2018 IEEE International Conference on Data Mining Workshops (ICDMW)

Voxelwise 3D Convolutional and Recurrent Neural Networks for Epilepsy and Depression Diagnostics from Structural and Functional MRI Data

Marina Pominova
Skolkovo Institute of Science and Technology
Moscow, Russia
marina.pominova@skoltech.ru

Alexey Artemov
Skolkovo Institute of Science and Technology
Moscow, Russia
a.artemov@skoltech.ru

Maksim Sharaev
Skolkovo Institute of Science and Technology
Moscow, Russia
m.sharaev@skoltech.ru

Ekaterina Kondrateva
Skolkovo Institute of Science and Technology
Moscow, Russia
ekaterina.kondrateva@skoltech.ru

Alexander Bernstein
Skolkovo Institute of Science and Technology
Moscow, Russia
a.bernstein@skoltech.ru

Evgeny Burnaev
Skolkovo Institute of Science and Technology
Moscow, Russia
e.burnaev@skoltech.ru

***Abstract*—In the field of psychoneurology, analysis of neuroimaging data aimed at extracting distinctive patterns of pathologies, such as epilepsy and depression, is well known to represent a challenging problem. As the resolution and acquisition rates of modern medical scanners rise, the need to automatically capture complex spatiotemporal patterns in large imaging arrays suggests using automated approaches to pattern recognition in volumetric images, such as training a classification models using deep learning. On the other hand, with typically scarce training data, the choice of a particular neural network architecture remains an unresolved issue. In this work, we evaluate off-the-shelf building blocks of deep voxelwise neural architectures with the goal of learning robust decision rules in computational psychiatry. To this end, we carry out a series of computational experiments, aiming at the recognition of epilepsy and depression on structural (3D) and functional (4D) MRI data. We discover that our investigated models perform on par with computational approaches known in literature, without the need for sophisticated preprocessing and feature extraction.**

***Index Terms*—structural MRI, functional MRI, epilepsy diagnostics, depression diagnostics, deep learning, convolutional neural networks, recurrent neural networks**

I. Introduction

Neurological brain diseases cause a significant decline in the quality of life, disability of millions of people around the world, and in addition, represent a global economic burden for society. Epilepsy is one of the most widespread neurological diseases, which affects 0.6–1.5% of the world's population [1], so accurate recognition and localization of epilepsy foci represent clear practical interest for clinical medicine.

It is known that the most frequent psychiatric comorbidities that complicate epilepsy are mood disorders, among which depression is prevalent. Approximately 1 to 3% of men and 2 to 9% of women in the general population suffer from depression disorder, however, it is significantly more prevalent in patients with epilepsy [2], [3]. The high comorbidity of epilepsy and depression may indicate the existence of shared pathological mechanisms [4]. However, such comorbidities often remain unrecognized and untreated as there are reciprocal interactions between the clinical manifestations of these disorders, and the presence of one of the diseases can complicate the diagnostics of another. Thus, distinguishing between epilepsy and comorbid epilepsy + depression relies on the identification of stable features (biological markers). Desired biomarkers of epilepsy would help correctly diagnose it for both patients with and without depression, and similarly, biomarkers of depression are required to not be dependent on the presence of epilepsy.

Nowadays, the information on pathological processes related to the structure and functional activity of the brain can be obtained using various scanning techniques, such as X-ray Computed Tomography (CT), High-Resolution Electroencephalography (HR-EEG), and structural and functional Magnetic Resonance Imaging (MRI, fMRI). In the present research, we focus on MRI-based modalities as existing research gives evidence that biomarkers of epilepsy can be present in structural neuroimaging data [5]–[8]. In addition, MRI scanning is included in contemporary epilepsy recommendations since high-resolution T1-weighted MRI has been demonstrated to reveal structural brain impairments associated with temporal lobe epilepsy (TLE) [9]. Similarly, depression is believed to strongly affect the functional activity of the brain. Thus, it can be detected by analyzing the data that contain information on aberrations in the neuronal network functioning [10]–[12].

Modern machine learning methods are successfully applied in medical studies, including the analysis of neuroimaging data [13]–[15]. However, challenges associated with data analysis in neuroimaging is that neuroimaging datasets available

2375-9259/18/$31.00 ©2018 IEEE
DOI 10.1109/ICDMW.2018.00050

for research typically contain a relatively small number of observations, while brain scanning techniques provide data of very high dimensionality. Application of common machine learning algorithms in such settings may lead to overfitting to the training set without finding actual dependence between data and target variable. Therefore, most of the existing approaches are forced to reside to preliminary extraction of the features which describe the dominant patterns in data, but are lower dimensional than original brain images. While this reduction in dimensionality can be performed with conventional methods, e.g. weighted PCA [16], or by computation of clinically meaningful features using specialized software [17]–[19], it still results in loss of discriminative power for predictive methods.

An alternative approach to the extraction of discriminative features from structural and functional MRI data is based on deep learning methods that have become widespread in the last few years and shown outstanding performance on a wide range of problems, particularily in image recognition [20]. The advantage of deep learning methods lies in their ability to automatically derive complex and informative problem-specific features from the raw data during the training process. It allows training a neural network directly on high-dimensional brain imaging data skipping the feature extraction step.

By design, neural architectures for deep learning are built in a modular way, with basic building blocks, such as composite convolutional layers, typically reused across many models and applications. This enables the standardization of deep learning architectures, with much research devoted to the exploration of pre-built layers and pre-trained activations (cf. [21], [22] for transfer learning, image retrieval, etc.). However, research aimed at the recognition of brain pathologies via application of deep neural networks to neuroimaing data is still in its infancy. Thus, the choice of appropriate building blocks targeting specific clinical applications such as epilepsy or depression recognition remains an open problem and requires further investigation.

In the present study, we carry out an extensive experimental evaluation of off-the-shelf layers for deep voxelwise neural architectures with an ultimate goal of designing effective decision rules in computational psychiatry. For this purpose, we consider the problems of recognition of epilepsy and depression based on the whole brain structural and functional MRI data. In this context, we pose four binary classification problems with clear clinical relevance. To identify the promising building blocks, construct our models, and setup the classification tasks, we survey a body of existing research and design a series of deep voxelwise neural architectures tailored for 3D and 4D data. We also attempt to select the most significant features for depression and epilepsy recognition via visualizing the attention of our trained neural networks.

