed the original spatial relationship between regions. Therefore, we used the kernels including the entire column of the connectivity matrix so that all connections from one region to all other regions can be included. The rearrangement only affects the order of regions in each column. A kernel including the entire column is invariable to the inclusion of the regions. The results in our study showed that such kernel is better than the square kernel. In the future research, higher-order convolution (e.g., 3D) could be employed when connections are assembled into a higher-order tensor (e.g., third-order tensor), which may retain the original spatial relationship between brain regions.

The atlas (i.e., AAL) used for brain parcellation in this study is a validated parcellation template, which has been employed in numerous neuroscience studies. Additionally, this atlas was used to reveal significant alterations of functional connectivity in patients with schizophrenia [29]. However, the AAL suffers

from a drawback of that it was developed only based on a single subject. Currently, there are other available atlases and these atlases might bring different benefits to the schizophrenia identification. This hypothesis should be verified with additional studies. When a different atlas is applied, the proposed model can be easily adapted by adjusting kernel sizes according to the parcellation of the atlas.

Although our model was proposed for schizophrenia identification in this study, it is applicable to the identification of other brain diseases or the classification other than disease diagnosis after adaptions based on applications. The extent of adaption depends on how an application differs from the case in this study. Less work in the adaption is required if an application is similar to our case. In this study, we included the models that have been used for schizophrenia identification and several other relevant models for performance comparisons, demonstrating deep learning models should be adapted to comply with underlying mechanisms associated with diseases in order to improve performance. We believe that other deep learning models with significant adaption based on neural mechanisms could also enhance performance.

