Insert here your thesis' task.

FACULTY
OF INFORMATION
TECHNOLOGY
CTU IN PRAGUE

Master's thesis

Imaging-Based Diagnostic Classification of ADHD

Bc. Ester Primasová

Department of Applied Mathematics
Supervisor: Ing. Magda Friedjungová, Ph.D.

September 15, 2022

Acknowledgements

I would like to thank my family and friends for support during writing this thesis.

Declaration

I hereby declare that the presented thesis is my own work and that I have cited all sources of information in accordance with the Guideline for adhering to ethical principles when elaborating an academic final thesis.

I acknowledge that my thesis is subject to the rights and obligations stipulated by the Act No. 121/2000 Coll., the Copyright Act, as amended. In accordance with Article 46 (6) of the Act, I hereby grant a nonexclusive authorization (license) to utilize this thesis, including any and all computer programs incorporated therein or attached thereto and all corresponding documentation (hereinafter collectively referred to as the "Work"), to any and all persons that wish to utilize the Work. Such persons are entitled to use the Work in any way (including for-profit purposes) that does not detract from its value. This authorization is not limited in terms of time, location and quantity. However, all persons that makes use of the above license shall be obliged to grant a license at least in the same scope as defined above with respect to each and every work that is created (wholly or in part) based on the Work, by modifying the Work, by combining the Work with another work, by including the Work in a collection of works or by adapting the Work (including translation), and at the same time make available the source code of such work at least in a way and scope that are comparable to the way and scope in which the source code of the Work is made available.

In Prague on September 15, 2022 . . . . . .

Czech Technical University in Prague
Faculty of Information Technology

Citation of this thesis

Primasová, Ester. *Imaging-Based Diagnostic Classification of ADHD*. Master's thesis. Czech Technical University in Prague, Faculty of Information Technology, 2022.

Abstrakt

Diplomová práce s názvem Diagnostická klasifikace ADHD na základě zobrazovacích metod se zabývá výzkumem existujících technik diagnostiky ADHD se speciálním zřetelem na snímky funkční magnetické rezonance (fMRI). Autorka diplomové práce nabízí vlastní model pro klasifikaci ADHD diagnostiky, a to skrze využití zobrazovací metody fMRI a hlubokých neuronových sítí. V diplomové práci jsou následně porovnány výsledky autorčina modelu s existujícími klasifikačními modely používající data z Celosvětové soutěže ADHD-200. Autorka diplomové práce potvrzuje funkčnost vlastního modelu, který nabízí výsledky diagnostiky ADHD srovnatelné s existujícími metodami zobrazení. Diplomová práce nabízí prostor pro další a hlubší výzkum, který by autorčin model mohl etablovat v prostředí medicínské diagnostiky.

Klíčová slova Diagnostická klasifikace ADHD, rs-fMRI, Lékařské zobrazování, Hluboké učení, CNN, RNN, LSTM, GRU

Abstract

The thesis entitled Imaging-Based Diagnostic Classification of ADHD deals with the research of existing techniques for ADHD diagnosis with special reference to functional magnetic resonance imaging (fMRI). The author of the

thesis proposes her own model for ADHD diagnostic classification through the use of fMRI imaging and deep neural networks. The thesis then compares the results of the author's model with existing classification models using data from the ADHD-200 Worldwide Competition. The author's thesis confirms the performance of her own model, which offers ADHD diagnostic results comparable to existing imaging methods. The thesis offers scope for further and deeper research to establish the author's model in a medical diagnostic setting.

Keywords ADHD Diagnostic Classification, rs-fMRI, Medical Imaging, Deep Learning, CNN, RNN, LSTM, GRU

Contents

List of Figures

List of Tables

Introduction

Motivation and objectives

Attention deficit hyperactivity disorder (ADHD) is one of the most common neurodevelopmental disorder among children and adolescent that has a detrimental impact on the brain neurodevelopment as compared to controls. ADHD individuals exhibit inattention, impulsiveness, hyperactivity or its combination. This condition, especially with the hyperactivity, is more common in males compared with females.[4]

It has early onset on life and usually persists into adulthood, although young adults often show reduced hyperactivity and impulsivity while retaining symptoms of inattention. Persisting to adulthood in the majority of patients ADHD is associated with serious psychosocial impairment and a high comorbidity rate including anxiety disorder, depressive disorder and learning problem, yet it is currently underdiagnosed and treated in many European countries. This leads to great levels of personal suffering, ineffective treatment when ADHD is misdiagnosed and higher societal costs, if left unidentified and untreated.[5]