This paper is structured as follows. In Section II, we survey the existing literature on the analysis of neuroimaging with a focus on deep learning-based approaches. Section III gives details on the sample of neuroimaging data we experiment with, specifies neural network architectures investigated, and lists the conducted experiments. In Section IV, we provide the obtained resluts of our evaluation, including the visualizations of neural network attention. We conclude with a discussion of results in Section V.

II. Related work

A. Neuroimaging-based diagnostics in psychoneurology

Brain changes associated with temporal lobe epilepsy (TLE) include asymmetrical distribution of temporal lobe abnormalities on the same brain hemisphere in the hippocampus, parahippocampal gyrus, and entorhinal cortex [23]. Therefore many forthcoming studies attempted to find biomarkers of MTLE (Medial Temporal Lobe Epilepsy) in these particular regions. For instance, [24] found reductions in both functional and structural connectivity between hippocampal structures and adjacent brain regions, as well as connectivity among the default mode network (DMN). It is believed that degeneration of structural connectivity may help explain the pathophysiological mechanism of the impaired cognitive functions. In voxel-level fMRI analysis, the basic experimental methodology involves defining regions of interest (ROIs) to be used as masks, applying them onto the residual images to extract the mean signal time-courses from each predefined ROI, and computing correlation coefficients between pairs of signal time-courses. The latter are then utilized to test hypotheses [25].

Functional MRI-based research into TLE is typically performed with presented visual stimuli (words, faces or scenes), and the goal of the patient is to respond, for instance, by pushing the button [26], [27]. Assessing the functional reserve of key memory structures, which remains a challenge for presurgical patients with intractable temporal lobe epilepsy, is central for a number of studies [27]. They provide evidence for detection of predictable patterns of activity across voxels associated with specific memories in MTL structures, including the hippocampus. Overall, their findings indicate that MVPA-fMRI could prove a useful non-invasive method of assessing pre-surgical memory capacity within the MTL. Patients with long-standing epilepsy may have variable anatomic localization of neurologic functions, such as memory, because of cerebral reorganization induced by the disease process [26]. Understanding this functional anatomy is vital when planning surgical resections and relies on complex preoperative evaluations. fMRI is a valid tool for assessing of memory lateralization in patients with MTLE and may therefore allow noninvasive preoperative evaluation of memory lateralization. Another issue is the development of neuroimaging measures that prove to be strongly predictive neuroimaging markers in pattern classification between healthy controls and general epileptic patients [28]. Using modern pattern-recognition approaches like sparse regression and support vector machine, they have achieved a cross-validated classification accuracy of 83.9% (specificity: 82.5%; sensitivity: 85%) across individuals from a large dataset consisting of 180 healthy controls and epileptic patients.

B. Deep learning-based pathology recognition

There are several examples of successful application of convolutional neural networks (CNN) with various architectures to the segmentation of volumetric MRI data [29], [30]. Many of those solutions are based on adaptation existing approaches to 2D-image analysis for handling volumetric data. For instance, [31] proposed for brain segmentation a ResNet-like architecture which extends the deep residual learning for processing of volumetric MRI data using 3D-filters in convolutional layers. Their model called VoxResNet was composed of volumetric residual blocks (VoxRes blocks) containing a total of 25 convolutional layers and also 4 deconvolutional layers. They demonstrated the potential of ResNet-like volumetric architectures, achieving better results compared to many of state-of-the-art methods for MRI images segmentation.

Convolutional neural networks have also shown good classification results in problems related to other psychoneurological diseases such as Alzheimer [32]. In a recent paper [33], two different 3D convolutional network architectures were proposed for a problem of distinguishing patients with Alzheimers disease from normal cohort based on their MRI scans. Firstly, they constructed a model with VGG-like architecture called VoxCNN. It is composed of four blocks with volumetric convolutional layers followed by two fully connected layers with batchnorm and dropout. Secondly, they derived classification model from VoxResNet architecture, previously proposed for segmentation. Both networks showed similar performance of 0.88 and 0.87 ROC AUC respectively for classification subjects with Alzheimers disease versus mild cognitive impairment and normal controls. These results are comparable to earlier approaches that use preliminary extracted features. Thus, it indicates that convolutional networks can be applied directly to the raw neuroimaging data without loss of classification performance, which allows skipping the preprocessing stage. However, to the best of our knowledge, there are still no studies devoted to the application of convolutional networks for the recognition of epilepsy based on MRI imaging.

When applied to medical tasks, recurrent neural networks (RNN) can be effectively used to analyze sequential data, such as electroencephalogram (EEG) or fMRI scanning results. Dvornek, Ventola, Pelphrey, and Duncan utilized recurrent networks with long short term memory (LSTM) units to find biomarkers for autism spectrum disorders (ASD) from resting-state functional MRI data taking into account its sequential nature [34]. They demonstrated classification accuracy of 0.69 in the problem of distinguishing individuals with ASD from healthy control. Moreover, they performed the comparison with previous approaches on the largest dataset available for ASD, which showed that RNN exceeded existing results by 9%. However, their network takes as input time series of ROI activations, which should be preliminarily computed from the raw fMRI data.

Dakka et al. proposed the model combining the advantages of recurrent and convolutional networks to classify fMRI series directly in the form of volumetric images sequence for the diagnosis of schizophrenia [35]. They examined multiple configurations consisting of several convolutional blocks with 3D-filters and two LSTM layers. The most successful of their models resulted in 0.65 and 0.63 classification accuracy and showed significant improvement of performance compared with baseline prediction methods such as support vector machines (SVM), and competitive results among other deep learning method applied to that problem.