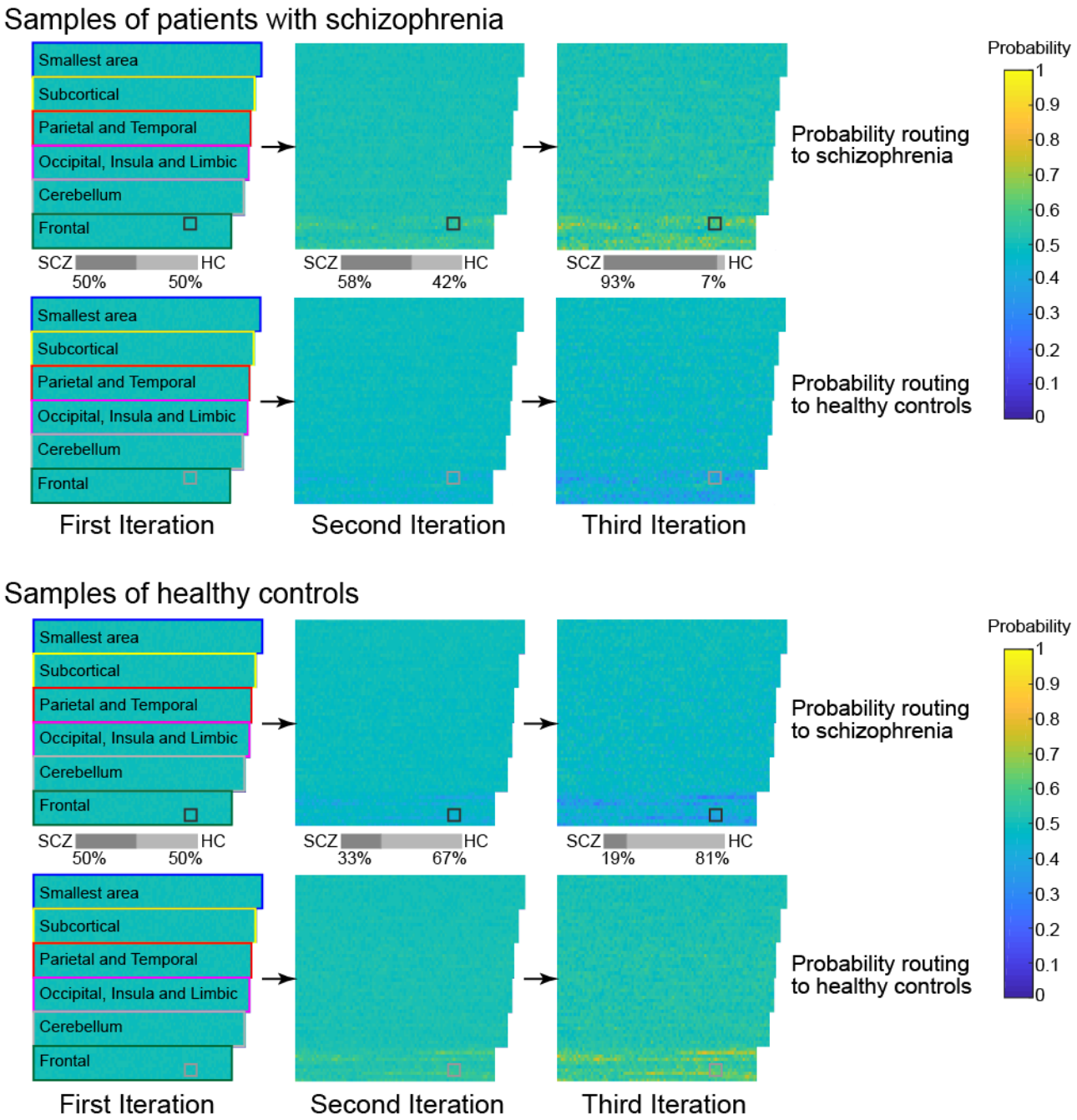

Fig. 3. The visualization of the routing process. The subplots in the first and second rows depict the evolution of c_{ij} for the samples of patients with schizophrenia while the subplots in the third and fourth rows are for the samples of healthy control. The probabilities evolved with the interations are illustrated for the example elements. It can be seen that the probability of the ground-truth class is increased with the interations.

Considering the optimal parameter settings are crucial to the success of the model, grid search was utilized. The grid search is a frequently-adopted strategy in the parameter setting of a neural network model, which really contributes to model performance improvement. This strategy of grid search was also implemented in all the compared methods in order to make fair comparison. The main parameters of compared methods are set as illustrated in the original reference papers. As there are massive parameters we have to set, it would take an unacceptable amount of time if all parameters are optimized together by the grid search. Therefore, we grouped the parameters and set the parameters with the reference of the settings in other deep learning models and the original capsule neural network. Because the sequence of setting parameters also matters and affects the model performance, we optimized the parameters in the order of their importance. The number of routing iterations was determined in the first place due to the critical use of routing process in the capsule network. The number of routing iterations was finally set as 3 based on the model performance (searching among 1, 2, 3, 4, and 5), which is the same to the setting described in the Sabour et al.'s original paper [34]. The crucial training parameters, activation function (including softplus [57], rectified linear units [58], and sigmoid [59]) and learning rate (among 0.1, 0.05, 0.01, 10^{-3}, and 10^{-4}), which affect the backpropagation process and the convergence of training phase, were then determined individually after the routing iterations was settled down. The epoch number (among 50, 100, 500, and 1000) and batchsize (among 2, 3, 8, 16, and 32), which affect the time required for training sessions and are relevant to the overfitting of training phase, were considered simultaneously. Then, the dropout rate (among 0.3, 0.4, 0.5, 0.6, and 0.7) that prevent the model overfitting was decided. Other insensitive parameters, whose minor variation does not lead to significant change in the model performance, are searched at last. The filter/slice size (among 8, 16, 32, 64, 128, and 256) and stride (among 1, 2, 3, and 4) that affect the feature size, and channel number (among 6, 12, 32, and 64) and vector length (among 8, 12, 20, 32, and 64) that affect the complexity of capsules were considered at the last step of the grid search procedure.