Patients with ADHD often experience neuropsychological difficulties such as deficient inhibition, memory, executive functioning, decision making, and emotional dysregulation which may have negative consequences for adult's self-esteem and the quality of relationships.[6]

A meta-analysis of 19 studies with over 55,000 participants found that in children and adolescents (5.9-7.1 %) meet diagnostic criteria for ADHD and by self-report measures (5.0 %) in young adults.[7] The diagnosis of ADHD mainly depends on a clinical evaluation of behavioral symptoms with responses from patients, parents and informants.[8]

"Most diagnostic guidelines require that ADHD be assessed and diagnosed by relying on information provided via a variety of methods (e.g., clinical interviews, observations and ratings) and collected from multiple sources (e.g.,

parents and teachers). However, using subjective measures always incorporates the risk ofinformant biases and clinicians are often confronted with great inconsistencies between ratings obtained from different sources."[9] In response to all these concerns, usage of objective measures might income valuable information for diagnosing ADHD and researchers .

Problem statements

For the reason that convention-based diagnostic systems not undoubtly reflecting neurobiological pathomechanisms, the appeal to identify clinically useful biomarkers for objective diagnosis is growing. Criteria for these biological biomarkers were defined by the workers of the World Federation of ADHD. The biomarker should be reliable, non-invasive, reproducible, inexpensive, easy to use, and confirmed by at least two independent studies counceled by qualified investigators and also have at least 80% sensitivity and at least 80% specificity. Unfortunatelly no reliable biomarker met these criteria yet.[10]

In 2011, the ADHD-200 Consortium announced "ADHD-200 Global Competition" with the aim of identifying biomarkers and developing diagnostic classification tools of ADHD disorder and its subtypes from resting-state functional magnetic resonance imaging (rs-fMRI), structural MRI (sMRI) of the brain and personal characteristics (phenotypic information).

The ADHD-200 dataset was collected on 776 children, adolescents and young adults from eight independent imaging sites and included 491 datasets collected from typically developing individuals and 285 from children and adolescents diagnosed with ADHD.[11] It was the first publicly available dataset with fMRI scans with huge amount of participants, including both psychiatric patients and healthy controls, compared to all previous studies that included from 20 to 104 participants.[12]

Fifty teams from around the world, with diverse backgrounds, including mathematics, statistics, computer science and neuroscience, joined the competition and 21 of them finally submitted diagnostic labels with predomination of machine learning models.[13] Considering the fact that many competitors lacked the specialized knowledge of neuroimaging methods to implement the necessary steps of data preparation for rs-fMRI or access to computing resources with high performance, the Neuro Bureau collaborated with all competitors by preprocessing the data and sharing the "ADHD-200 Preprocessed" version at the Neuroimaging Informatics Tools and Resources Clearinghouse NITRC.[13]

The winner was a team from Johns Hopkins University with an ensemble model which achieved a 61% accuracy, 94% specificity and 21% sensitivity.

That shows us that diagnosic classification based on fMRI images can be developed with a very low risk of false positives.[11] In spite of the quite humble accuracy, the ADHD-200 dataset has become the main source of data for machine learning based development for ADHD diagnostic classification and many more researchers did their best to overcome it. Especially deep learning methods had promising outcome in achieving improved classification performance, though there is still room for improvement.[13]

The aim of this thesis is to research the medical imaging based on MRI, describe the state-of-the-art methods for ADHD classification and also implement my own prototype for ADHD diagnostic classification using the publically available ADHD-200 dataset.

Chapter **??** and 3 introduce the reader to medical imaging domain focused on magnetic resonance imaging, ADHD disorder, related work on its classification focused on ADHD-200 dataset and state-of-the-art techniques. There is also provided necessary machine learning background though some knowledge in this area is expected from the reader. In Chapter 4, the main freely available dataset for ADHD diagnostic classification, ADHD-200, is described in detail. That include data analysis, the spread of the data and preprocessing methodologies. Chapter **??** describe proposed network architecture inspired by the selected paper from Chapter 3, process of its implementation, experiments, hyperparameter optimization and evaluation. Finally, the results of the experiments are summarized and discussed alongside with reflection on future work are in Chapter 6.