Earlier, a similar approach was effectively applied for analysis of EEG data by Bashivan, Rish, Yeasin, and Codella [36]. In contradistinction to fMRI, EEG time-series consist of two-dimensional images, therefore the authors used standard 2D-convolutions to extract spatial features. First, they evaluated several CNN architectures classifying distinct images sampled from EEG data. Then the best performed one was extended to handle the whole time-series. The authors compared various approaches to deal with temporal component of the data including max-pooling over time, 1D temporal convolution and recurrent part with one or several LSTM layers. The latter achieved a test error of 8.89% on the four-class classification demonstrating the best accuracy among the other variants considered and significant performance improvements over the state-of-the-art results.

III. Materials and Methods

In this section, we present the dataset that we used in our experiments and describe the models examined for depression and epilepsy detection problems.

A. Data used in the research

The brain imaging data used in this research include resting state functional T2-weighted MRI EPI series of 100 patients and also structural T1-weighted MP-RAGE images for 90 of them. The original resolution of the structural MRI images is $512 \times 512 \times 340$ voxels. Functional MRI series consists of 133 images recorded with an interval of 3.7 seconds.

Details on data preprocessing. Structural MRI data was preprocessed with the use of SPM toolbox[1]. During the preprocessing, the size of each image was reduced from the original $512 \times 512 \times 340$ voxels to $180 \times 216 \times 180$ voxels.

Functional data was preprocessed in SPM toolbox and denoised in two different ways: automatic elimination of high and low frequencies (spectral denoise) and exclusion of the noisy independent components obtained using ICA (for applications of independent component analysis to denoising fMRI data see [37], [38]) from FSL Melodic (manual denoise, see [39], [40]) toolbox[2]. Dimensionality of the resulting sequences was reduced from $64 \times 64 \times 30$ to $52 \times 62 \times 52$ voxels for an image at each time point. Moreover, fMRI scans of two patients were removed from the dataset due to the low quality of the captured scans as evaluated by the manual investigation.

Each subject in our dataset belongs to one of the 4 groups:

- the healthy control group (H),

[1]https://www.fil.ion.ucl.ac.uk/spm/ext/

[2]https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/MELODIC

TABLE I: Distribution of neuroimaging data in our dataset across subject groups and imaging modalities

Group	# Structural scans	# Functional scans
H	23	25
D	24	25
E	21	25
DE	22	23

- patients with major depressive disorder (D) experiencing an acute depressive episode,
- patients diagnosed with epilepsy (E), and
- patients diagnosed with both epilepsy and major depressive disorder (DE).

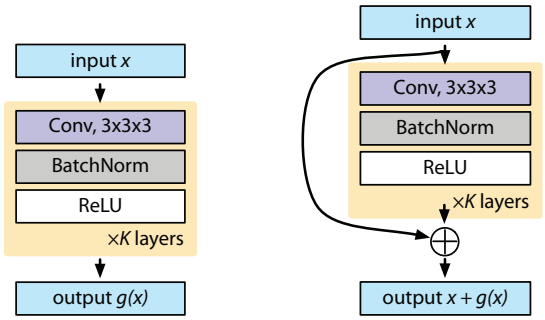
Each group included adult subjects (under 70 years old) of both genders. The distribution of patients across the groups is presented in the Table I for structural and functional images.

B. Specifications of the investigated models

In the present work, we investigate three off-the-shelf building blocks for our models, as identified in Section II:

- Volumetric convolutional modules composed of K repeated 3D convolutions with F filters followed by non-linearities and batch normalization transformations. The architecture of the resulting employed modules is shown in Figure 1a. Inspired by the VGG architecture [41], repeated application of volumetric convolutions allows fast growing of the receptive field while keeping the number of parameters at the neccessary minimum, thus making the whole architecture more computationally efficient. We also apply batch normalization after the first fully connected layer and all the convolutional layers in order to accelerate network convergence [42].
- Volumetric residual (VoxRes) blocks composed of K repeated 3D convolutions with F filters followed by non-linearities and batch normalization transformations. In each VoxRes module, the input feature x and transformed feature $g(x)$ are added together with skip connection as shown in Figure 1b, hence the information can be directly propagated in the forward and backward passes [31]. In our interpretation, VoxRes modules can be viewed as volumetric convolutional modules with skip connections.
- To deal with the temporal dependencies, we opt for the standard long short-term memory (LSTM) cells [43]. As the input to our network is composed of 3D images, we transform it into a feature map via a 3D CNN, which leads us to ConvLSTM units [34].

Starting from these components, we specify 12 architecture variants of 3D convolutional and recurrent-3D convolutional neural networks, as summarized in Tables II–III. We further describe the proposed network architectures below. Note that VoxCNN-based configurations (VoxCNN-A and VoxCNN-B) and VoxResNet-based configurations (VoxResNet-A through D) are only applied to structural MRI data, while models RCNN-A through F deal with functional MRI data only.



(a) (b)

Fig. 1: A schematic visualization of modules used in the construction of the investigated architectures: (a) volumetric convolutional modules, (b) volumetric residual modules.

a) VoxCNN: We examine two network configurations based on the VGG-like architecture with volumetric convolutions [33] (A–B, see Table II, (a)). Both VoxCNN-A and VoxCNN-B networks consist of four volumetric convolutional blocks with different number of Conv3D layers. We start with a somewhat more computationally efficient VoxCNN-A model and continue with a richer but computationally more demanding VoxCNN-B. All of the Conv3D layers contain filters of size $3 \times 3 \times 3$ voxels with the number of filters increasing by the factor of two in each next block. The last layer of the network contains two units and uses softmax activation function to output probabilities for binary classification.

b) VoxResNet: We compare several VoxResNet configurations (A–D, see Table II, (b)) with various numbers of residual blocks and numbers of filters in each convolutional layer. When building these models, we are inspired by the successful application of VoxResNets for the clinical diagnostics [31], [33]. As was the case with VoxCNN architectures, we start with computationally cheap VoxResNet-A and VoxResNet-B models, differing in depth but similar in terms of number of convolutional filters. We further investigate the influence of network capacity on the recognition performance by doubling the nuber of filters in each convolution layer for configurations VoxResNet-C and VoxResNet-D. However, we also use filters with stride 2 in the first convolutional layer of these networks in order to reduce the dimension of the original image to meet the limitations of the GPU memory.

c) Recurrent-Convolutional Neural Network: Since each fMRI scanning result is represented as a sequence of 3D images, we investigate a series of recurrent-convolutional neural networks (RCNNs) inspired by their recent success in recognition of schizophrenia [35].