Based on the comparison results of different norms used in the loss function, the performance is differential. In our case, L2-norm loss was better than L1-norm loss. All three assessment indicators showed superior accuracy when the L2-norm loss was utilized (L2-norm vs. L1-norm, accuracy, 82.42% vs. 69.34%; sensitivity, 88.57% vs. 75.89%; specificity, 75.00% vs. 61.67%). L1-norm loss exhibited fluctuant during searching an optimal solution and was less convergent. In contrast, L2-norm loss showed relatively stable convergence. In the capsule layer, we brought capsule dropout strategy to improve training effectiveness. Other than the scalar dropout strategy used in the image or video processing, we randomly discarded vectors. Moreover, we separately set the dropout rate for each channel (corresponding to each kernel) so that the number of vectors discarded in each channel can be kept identical. Compared to the vector dropout strategy (the dropout was not separately set for each channel and the number of vectors discarded in one channel might be more than that of another channel.), the classification performance was improved by 0.6 % when the capsule dropout strategy was used. The improvement was 1.4 % when compared to the scalar dropout scalar strategy which has been widely utilized in the deep learning models when processing the image or video data. These results demonstrated that our strategy discarding entire vectors is better than that of discarding elements of a vector in the capsule network. Moreover, the separate dropout in each channel gives the advantage that the dropout rate would not imbalanced across channels. Therefore, it avoids that there is excessive dropout in some channels whereas there is a lack in the others.

V. Conclusion

In this study, we proposed a multi-kernel capsule network to identify schizophrenia. The proposed method was compared not only to the methods that had been used in the functional connectivity-based schizophrenia identification, but also to the methods that achieved excellent performance in other classification problems. According to the comparison results, our proposed method is the best one among all these methods. In order to provide insights into the model architecture of the proposed method, we compared different architecture settings and demonstrated the outstanding performance of the proposed method. Moreover, we gave the parameter tuning suggestions in this paper based on our empirical experience, which might be helpful to the researchers who engage in the studies using a capsule network. Our study is the first attempt to identify schizophrenia based on functional connectivity by a capsule network and could give heuristic cues for further studies.

Due to that it is at a very early stage to develop a capsule network for disease detection, there is a large space to improve performance from many angles. For instance, vector representation in the capsule network can be replaced by tensor representation. In this case, additional information can be represented besides the direction and probability that have been represented by a vector.

In summary, our study demonstrated that capsule network was feasible and promising in the identification of schizophrenia. This model can also be extended to detect other diseases after appropriate adaption. Further efforts are required to improve the performance and broaden applications of the capsule networks.

References

[1] T. Walsh *et al.*, "Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia," *Science*, vol. 320, no. 5875, pp. 539–543, 2008.

[2] S. Saha, D. Chant, J. Welham, and J. McGrath, "A systematic review of the prevalence of schizophrenia," *PLoS Med.*, 2(5), e141, 2005.

[3] O. Marín, "Interneuron dysfunction in psychiatric disorders," *Nat. Rev. Neurosci.*, vol. 13, no. 2, p. 107, 2012.

[4] A. Ulloa, S. Plis, and V. Calhoun, "Improving classification rate of schizophrenia using a multimodal multi-layer perceptron model with structural and functional MR," *arXiv Prepr. arXiv1804.04591*, Apr. 2018.

[5] G. Deshpande, P. Wang, D. Rangaprakash, and B. Wilamowski, "Fully connected cascade artificial neural network architecture for attention deficit hyperactivity disorder classification from functional magnetic

resonance imaging data," *IEEE Transactions on Cybernetics*, vol. 45, no. 12, pp. 2668 – 2679, 2015.

[6] P. Yang *et al.*, "Fused sparse network learning for longitudinal analysis of mild cognitive impairment," *IEEE Transactions on Cybernetics.* pp. 1 – 14, 2019.

[7] J. Wang, Q. Wang, H. Zhang, J. Chen, S. Wang, and Di. Shen, "Sparse multiview task-centralized ensemble learning for ASD diagnosis based on age- and sex-related functional connectivity patterns," *IEEE Transactions on Cybernetics*, vol. 49, no. 8. pp. 3141 – 3154, 2019.

[8] A. Savio and M. Graña, "Local activity features for computer aided diagnosis of schizophrenia on resting-state fMRI," *Neurocomputing*, vol. 164. pp. 154 – 161, 2015.