Chapter 1

Machine Learning background

This chapter deals with information from the field of machine learning / deep learning with emphasis on explaining the basic principles of recurrent neural networks.

Pocita se s tim, ze je ctenar obeznamen se zaklady machine learningu. V teto kapitole budu zbezne venovat vysvetleni RNN, LSTM, GRU, Attention. V pripade potreby doporucuji prostudovat [14] pripadne literaturu doporucenou v jednotlivych sekcich.

1.1 Recurrent neural networks

In this section will be introduced recurrent neural networks (RNNs), that are primarily used to detect patterns in a sequence of data.[14]

Standard feed-forward ANNs have inability to remember historical data and has no knowledge of order in time. If each part of input is generated independently, it can works well, but if they are related in time or space or are anyhow dependent, the memory inside the model is needed.

RNNs contain an internal memory and may create recurrent connections between hidden units. Current value of a variable can influence its own value at a future. Another feature of most RNNs is possibility to process sequences of variable length and parameter sharing across model.[14]

That suits for problems such as natural language translation, speech recognition, video analysis, image captioning, time series predictions, handwriting recognition and many others, where the data sequences are processed.[15, 1, 16]

1.1.1 Backpropagation Through Time

Backpropagation Through Time (BPTT) is algorithm widely used in RNNs. By knowledge, that for every RNN with a finite number of operations, there exists an corresponding traditional feedforward network with the same parameters repeated through, usual backpropagation training algorithm can be applied to the unfolded RNNs.[17]

If the backpropagation algorithm is applied to the whole unrolled computational graph with respect to each parameter's occurence, the gradient is obtained and may be further used by any gradient based techniques to train an RNN.[14]

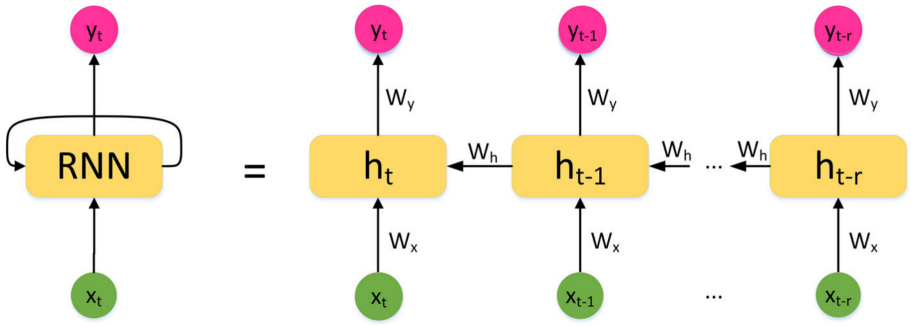


Figure 1.1: An standard RNN and the same network unfolded over time t. Picture adapted from [1].

Figure 1.1.1 shows the standard structure of RNN and the corresponding unfolded neural network in time t.
The input at time t is x_t and the real oputput at time t is y_t.
W_h is the weight matrix that connects two consecutive hidden states h_{t-1} and h_t at time t.
W_x is the weight matrix that connects the input layer with the hidden layer and W_y is the weight matrix that connects the hidden layer with the output layer. An activation function is represented by σ and b_h is bias vector of hidden layer.
As the picture shows, the hidden layer is affected both by its input x_t and by the hidden state h_{t-1}.
The hidden state h_t value of RNN can be obtained by the following equation

$$h_t = \sigma(W_x x_t \dot{+} W_h h_{t-1} \dot{+} b_h)[1] \tag{1.1}$$

When training RNNs with long-term dependencies, problems may occure. The reason is, that propagating the gradient over too many layers tends to vanishingly small values, that is called vanishing gradient problem and happens most of the time, or tends to explode to extremely large values, that is called exploding gradient problem. If the gradient is too small, inital layer's weights are not be updated effectively and it can lead to the overall inaccuracy. When the huge error gradients accumulate exponentially from layer to layer, weights updates during training can not be done effectively and the whole network may become highly unstable.[14, 18]

To learn more about RNNs, the backpropagation through time and its computation, the following literature is recommended[14, 1, 15].

1.1.2 LSTM

The most effective sequence modelling networks, that can solve the exploding and vanishing gradient problems, are gated RNNs. First one to be introduces is the long short-term memory (LSTM).