The architecture consists of a convolutional part followed by a recurrent part. During the model training process, the input scans are sequentially fed to the convolutional part of the network, which aims to extract complex spatial features while simultaneously reducing the dimensionality of data. Then features describing 3D tensor of the input object, are transmitted to the recurrent part of the network, which pre-

TABLE II: Specification of the 3D network configurations applied to structural MRI data

(a)

VoxCNN	
A	**B**
Conv3D, 8 Conv3D, 8 Conv3D, 8	Conv3D, 8 Conv3D, 8
MaxPool3D	
Conv3D, 16 Conv3D, 16 Conv3D, 16	Conv3D, 16 Conv3D, 16
MaxPool3D	
Conv3D, 32 Conv3D, 32	Conv3D, 32 Conv3D, 32 Conv3D, 32
MaxPool3D	
Conv3D, 64 Conv3D, 64	Conv3D, 64 Conv3D, 64 Conv3D, 64
MaxPool3D	
FullyConnected, 128 Dropout	
FullyConnected, 64	
Output, 2 classes	

(b)

VoxResNet			
A	**B**	**C**	**D**
Conv3D, 16 Conv3D, 16	Conv3D, 16 Conv3D, 16	Conv3D, 32, stride 2 Conv3D, 32	Conv3D, 32, stride 2 Conv3D, 32
Conv3D, 32, stride 2		Conv3D, 64, stride 2	
VoxRes, 32 VoxRes, 32	VoxRes, 32 VoxRes, 32	VoxRes, 64 VoxRes, 64	VoxRes, 64 VoxRes, 64
Conv3D, 32, stride 2		Conv3D, 64, stride 2	
VoxRes, 32 VoxRes, 32	VoxRes, 32 VoxRes, 32	VoxRes, 64 VoxRes, 64	VoxRes, 64 VoxRes, 64
Conv3D, 64, stride 2		Conv3D, 128, stride 2	
VoxRes, 64 VoxRes, 64	VoxRes, 64 VoxRes, 64	VoxRes, 128 VoxRes, 128	VoxRes, 128 VoxRes, 128
	Conv3D, 64, stride 2		Conv3D, 128, stride 2
	VoxRes, 64 VoxRes, 64		VoxRes, 128 VoxRes, 128
MaxPool3D			
FullyConnected, 128 Dropout			
Output, 2 classes			

TABLE III: Specification of the 4D network configurations applied to functional MRI data

RCNN					
A	**B**	**C**	**D**	**E**	**F**
Conv3D, 16 Conv3D, 16	Conv3D, 16 Conv3D, 16	Conv3D, 16 Conv3D, 16	Conv3D, 16 Conv3D, 16	Conv3D, 16 Conv3D, 16	Conv3D, 16 Conv3D, 16
MaxPool3D					
	Conv3D, 32 Conv3D, 32	Conv3D, 32 Conv3D, 32		Conv3D, 32 Conv3D, 32	Conv3D, 32 Conv3D, 32
	MaxPool3D			MaxPool3D	
		Conv3D, 64			Conv3D, 64
		MaxPool3D			MaxPool3D
LSTM					
			LSTM		
FullyConnected, 512 Dropout					
Output, 2 classes					

dicts the class label, taking into account dependence between sequential changes of these features.

Similarly to the VoxCNN and VoxResNet models, we evaluate several configurations containing various numbers of convolutional blocks (A–F, see Table III) with increasing computational complexity. The recurrent part consists of one or more layers with LSTM units and takes as input sequence of the feature vectors extracted by the convolutional part from the images in the fMRI series. We tested configurations with up to two LSTM layers with 128 memory cells in each followed by the fully connected layer with 512 neurons. Only outputs of the last LSTM layer after it has seen all the elements of the sequence are used as inputs for the fully connected layer. Last layer of the network consist of two neurons and uses softmax activation function. Various dropout rates for the fully connected layer were also considered in order to control the network overfitting.

C. Training details

Training of all the presented models is carried out by optimizing the cross-entropy loss function. We use AdaM optimizer with learning rate of 1×10^{-4} for both VoxCNN and VoxResNet networks and 3×10^{-5} for RCNN. Values of first and second moments are set to 0.9 and 0.99, respectively.

Because of the large size of both MRI and fMRI objects, we train our models with batch size of 4 samples. Dropout probability is tuned for each network configuration of the described architectures and both classification problems. In general, we use more strong regularization when applying same model to the problem without patients with mixed pathology. It is based on the fact that in these tasks approximately half of the patients is removed and, hence, the model is more prone to overfitting.

We have trained VoxCNN for 400 epochs, VoxResNet for 50 epochs, and RCNN for 80 epochs for *DvsH* task and for 50 epochs for *DvND* task. These values were determined for each architecture depending on learning rate and dropout probability in such a way as to achieve maximal performance on the training dataset and stable results on the test set.

The quality of model predictions is estimated by metric ROC AUC (Area Under ROC-Curve). Because of the limited size of dataset, repeated 5-fold cross-validation with 3

TABLE IV: Summary of the performed experiments

Task	Data modality	Cleaning	Models
EvsNE	structural		VoxCNN-*, VoxResNet-*
EvsH	structural		VoxCNN-*, VoxResNet-*
DvsND	structural		VoxCNN-*, VoxResNet-*
DvsND	functional		RCNN-*
DvsND	functional	✓	RCNN-*
DvsH	structural		VoxCNN-*, VoxResNet-*
DvsH	functional		RCNN-*
DvsH	functional	✓	RCNN-*

repetitions is used in order to get more accurate estimation of classification performance.

Training details for RCNN. Because of the limitations of GPU memory (we have 16 GB GPU memory at our disposal) and large total size of the fMRI sequences, they were divided into smaller segments of length 16 time points ($16 \times 3.7 = 59.2$ seconds in total). Thus, from the series for each object, $130 - 16 + 1 = 115$ different subsequences were obtained and used to form training and test sets. However, training and evaluation are performed only on subsequences belonging to different objects in order to prevent information used for testing from getting into the training set.