[9] M. Hu, K. Sim, J. H. Zhou, X. Jiang and C. Guan, "Brain MRI-based 3D Convolutional Neural Networks for Classification of Schizophrenia and Controls,"*arXiv Prepr. arXiv2003.08818*, 2020.

[10] M. Liu, J. Zhang, C. Lian, and D. Shen, "Weakly supervised deep learning for brain disease prognosis using MRI and incomplete clinical scores," *IEEE Transactions on Cybernetics.* pp. 1 – 12, 2019.

[11] S. R. Sponheim, W. G. Iacono, P. D. Thuras, and M. Beiser, "Using biological indices to classify schizophrenia and other psychotic patients," *Schizophr. Res.*, vol. 50, no. 3, pp. 139–150, 2001.

[12] S. R. Sponheim, W. G. Iacono, P. D. Thuras, S. M. Nugent, and M. Beiser, "Sensitivity and specificity of select biological indices in characterizing psychotic patients and their relatives," *Schizophr. Res.*, vol. 63, no. 1–2, pp. 27–38, 2003.

[13] Y. Takayanagi *et al.*, "Differentiation of first-episode schizophrenia patients from healthy controls using ROI-based multiple structural brain variables," *Prog. Neuro-Psychopharmacology Biol. Psychiatry*, vol. 34, no. 1, pp. 10–17, 2010.

[14] Y. Takayanagi *et al.*, "Classification of first-episode schizophrenia patients and healthy subjects by automated MRI measures of regional brain volume and cortical thickness," *PLoS One*, vol. 6, no. 6, p. e21047, 2011.

[15] E. Castro, M. Martínez-ramón, G. Pearlson, J. Sui, and V. D. Calhoun, "Characterization of groups using composite kernels and multi-source fMRI analysis data : Application to schizophrenia," *Neuroimage*, vol. 58, no. 2, pp. 526–536, 2011.

[16] Y. Fan *et al.*, "Multivariate examination of brain abnormality using both structural and functional MRI," *Neuroimage*, vol. 36, no. 4, pp. 1189–1199, 2007.

[17] Y. Fan, D. Shen, R. C. Gur, R. E. Gur, and C. Davatzikos, "COMPARE: classification of morphological patterns using adaptive regional elements," *IEEE Trans. Med. Imaging*, vol. 26, no. 1, pp. 93 – 105, 2007.

[18] Y. Fan *et al.*, "Unaffected family members and schizophrenia patients share brain structure patterns: a high-dimensional pattern classification study," *Biol Psychiatry*, vol. 63, no. 1, pp. 118 – 124, 2008.

[19] M. Nieuwenhuis, N. E. M. Van Haren, H. E. H. Pol, W. Cahn, and H. G. Schnack, "Classification of schizophrenia patients and healthy controls from structural MRI scans in two large independent samples," *Schizophr. Res.*, vol. 61, no. 3, pp. 606 – 612, 2012.

[20] W. Pettersson-Yeo, P. Allen, S. Benetti, P. McGuire, and A. Mechelli, "Dysconnectivity in schizophrenia: where are we now?," *Neurosci. Biobehav. Rev.*, vol. 35, no. 5, pp. 1110–1124, 2011.

[21] E. Zarogianni, T. W. J. Moorhead, and S. M. Lawrie, "Towards the identification of imaging biomarkers in schizophrenia, using multivariate pattern classification at a single-subject level," *NeuroImage Clin.*, vol. 3, pp. 279–289, 2013.

[22] B. F. *et al.*, "Specific Substantial Dysconnectivity in Schizophrenia: A Transdiagnostic Multimodal Meta-analysis of Resting-State Functional and Structural Magnetic Resonance Imaging Studies," *Biological Psychiatry*, vol. 85, no. 7. pp. 573 – 583, 2019.

[23] J. Fitzsimmons, M. Kubicki, and M. E. Shenton, "Review of functional and anatomical brain connectivity findings in schizophrenia," *Curr. Opin. Psychiatry*, vol. 26, no. 2, pp. 172–187, 2013.