LSTM RNNs have LSTM memory blocks, which has the same inputs and outputs as an basic RNN, memory cell, gating system that controls the whole flow of informations and some other parameters.
The memory block has its output recurrently connected back to its input and all of the gates.
LSTM cell contains weight internal reccurence (self-loop) which allows the gradient flow over a long duration. The memory cell is protected by the input gate, output gate and forget gate. These gates regulate the input and output flows and thus control the changes in the memory cell - specifically they transforms a cell state.
Data that is irrelevant may not be allowed to get through the input and output gates. Only relevant information can pass in and affect the cell or pass out and affect whole network.[2, 14, 1]
The original LSTM block did not have forget gates.[19]
Figure 1.1.2 shows the standard LSTM memory block structure.

LSTM networks has been found successful in speech recognition, image captioning, handwriting generation, handwriting recognition and many others.[2]

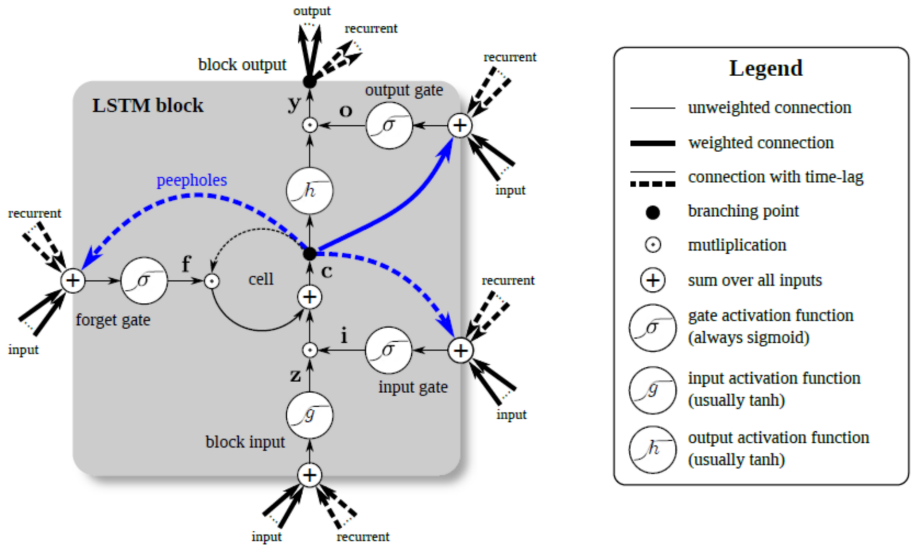


Figure 1.2: LSTM memory block. Picture adapted from [2].

Chapter 2

Medical imaging background

Neuroimaging is a term used for multiple imaging techniques serving to examine structure and function of brain. Further, in cognitive neuroscience they serve to examine the neurobiological correlates of human decision-making. These techniques are typically noninvasive, but some involve exposure to radiation or injection of radiolabeled substances.[20]

2.1 Positron Emission Tomography (PET)

Positron emission tomography (PET) involves injection of radioactive tracers into blood stream which can be then observed as emitted positrons are detected. Since the tracers are more processed in active parts of brain, these parts can be identified on PET and used to study brain function.

PET can also provide information on metabolic changes in brain. As an example, higher demands on glucose are typical for tumors where the usual metabolism is changed. And since injected radiolabeled glucose emits positrons, these tumors can be examined with PET. In addition to tumor evaluation, PET is also used for movement disorder diagnosis.[21, 22]

2.2 Electroencephalogram (EEG)

Electroencephalography (EEG) utilizes electrodes placed on head to measure electrical activity of neurons in brain. Variations in neuron activity are linked to brain wave changes. The product of the measurement is called electroencephalogram and shows changes in shape, magnitude and frequency of these waves.

Temporal resolution of detectable changes on EEG is in range of milliseconds, but spatial resolution is rather low compared to, MRI and PET.

The method is typically used for diagnostics of epilepsy, evaluation of seizures caused by other diseases and coma (Noggle and Davis 2021), but it is also used in research of neuropsychiatric disorders.[21, 20, 22]

2.3 Computed tomography (CT)

Computed tomography is an imaging method used to obtain structural information on brain. It is based on differences in x-rays absorption across different tissue types. Patient is irradiated with x-rays from multiple directions and the remaining x-rays travelling through the head are detected. This information is then used to reconstruct an image. Denser tissues such as bones absorb more x-rays, while water or air absorb the least.