D. Design of the computational experiments

In order to perform a comprehensive evaluation of the proposed architectures, we consider four different diagnostic tasks, that we formulate as binary classification problems:

- *EvsNE* – recognition of subjects diagnosed with epilepsy (including patients with depression) against subjects without epilepsy (including patients with depression) classification;
- *EvsH* – classification of subjects diagnosed with epilepsy versus healthy control;
- *DvsND* – classification of depression (including patients with epilepsy) against subjects without depression (including patients with epilepsy) classification;
- *DvsH* – classification of depression versus healthy control classification.

The summary of our experiments is in Table IV. We use data of structural MRI for the *EvsNE* and *EvsH* classifications and functional MRI data for the *DvsND* and *DvsH* tasks.

IV. Experimental results

This section presents the results obtained for our four binary classification tasks. The results for the problems which were solved with structural MRI data using VoxCNN and VoxResNet network architectures are shown in the Table V.

A. Epilepsy recognition performance

Performance of the two VoxCNN modifications is similar to that of VoxResNet-A and VoxResNet-B, but it should be noted that it took much longer for VoxCNN networks to converge, marking the advantages of residual architectures for practical use. Both VoxResNet-C and VoxResNet-D configuration show better classification results for *EvsNE* task than VoxResNet-A and VoxResNet-B in spite of applying downsampling with convolutions with stride of 2 in the first layer and potentially discarding a part of input information.

TABLE V: Performance obtained on the structural MRI classification tasks. Reported is ROC-AUC on the 5-fold cross-validation

Model	EvsNE	EvsH	DvsND	DvsH
VoxCNN-A	0.66 ± 0.09	0.69 ± 0.11	0.49 ± 0.11	0.6 ± 0.16
VoxCNN-B	0.69 ± 0.06	0.73 ± 0.10	0.55 ± 0.12	0.58 ± 0.17
VoxResNet-A	0.65 ± 0.11	0.76 ± 0.18	0.57 ± 0.11	0.66 ± 0.15
VoxResNet-B	0.69 ± 0.11	0.66 ± 0.20	0.52 ± 0.16	0.63 ± 0.20
VoxResNet-C	0.71 ± 0.12	0.69 ± 0.17	0.52 ± 0.12	0.61 ± 0.19
VoxResNet-D	0.73 ± 0.08	0.61 ± 0.19	0.54 ± 0.10	0.62 ± 0.19

TABLE VI: Performance obtained on the functional MRI classification tasks. Reported is ROC-AUC on the 5-fold cross-validation

Model	DvsND	DvsH	DvsND (cleaned)	DvsH (cleaned)
RCNN-A	0.53 ± 0.11	0.62 ± 0.12	0.56 ± 0.03	0.60 ± 0.10
RCNN-B	0.54 ± 0.11	0.68 ± 0.09	0.66 ± 0.12	0.72 ± 0.17
RCNN-C	0.63 ± 0.07	0.66 ± 0.10	0.67 ± 0.09	0.73 ± 0.17
RCNN-D	0.58 ± 0.10	0.64 ± 0.11	0.55 ± 0.07	0.63 ± 0.14
RCNN-E	0.64 ± 0.09	0.63 ± 0.18	0.67 ± 0.08	0.73 ± 0.11
RCNN-F	0.61 ± 0.12	0.70 ± 0.15	0.64 ± 0.10	0.69 ± 0.13

In average, the results of *EvsNE* task are comparable with those of *EvsH* task. Nonetheless, as can be seen by the standard deviation, most of the results in the latter task are very unstable and strongly depend on the subjects in the training set, which makes conclusions drawn on the basis on their comparison not sufficiently reliable. Moreover, a comparison of the best and worst results for this problem by Student's t-test did not reveal a significant difference at the significance level of 0.05. It may be caused by the training sample size, which is considerably smaller for *EvsH* and *DvsH* problems due to removing patients with mixed pathologies. In case of *EvsH* task, this effect can be additionally strengthened by diversity of types and focus localizations of the epilepsy represented in the provided dataset.

As can be observed, VoxCNN-B model achieves the highest results in *EvsNE* and *EvsH* tasks. It also demonstrates an acceptable class separation comparing to the best results both in the presence of patients with depression, and without them.

B. Depression recognition performance

The Table VI demonstrates the performance of various modifications of the RCNN model trained on the functional MRI data for *DvsND* and *DvsH* classification tasks. In these evaluations RCNN-A, RCNN-B, and RCNN-E models were unable to deliver satisfactory results. We hypothesize that this may be due to the fact that only two convolutional layers may