[24] K. Friston, H. R. Brown, J. Siemerkus, and K. E. Stephan, "The dysconnection hypothesis (2016)," *Schizophr. Res.*, vol. 176, no. 2–3, pp. 83–94, 2016.

[25] M. R. Arbabshirani, K. Kiehl, G. Pearlson, and V. D. Calhoun, "Classification of schizophrenia patients based on resting-state functional network connectivity," *Front. Neurosci.*, vol. 7, p. 133, 2013.

[26] J. Kim, V. D. Calhoun, E. Shim, and J.-H. Lee, "Deep neural network with weight sparsity control and pre-training extracts hierarchical features and enhances classification performance: Evidence from whole-brain resting-state functional connectivity patterns of schizophrenia," *Neuroimage*, vol. 124, pp. 127–146, 2016.

[27] H. Shen, L. Wang, Y. Liu, and D. Hu, "Discriminative analysis of resting-state functional connectivity patterns of schizophrenia using low dimensional embedding of fMRI," *Neuroimage*, vol. 49, no. 4, pp. 3110–3121, 2010.

[28] B. Rashid *et al.*, "Classification of schizophrenia and bipolar patients using static and dynamic resting-state fMRI brain connectivity," *Neuroimage*, vol. 134, pp. 645–657, 2016.

[29] J. Li, Y. Sun, Y. Huang, A. Bezerianos, and R. Yu, "Machine learning technique reveals intrinsic characteristics of schizophrenia : an alternative method," *Brain Imaging Behav. 1-11.*, pp. 1–11, 2018.

[30] M. R. Arbabshirani, K. A. Kiehl, G. D. Pearlson, and V. D. Calhoun, "Classification of schizophrenia patients based on resting-state functional network connectivity," *Frontiers in Neuroscience*, no. 7 JUL. 2013.

[31] H. Shen, L. Wang, Y. Liu, and D. Hu, "Discriminative analysis of resting-state functional connectivity patterns of schizophrenia using low dimensional embedding of fMRI," *Neuroimage*, vol. 49, no. 4, pp. 3110 – 3121, 2010.

[32] X. Ji, W. Quan, L. Yang, J. Chen, J. Wang and T. Wu, "Classification of Schizophrenia by Seed-based Functional Connectivity using Prefronto-Temporal Functional Near Infrared Spectroscopy," *Journal of Neurocience Methods*, 2020, 344.

[33] C. R. Phang, C. M. Ting, F. Noman, H. Ombao, "Classification of EEG-Based Brain Connectivity Networks in Schizophrenia Using a Multi-Domain Connectome Convolutional Neural Network," *arXiv Prepr. arXiv1903.08858, 2019.*

[34] S. Sabour, N. Frosst, and G. E. Hinton, "Dynamic routing between capsules," *The 31st Conference on Neural Information Process Systems (NIPS 2017)*, Long Beach, USA, 2017.

[35] A. Mobiny and H. Van Nguyen, "Fast CapsNet for lung cancer screening," *medical image computing and computer assisted intervention*, vol. 11071 LNCS. pp. 741 – 749, 2018.

[36] T. Iesmantas and R. Alzbutas, "Convolutional capsule network for classification of breast cancer histology images," *International Conference Image Analysis and Recognition*, 2018, pp. 853–860.

[37] C. Xia, C. Zhang, X. Yan, Y. Chang, and P. S. Yu, "Zero-shot user intent detection via capsule neural networks," *arXiv Prepr. arXiv1809.00385*, 2018.

[38] W. Zhao, J. Ye, M. Yang, Z. Lei, S. Zhang, and Z. Zhao, "Investigating capsule networks with dynamic routing for text classification," *arXiv Prepr. arXiv1804.00538*, 2018.

[39] K. A. Islam, D. Pérez, V. Hill, B. Schaeffer, R. Zimmerman, and J. Li, "Seagrass detection in coastal water through deep capsule networks," *Chinese Conference on Pattern Recognition and Computer Vision (PRCV)*, 2018, pp. 320–331.