CT is relatively cheap compared to other imaging methods and it is a common method of choice for acute brain trauma. CT is also considered as an non-invasive technique, but it needs to be considered that x-rays are applied.[21, 22]

2.4 Magnetic Resonance Imaging (MRI)

Magnetic resonance imaging is a non-invasive diagnostic technique using radio frequency radiation and magnetic field to obtain information on structure and function in patients body providing a possibility to detect various pathologies, injuries and illnesses. In order to be able to see a signal on MRI, a nucleus possessing a magnetic moment (such as proton 1H) is needed. In medicine, proton of water is the most commonly detected nucleus.[20, 23]

In presence of external magnetic field, spins of the active nuclei get aligned along the magnetic field. Once aligned, a radio-frequency pulse is applied causing the spins to flip. The resonant frequency depends on the nucleus and the strength of the magnetic field. Following the pulse, spins relax back to the equilibrium state while reemitting the radiation giving rise to signal.[23].

There are two types of relaxation - longitudinal (T1) relaxation and transverse (T2) relaxation. Longitudinal (spin-lattice) relaxation is a process in which spins populations returns back to thermal equilibrium state (which has been previously disturbed by radiofrequency pulse) by transferring energy to its surroundings. In process of transverse (spin-spin) relaxation the magnetization in xy plane decays.[24]

Differences in T1 and T2 relaxation times are employed in MRI to obtain contrast as different relaxation times leads to differences in signal intensities obtained at certain time point. For example, brain has a shorter T2 value

compared to cerebrospinal fluid allowing the contrast to appear between the two.[23]

In a constant magnetic field B0 of the MRI scanner, use of field gradients allows to spatially localize the signal as the field varies across the body. A gradient magnetic field provided by gradient coils is applied causing the nuclei to experience different fields at different locations. As the field differs, the resonant frequency of the nuclei differs. This allows us to select slices (slice select gradient) as well as location along the other two axis (frequency encoding and phase encoding gradients). In this way, 3 dimensions are obtained and localization is made possible.[23]

In addition to MRI, MR spectroscopy imaging (MRSI) can provide information on changes in biochemistry in different parts of brain that can be linked to disease, injury or other changes. In MRSI experiment, spectra for individual voxels are obtained and metabolic changes in various parts of brain can be visualized as an image (so-called metabolite map) to provide an information on local biochemical changes in brain.[25]

2.4.1 fMRI

In functional magnetic resonance imaging (fMRI) brain activity (functional changes) is measured employing changes in blood oxygenation, tissue perfusion and blood volume changes. These changes occur in response to neural activity and can therefore provide some information on brain functioning. Parts of the brain that get active can be identified in the process and represented in form of an activation map.
In fMRI additionally to 3 dimensions that were already introduced, the fourth dimension – time – plays a role, we talk about temporal resolution. [26]

Echo planar imaging (EPI) is a fast MRI acquisition method using multiple echoes in different phase steps. Different types of contrast can be obtained with EPI: T2* weighting in GE-EPI (gradient echo EPI), T2 weighting in SE-EPI (spin echo EPI) or T1 weighting in IR-EPI (inversion recovery EPI). Additionally, diffusion weighted EPI (DW EPI) can be used.[27]

When activated, neurons are supplied with increased blood flow to obtain more oxygen and glucose. This is called hemodynamic response. Oxyhemoglobin and deoxyhemoglobin have different magnetic susceptibilities. As hemodynamic response results in changes in levels of the two types of hemoglobin, it can be detected with MRI. This is a basis for blood oxygen level dependent (BOLD) contrast imaging.

2.4.2 task-based fMRI

The aim of task-based fMRI experiment is to determine parts of the brain involved in a certain task. For this purpose, patients are required to perform this task which leads to neural activation. This can be followed as changes in blood flow by MRI and brain structures linked to this specific task can be localized.

2.4.3 rs-fMRI

Resting state MRI (rs-fMRI) is a method based on spontaneous low frequency fluctuations in BOLD during resting state. In contrast to task-based fMRI, patients do not perform any tasks and remain resting during rsfMRI experiments as spontaneous neural activity is examined. The goal is to obtain more information on functional connectivity in brain in form of resting state networks (RSN).[28, 29, 30]

2.4.4 sMRI

Structural MRI (sMRI) is a tool for studying brain anatomy, structures of different tissue types and their alterations due to injuries, disease or pathologies such as tumor growth. sMRI is not directly linking the brain structures to neural activity but provides information on the parts of brain (and, for example, their volume) that are known to be associated with certain disease. The knowledge obtained from sMRI and fMRI can then be combined.[31]

Neuroimaging data are usually obtained in Nifti format. It consists of an image data and header part which contains information about measurement parameters. Prior to analysis, images are processed...