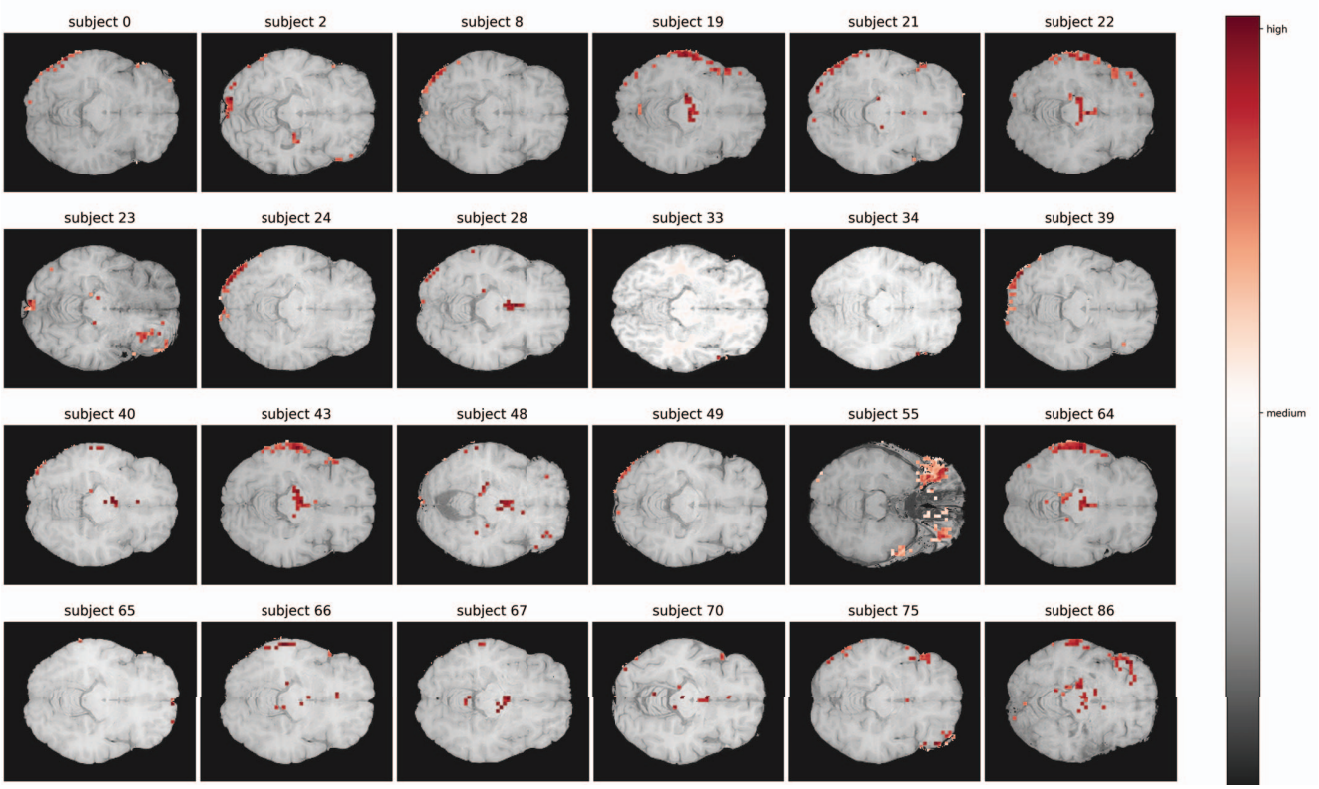


Fig. 2: Visualization of neural network attention for subjects with medial temporal lobe epilepsy.

be insufficient to extract features that describe the data well. We note, however, that extracting only the relevant components from the data improves the results significantly, see Table VI, cleaned data.

In spite of a limited sample size, all of the evaluated configurations perform consistently better in the problem of depression recognition without adding patients with mixed pathology. This is consistent with the fact that the presence of patients with both diseases complicates detection of one of them.

The highest ROC-AUC scores and adequate standard deviation for both *DvsND* and *DvsH* classification tasks are reported by RCNN-F configuration. However, analysis of the features obtained in this tasks has shown that they are related to different brain regions and thus can not be considered as stable depression biomarkers.

C. Visualization of the feature importance

The network attention is visualized as heatmap by the Grad-CAM method [44]. Fig. 2 demonstrate the regions that were considered the most important by the classification model for *EvsNE* task in subjects diagnosed with medial temporal lobe epilepsy (MTLE). However, although some regions coincide and can be considered as potentially stable biomarkers, further investigation is required to draw the final conclusions.

V. Discussion and future work

In the scope of this study, we sought to evaluate off-the-shelf building modules for deep learning architectures aimed at the recognition of pathologies in neuroimaging data.

With this goal, we examined three classes of deep neural network architectures, which were demonstrated to perform well for MRI and fMRI data analysis in the literature. Namely, we tested various modifications of VoxCNN and VoxResNet models in order to solve epilepsy recognition task using structural MRI scans and voxelwise recurrent-convolutional networks for classification of patients with and without depression based on their functional MRI series.

We have achieved 0.73 ROC-AUC score when classifying depression versus a control group, but adding a sample of patients with epilepsy makes classification more challenging. From a medical point of view, this could be explained by the fact that depression and epilepsy potentially share common pathological pathways.

In case of epilepsy, classifications with and without the presence of patients suffering from depression showed similar results. This may be caused by the fact that patients with epilepsy can differ in the type of disease, and in this case, the removal of patients with mixed pathology from the sample makes it inadequate for a more successful classification.

Next, we identify most informative features for depression and epilepsy classification problems using a method for CNN

attention visualization. We compared areas obtained in the *EvsH* and *EvsNE* classification tasks to figure out whether they represent same brain regions. The same was performed for depression.

For epilepsy we revealed some potentially significant brain regions, which were found significant for both Epilepsy vs Healthy Control and Epilepsy vs No Epilepsy (including patients with major depression disorder) and could represent the aim for further investigations.

Further algorithmic improvement of the obtained results can be done through usage of advanced computationally efficient sparse convolutional neural networks [45], combined with efficient approaches to initialization [46] and construction of ensembles [47]. In order to tackle natural class imbalance we will use resampling approaches [48]–[50] along with uncertainty quantification based on conformal prediction framework [51], [52].

The considered problem was stated in the scope of the Project "Machine Learning and Pattern Recognition for the development of diagnostic and clinical prognostic prediction tools in psychiatry, borderline mental disorders, and neurology" (a part of the Skoltech Biomedical Initiative program).

ACKNOWLEDGMENT

The work was supported by The Ministry of Education and Science of Russian Federation, grant No. 14.615.21.0004, grant code: RFMEFI61518X0004.

REFERENCES

[1] G. S. Bell, A. Neligan, and J. W. Sander, "An unknown quantitythe worldwide prevalence of epilepsy," *Epilepsia*, vol. 55, no. 7, pp. 958–962, 2014.

[2] C. L. Harden and M. A. Goldstein, "Mood disorders in patients with epilepsy," *CNS drugs*, vol. 16, no. 5, pp. 291–302, 2002.

[3] M. Dudra-Jastrzêbska, M. M. Andres-Mach, J. J. Luszczki, and S. J. Czuczwar, "Mood disorders in patients with epilepsy," *Pharmacological reports*, vol. 59, no. 4, p. 369, 2007.