[40] D. R. de Jesus, J. Cuevas, W. Rivera, and S. Crivelli, "Capsule Networks for Protein Structure Classification and Prediction," *arXiv Prepr. arXiv1808.07475*, Aug. 2018.

[41] C. Xiang, L. Zhang, Y. Tang, W. Zou, and C. Xu, "MS-CapsNet: A Novel Multi-Scale Capsule Network," *IEEE Signal Process. Lett.*, vol. 25, no. 12, pp. 1850–1854, 2018.

[42] P. Afshar, A. Mohammadi, and K. N. Plataniotis, "Brain tumor type classification via capsule networks," *IEEE International Conference on Image Processing (ICIP)*, 2018, pp. 3129–3133.
[43] A. R. Mayer *et al.*, "Functional imaging of the hemodynamic sensory gating response in schizophrenia," *Hum. Brain Mapp.*, vol. 34, no. 9, pp. 2302–2312, 2013.
[44] X.-W. Song *et al.*, "REST: A Toolkit for Resting-State Functional Magnetic Resonance Imaging Data Processing," *PLoS One*, vol. 6, no. 9, p. e25031, Sep. 2011.
[45] Y. Chao-Gan and Z. Yu-Feng, "DPARSF：A MATLAB Toolbox for" Pipeline" Data Analysis of Resting-State fMRI," *Frontiers in Systems Neuroscience*, 4, 13.
[46] N. Tzourio-Mazoyer *et al.*, "Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain," *Neuroimage*, vol. 15, no. 1, pp. 273–289, 2002.
[47] T. R. Sivapriya, K. A. R. N. Banu, and T. P. R. Jeba, "Ensemble Merit Merge Feature Selection for Enhanced Multinomial Classification in Alzheimer's Dementia," *Comput. Math. Methods Med.*, vol. 2015, pp. 1–11, 2015.
[48] K. Oppedal, T. Eftestøl, K. Engan, M. K. Beyer, and D. Aarsland, "Classifying dementia using local binary patterns from different regions in magnetic resonance images," *International Journal of Biomedical Imaging*, vol. 2015. 2015.
[49] G. Lemaitre *et al.*, "A boosting approach for prostate cancer detection using multi-parametric MRI," *Twelfth International Conference on Quality Control by Artificial Vision 2015*. International Society for Optics and Photonics, 2015.
[50] G. J. Katuwal, N. D. Cahill, S. A. Baum, and A. M. Michael, "The predictive power of structural MRI in Autism diagnosis," *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, vol. 2015-Novem. pp. 4270–4273, 2015.
[51] X. Zhao, F. Zhou, L. Ou-Yang, T. Wang, and B. Lei, "Graph convolutional network analysis for mild cognitive impairment prediction," *Proceedings - International Symposium on Biomedical Imaging*, vol. 2019-April. pp. 1598–1601, 2019.
[52] Z. Wang, Y. Tong, and X. Heng, "Phase-Locking Value Based Graph Convolutional Neural Networks for Emotion Recognition," *IEEE Access*, vol. 7. pp. 93711–93722, 2019.
[53] M. Liu and Y. Kim, "Classification of Heart Diseases Based on ECG Signals Using Long Short-Term Memory," *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, vol. 2018-July. pp. 2707–2710, 2018.
[54] P. Qiao, Y. Zhang, D. Chen, and G. Xu, "Character-Based Convolutional Grid Neural Network for Breast Cancer Classification," *2017 International Conference on Green Informatics (ICGI)*, 2017.
[55] J. Zhao *et al.*, "Functional network connectivity (FNC)-based generative adversarial network (GAN) and its applications in classification of mental disorders," *Journal of Neuroscience Methods*, vol. 341. 2020.
[56] J. Li, C. Li, A. Cichocki, "Canonical Polyadic Decomposition with Auxiliary Information for Brain Computer Interface," *IEEE Journal of Biomedical & Health Informatics*, 2017, 21(1): 263-271.
[57] Zheng H, Yang Z, Liu W, Liang J and Li Y, "Improving deep neural networks using softplus units," *International Joint Conference on Neural Network*, 2015: 1-4.
[58] J. Schmidt-Hieber, "Nonparametric regression using deep neural networks with ReLU activation function," *arXiv Prepr. arXiv*1708.06633, 2017.
[59] G. Cybenko, "Approximation by superpositions of a sigmoidal function," *Mathematics of Control, Signals, and Systems*, vol. 2, no. 4. pp. 303–314, 1989.