Chapter 3

Review of existing approaches

In this chapter, the current literature on statistical, machine learning and deep learning studies on ADHD diagnostic classification with identification of the various diagnostic techniques used, is researched. Focused on brain magnetic resonance imaging along with ADHD-200 dataset.

Due to the large number of existing studies and the need for their mutual comparability, only related works with the two-class classification were selected to differentiate between subjects with ADHD diagnosis and controls. The accuracy metric was selected for their comparacy though some of them also provide specificity and sensitivity measures. The list of existing studies is not exhaustive and includes state-of-the-art techniques in the area.

For ADHD diagnostic classification, *Zou et al.*[32] proposed joint multi-modality **3D-CNN** model with two-branches to combine s-MRI and fMRI data and learn from each one unique modality. The model was inspired by the way that radiologists examine brain images, learned latent 3D local patterns from individual 3D features (including ReHo, fALFF and VMHC, as well as the density of GM, WM and CSF in MNI space) and by hold-out validation obtained the state-of-the-art performance 69.2%.

In addition, *Riaz et al.*[33] proposed firstly **FCNet**, an CNN based network for calculating functional connectivity from fMRI signals obtaining overall 60.4% accuracy. Only Peking, NeuroImage and NYU sites data were used in this model and obtained corresponding accuracies 62.7%, 60% and 58.5%

Complementing this, *Riaz et al.*[34] proposed **DeepfMRI**, a novel end-to-end deep network for ADHD classification using feature extraction layer for functional connectivity features, similarity network and classification network with SVM. The network segmented brain into 90 ROIs using AAL atlas and computed similarity between regions features. Again, only Peking, NeuroImage and NYU sites data were used in this model and obtained corresponding accuracies 62.7%, 67.9% and 73.1%.

Moreover, *Mao et al.*[35] created **4D-CNN** network based on granular computing at a coarse level trained on derivative changes in entropy. Their work started with spatial information extraction of each fMRI image frame by 3D-CNN model and temporal information extraction by feature pooling and LSTM. Then final 4D-CNN model extracted both spatial and temporal information at the same time and obtained the accuracy 71.3%.

Finally, *Zhang et al.*[36] created **SC-CNN-Attention** network in order to overcome disadvantages of the conventional methods that depend on the single channel model and static computations and may cause the deficit of fMRI intrinsic information. This network fuse a separated channel convolutional neural network (SC-CNN) with an attention-based network to classify subjects with ADHD. Separated channels learn the temporal feature from all of the time series signals from one brain region (ROI) and then capture the temporal features among all the ROIs. There were used Peking, KKI, NYU, OHSU and NI sites data and obtained overall accuracy 68.6% evaluated with the "leave-one-site-out" CV (LOSOCV) for clinical relevance evaluation.

3.1 Training and testing data

Most of the studies retrieved MRI data from the Neuro Bureau ADHD-200 Preprocessed repository (ADHD-200) preprocessed by one of three available preprocessing data strategy. Pipelines mostly differ in the preprocessing toolset, algorithms, chosen parameters and computed statistical derivatives, but many of the studies did not mention which one had been chosen.[13] For this reason table no. XY does not include this information although may be mentioned, that the Athena pipeline was mentioned most often, followed by the NIAK pipeline, and lastly Burner pipeline.
The reviewed studies varied in used MRI data, some of them used along with MRI data an phenotypic information including patient's sex, age, IQ etc.