[4] A. M. Kanner, S. C. Schachter, J. J. Barry, D. C. Hersdorffer, M. Mula, M. Trimble, B. Hermann, A. E. Ettinger, D. Dunn, R. Caplan, *et al.*, "Depression and epilepsy: epidemiologic and neurobiologic perspectives that may explain their high comorbid occurrence," *Epilepsy & Behavior*, vol. 24, no. 2, pp. 156–168, 2012.

[5] B. C. Bernhardt, S.-J. Hong, A. Bernasconi, and N. Bernasconi, "Magnetic resonance imaging pattern learning in temporal lobe epilepsy: classification and prognostics," *Annals of neurology*, vol. 77, no. 3, pp. 436–446, 2015.

[6] J. D. Rudie, J. B. Colby, and N. Salamon, "Machine learning classification of mesial temporal sclerosis in epilepsy patients," *Epilepsy research*, vol. 117, pp. 63–69, 2015.

[7] V. Sujitha, P. Sivagami, and M. Vijaya, "Support vector machine based epilepsy prediction using textural features of mri," *Procedia Computer Science*, vol. 2, pp. 283–290, 2010.

[8] C. Rummel, N. Slavova, A. Seiler, E. Abela, M. Hauf, Y. Burren, C. Weisstanner, S. Vulliemoz, M. Seeck, K. Schindler, *et al.*, "Personalized structural image analysis in patients with temporal lobe epilepsy," *Scientific reports*, vol. 7, no. 1, p. 10883, 2017.

[9] J. Wellmer, C. M. Quesada, L. Rothe, C. E. Elger, C. G. Bien, and H. Urbach, "Proposal for a magnetic resonance imaging protocol for the detection of epileptogenic lesions at early outpatient stages," *Epilepsia*, vol. 54, no. 11, pp. 1977–1987, 2013.

[10] X. Wang, Y. Ren, and W. Zhang, "Depression disorder classification of fmri data using sparse low-rank functional brain network and graph-based features," *Computational and mathematical methods in medicine*, vol. 2017, 2017.

[11] C. H. Fu, J. Mourao-Miranda, S. G. Costafreda, A. Khanna, A. F. Marquand, S. C. Williams, and M. J. Brammer, "Pattern classification of sad facial processing: toward the development of neurobiological markers in depression," *Biological psychiatry*, vol. 63, no. 7, pp. 656–662, 2008.

[12] L.-L. Zeng, H. Shen, L. Liu, L. Wang, B. Li, P. Fang, Z. Zhou, Y. Li, and D. Hu, "Identifying major depression using whole-brain functional connectivity: a multivariate pattern analysis," *Brain*, vol. 135, no. 5, pp. 1498–1507, 2012.

[13] A. P. Zijdenbos, R. Forghani, and A. C. Evans, "Automatic" pipeline" analysis of 3-d mri data for clinical trials: application to multiple sclerosis," *IEEE transactions on medical imaging*, vol. 21, no. 10, pp. 1280–1291, 2002.

[14] M. de Bruijne, "Machine learning approaches in medical image analysis: From detection to diagnosis," 2016.

[15] F. Pereira, T. Mitchell, and M. Botvinick, "Machine learning classifiers and fmri: a tutorial overview," *Neuroimage*, vol. 45, no. 1, pp. S199–S209, 2009.

[16] S. Chernova and E. Burnaev, "On an iterative algorithm for calculating weighted principal components," *Journal of Communications Technology and Electronics*, vol. 60, pp. 619–624, Jun 2015.

[17] B. Fischl, "Freesurfer," *Neuroimage*, vol. 62, no. 2, pp. 774–781, 2012.

[18] A. Abraham, F. Pedregosa, M. Eickenberg, P. Gervais, A. Mueller, J. Kossaifi, A. Gramfort, B. Thirion, and G. Varoquaux, "Machine learning for neuroimaging with scikit-learn," *Frontiers in neuroinformatics*, vol. 8, p. 14, 2014.

[19] S. Whitfield-Gabrieli and A. Nieto-Castanon, "Conn: a functional connectivity toolbox for correlated and anticorrelated brain networks," *Brain connectivity*, vol. 2, no. 3, pp. 125–141, 2012.

[20] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *nature*, vol. 521, no. 7553, p. 436, 2015.

[21] A. Sharif Razavian, H. Azizpour, J. Sullivan, and S. Carlsson, "Cnn features off-the-shelf: an astounding baseline for recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp. 806–813, 2014.

[22] A. Babenko, A. Slesarev, A. Chigorin, and V. Lempitsky, "Neural codes for image retrieval," in *European conference on computer vision*, pp. 584–599, Springer, 2014.

[23] S. S. Keller and N. Roberts, "Voxel-based morphometry of temporal lobe epilepsy: An introduction and review of the literature," *Epilepsia*, vol. 49, no. 5, pp. 741–757, 2008.

[24] W. Liao, Z. Zhang, Z. Pan, D. Mantini, J. Ding, X. Duan, C. Luo, Z. Wang, Q. Tan, G. Lu, *et al.*, "Default mode network abnormalities in mesial temporal lobe epilepsy: a study combining fmri and dti," *Human brain mapping*, vol. 32, no. 6, pp. 883–895, 2011.

[25] G. Bettus, F. Bartolomei, S. Confort-Gouny, E. Guedj, P. Chauvel, P. J. Cozzone, J.-P. Ranjeva, and M. Guye, "Role of resting state functional connectivity mri in presurgical investigation of mesial temporal lobe epilepsy," *Journal of Neurology, Neurosurgery & Psychiatry*, pp. jnnp–2009, 2010.

[26] A. J. Golby, R. A. Poldrack, J. Illes, D. Chen, J. E. Desmond, and J. D. Gabrieli, "Memory lateralization in medial temporal lobe epilepsy assessed by functional mri," *Epilepsia*, vol. 43, no. 8, pp. 855–863, 2002.

[27] H. M. Bonnici, M. Sidhu, M. J. Chadwick, J. S. Duncan, and E. A. Maguire, "Assessing hippocampal functional reserve in temporal lobe epilepsy: a multi-voxel pattern analysis of fmri data," *Epilepsy research*, vol. 105, no. 1-2, pp. 140–149, 2013.