Tian Wang received his M.S. degree in electrical and computer engineering from the National University of Singapore, Singapore, in 2016. He was a Research Assistant at the National University of Singapore, Singapore, and was a Research Assistant at the Wuyi University, Jiangmen, China, before moving to industry. His research interests include machine learning, data mining and neuropyhsiological data analytics.

Anastasios (Tassos) Bezerianos, BS, MS, PhD
He is Senior Researcher at the Hellenic Institute of Transport of Centre for Research and Technology Hellas (CERTH), Thessaloniki, Greece. He has been the Professor of Medical Physics at the Medical School of Patras University, Patras Greece since 2004, head of Cognitive Engineering (COGEN) Laboratory at the N.1 Health Institute, National University of Singapore, from 2012 to 2020 and a Visiting Professor in the Computer Science Department of New South Wales University (NSWU) -Canberra, AUS from 2016 to 2020. He studied Physics at Patras University and Telecommunications at Athens University, and he received his Ph.D. on Bioengineering from the University of Patras. His research entails diverse areas spanning from Artificial Intelligence and Robotics to Biomedical Signal Processing and Brain Imaging as well as Mathematical Biology and Systems Medicine and Bioinformatics. His work is summarized in 164 journal and 245 conference proceedings publications, 1 book and 2 patents. He has research collaborations with research institutes and universities in Japan, China, and Europe, and he is Associate Editor of IEEE TNSRE and PLOS ONE Neuroscience journals and a reviewer for several international scientific journals.. He is a Fellow of The European Alliance for Medical and biological Engineering &Science (EAMBES) and Senior Member of IEEE and Founder and Chairman of the biannual International Summer School on Emerging Technologies in Biomedicine and the International Workshop on Human intELligence machIne cOexiStence (HELIOS)

Andrzej Cichocki received the M.Sc. (with honors), Ph.D. and Dr.Sc. (Habilitation) degrees, all in electrical engineering from Warsaw University of Technology (Poland). He spent several years at University Erlangen (Germany) as an Alexander-von-Humboldt Research Fellow and Guest Professor. He was a Senior Team Leader and Head of the laboratory for Advanced Brain Signal Processing, at RIKEN Brain Science Institute (Japan) and now he is a Professor in the Skolkovo Institute of Science and Technology - SKOLTECH (Russia). He is author of more than 500 technical journal papers and 6 monographs in English (two of them translated to Chinese). He served as Associate Editor of, IEEE Trans. on Signals Processing, IEEE Trans. on Neural Networks and Learning Systems, IEEE Trans on Cybernetics, Journal of Neuroscience Methods and he is founding Editor in Chief for Journal Computational Intelligence and Neuroscience. Currently, his research focus on multiway blind source separation, tensor decompositions, tensor networks, deep learning, human robot interactions and brain computer interface. His

publications currently report over 46,000 citations according to Google Scholar, with an h-index of 97. He is Fellow of the IEEE since 2013.

Junhua Li (S'12-M'13-SM'18) received his PhD from the Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai, China, in 2013.

He is currently a Lecturer in the School of Computer Science and Electronic Engineering, University of Essex, Colchester, UK. He was a Senior Research Fellow at the National University of Singapore, Singapore, before moving to the UK. His research interests include computational neuroscience, brain-computer interface, machine learning, neurophysiological data analytics, and their practical applications.

He is an Associate Editor for IEEE Access, a Review Editor for Frontiers in Human Neuroscience, and served as a Guest Associate Editor for several journal special issues related to his research interests.