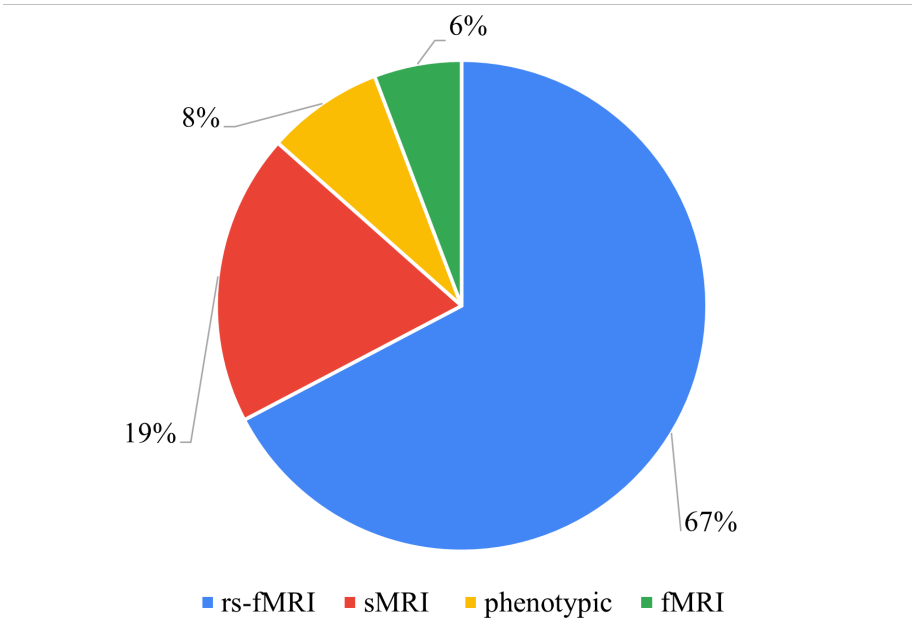


Figure 3.1: Data features of referenced literature

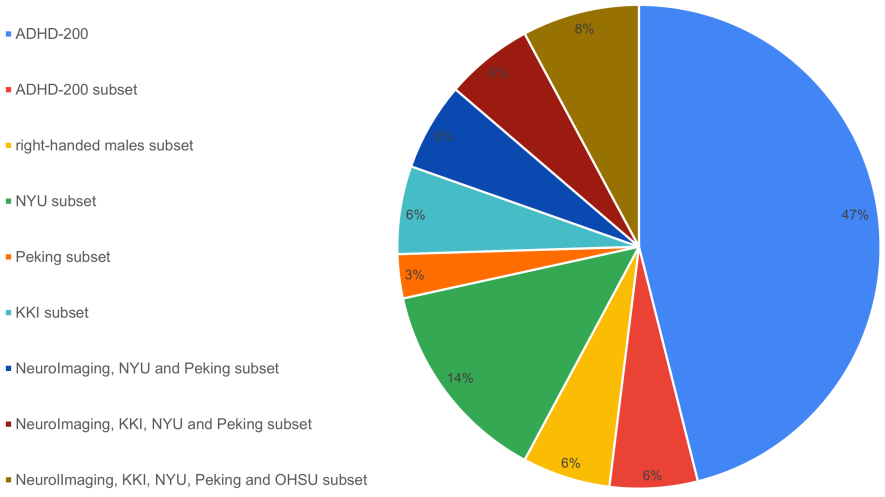


Figure 3.2: Dataset subset's of referenced literature

3.2 Validation test type

Various model evaluation approaches were found across the articles under review. The methods can be divided by hold-out and cross-validation approach. Hold-out validation divide the dataset into training, validation and testing data. Training data are used for training the model, validation data are used for tuning the model and test data for model's performance evaluation. K-fold cross validation differ in the number of folds that determine how many times algorithm iterates to confirm that all of the folds were used to train and test the model.

Many of the reviewed studies used leave-one-out CV (LOOCV), where the k number from K-fold CV is the number of all samples from original dataset. There is also another approach of cross validation that leave-one-site-out (LOSOCV) for testing while training on the rest of the data. This repeats for all of the dataset sites.

To get credible evaluation, training and testing datasets need to be independent. In diagnosis scenario this last LOSOCV approach, sometimes called as subject-wise, is recommended to be used for clinical relevance evaluation. In our case, universality from one clinical site to another is wanted, thus CV needs to be across these sites.[37]