[28] J. Zhang, W. Cheng, and et al., "Pattern classification of large-scale functional brain networks: identification of informative neuroimaging markers for epilepsy," *PloS one*, vol. 7, no. 5, p. e36733, 2012.

[29] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3d u-net: learning dense volumetric segmentation from sparse annotation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 424–432, Springer, 2016.

[30] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *3D Vision (3DV), 2016 Fourth International Conference on*, pp. 565–571, IEEE, 2016.

[31] H. Chen, Q. Dou, L. Yu, and P.-A. Heng, "Voxresnet: Deep voxelwise residual networks for volumetric brain segmentation," *arXiv preprint arXiv:1608.05895*, 2016.

[32] E. Hosseini-Asl, G. Gimel'farb, and A. El-Baz, "Alzheimer's disease diagnostics by a deeply supervised adaptable 3d convolutional network," *arXiv preprint arXiv:1607.00556*, 2016.

[33] S. Korolev, A. Safiullin, M. Belyaev, and Y. Dodonova, "Residual and plain convolutional neural networks for 3d brain mri classification," in *Biomedical Imaging (ISBI 2017), 2017 IEEE 14th International Symposium on*, pp. 835–838, IEEE, 2017.

[34] N. C. Dvornek, P. Ventola, K. A. Pelphrey, and J. S. Duncan, "Identifying autism from resting-state fmri using long short-term memory networks," in *International Workshop on Machine Learning in Medical Imaging*, pp. 362–370, Springer, 2017.

[35] J. Dakka, P. Bashivan, M. Gheiratmand, I. Rish, S. Jha, and R. Greiner, "Learning neural markers of schizophrenia disorder using recurrent neural networks," *arXiv preprint arXiv:1712.00512*, 2017.

[36] P. Bashivan, I. Rish, M. Yeasin, and N. Codella, "Learning representations from eeg with deep recurrent-convolutional neural networks," *arXiv preprint arXiv:1511.06448*, 2015.

[37] M. J. McKeown, L. K. Hansen, and T. J. Sejnowsk, "Independent component analysis of functional mri: what is signal and what is noise?," *Current opinion in neurobiology*, vol. 13, no. 5, pp. 620–629, 2003.

[38] L. Griffanti, G. Douaud, J. Bijsterbosch, S. Evangelisti, F. Alfaro-Almagro, M. F. Glasser, E. P. Duff, S. Fitzgibbon, R. Westphal, D. Carone, *et al.*, "Hand classification of fmri ica noise components," *Neuroimage*, vol. 154, pp. 188–205, 2017.

[39] M. Sharaev, A. Andreev, A. Artemov, E. Burnaev, E. Kondratyeva, S. Sushchinskaya, I. Samotaeva, V. Gaskin, and A. Bernstein, "Pattern recognition pipeline for neuroimaging data," in *Artificial Neural Networks in Pattern Recognition (ANNPR-2018). Lecture Notes in Computer Science* (L. Pancioni, F. Schwenker, and E. Trentin, eds.), vol. 11081, pp. 306–319, Springer, 2018.

[40] M. Sharaev, A. Artemov, E. Kondratyeva, S. Sushchinskaya, E. Burnaev, A. Bernstein, R. Akzhigitov, and A. Andreev, "Mri-based diagnostics of depression concomitant with epilepsy: in search of the potential biomarkers," in *2018 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, IEEE.

[41] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[42] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, 2015.

[43] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[44] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization.," in *ICCV*, pp. 618–626, 2017.

[45] A. Notchenko, Y. Kapushev, and E. Burnaev, "Large-scale shape retrieval with sparse 3d convolutional neural networks," in *Analysis of Images, Social Networks and Texts* (W. M. van der Aalst, D. I. Ignatov, M. Khachay, S. O. Kuznetsov, V. Lempitsky, I. A. Lomazova, N. Loukachevitch, A. Napoli, A. Panchenko, P. M. Pardalos, A. V. Savchenko, and S. Wasserman, eds.), (Cham), pp. 245–254, Springer International Publishing, 2018.

[46] P. D. Erofeev and E. V. Burnaev, "The influence of parameter initialization on the training time and accuracy of a nonlinear regression model," *Journal of Communications Technology and Electronics*, vol. 61, pp. 646–660, Jun 2016.

[47] P. V. Prikhod'ko and E. V. Burnaev, "On a method for constructing ensembles of regression models," *Automation and Remote Control*, vol. 74, pp. 1630–1644, Oct 2013.

[48] D. Smolyakov, P. Erofeev, and E. Burnaev, "Model selection for anomaly detection," in *Proc. SPIE 9875, Eighth International Conference on Machine Vision, Barcelona, Spain (December 8, 2015)* (A. Verikas, P. Radeva, and D. Nikolaev, eds.), vol. 9875, SPIE, 2015.

[49] A. Papanov, P. Erofeev, and E. Burnaev, "Influence of resampling on accuracy of imbalanced classification," in *Proc. SPIE 9875, Eighth International Conference on Machine Vision, Barcelona, Spain (December 8, 2015)* (A. Verikas, P. Radeva, and D. Nikolaev, eds.), vol. 9875, SPIE, 2015.

[50] D. Smolyakov and E. Burnaev, "One-class SVM with privileged information and its application to malware detection," in *IEEE International Conference on Data Mining Workshops, ICDM Workshops 2016, December 12-15, 2016, Barcelona, Spain.* (C. Domeniconi, F. Gullo, F. Bonchi, J. Domingo-Ferrer, R. A. Baeza-Yates, Z. Zhou, and X. Wu, eds.), pp. 273–280, IEEE Computer Society, 2016.

[51] V. Vovk and E. Burnaev, "Efficiency of conformalized ridge regression," *CoRR*, vol. abs/1404.2083, 2014.

[52] I. Nazarov and E. Burnaev, "Conformalized kernel ridge regression," in *15th IEEE International Conference on Machine Learning and Applications, ICMLA 2016, Anaheim, CA, USA, December 18-20, 2016*, pp. 45–52, IEEE, 2016.