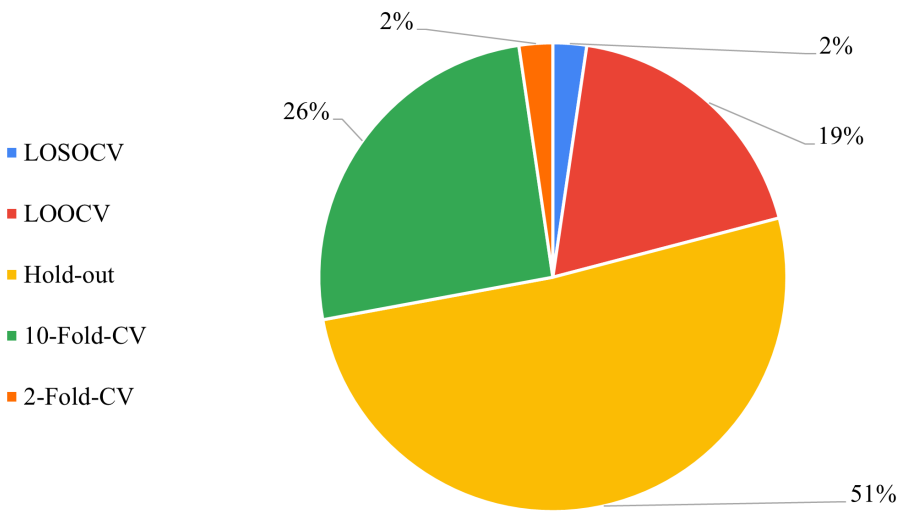


Figure 3.3: Validation type of reviewed literature

Chapter 4

Dataset description

For this thesis, ADHD-200 Sample dataset[1] was selected that was originally released for the ADHD-200 Global competition in 2011 and that is dedicated to accelerate researchers's understanding of the neurobiological basis of ADHD.

In order to make the competition accessible to wider range of scientists, perceiving that many of them lack the specialized knowledge of neuroimaging methods to implement the necessary steps of data preparation for rs-fMRI data and does not have access to computing resources with high performance, The Preprocessed Connectomes Project (PCP)[2] began openly sharing the results of the work on preprocessing the data at the Neuroimaging Informatics Tools and Resources Clearinghouse (NITRC)[3] .[13]

Many competitors used this preprocessed data, including the winning team and by reason of the aforementioned complications with the basic neuroimaging preprocessing the same data were used for this thesis.

4.1 ADHD-200 dataset

Dataset is publicly accessible and contains resting-state functional magnetic resonance imaging (rs-fMRI) and structural magnetic resonance imaging (sMRI) of the brain data along with phenotypic information for ADHD classification and biomarkers recognition.

Data was collected on 776 children, adolescents and young adults from eight independent imaging sites including 491 samples collected from typically developing individuals and 285 from children and adolescents diagnosed with

[1] http://fcon_1000.projects.nitrc.org/indi/adhd200/
[2] http://preprocessed-connectomes-project.org/
[3] https://www.nitrc.org/

ADHD, all anonymously. The phenotypic information includes the diagnosis, ADHD symptom measures, sex, age, intelligence quotient and their medication status. Not all samples includes all the phenotypic information. The preprocessed data were preprocessed by three different analytical pipelines. Pipelines mostly differ in the preprocessing toolset, algorithms, chosen parameters and computed statistical derivatives.[13]

The Athena[4] pipeline is using the combination of FSL[5] and AFNI[6] neuroimaging toolkits, NIAK[7] pipeline is based on the NeuroImaging Analysis Kit[8] on CBRAIN[9] software and the last pipeline called Burner[10] is based on the voxel-based-morphometry style analysis with SPM8[11] tool and uses only s-MRI data while both Athena and NIAK pipelines use rs-fMRI and s-MRI.

4.2 Data preprocessing and Feature Extraction

For this work, ADHD-200 dataset was used preprocessed with the Athena pipeline that uses custom BASH script to combine two neuroimaging toolkits - AFNI and FSL.

The Athena pipeline steps of the preprocessing for functional data are as follows:

- Removing first four volumes
- Slice timing correction where the separate slices of voxel time series are aligned to the middle one. This was done separately for each dataset site
- Replacing the transformation matrix in header with cardinal matrix
- Reorienting the dataset into RPI coordination
- Warping volumes for the motion correction
- Mask the data to get brain-only dataset

[4] https://www.nitrc.org/plugins/mwiki/index.php?title=neurobureau:AthenaPipeline

[5] https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FSL

[6] https://afni.nimh.nih.gov/

[7] https://www.nitrc.org/plugins/mwiki/index.php?title=neurobureau:NIAKPipeline

[8] https://www.nitrc.org/projects/niak/

[9] http://www.cbrain.ca/

[10] https://www.nitrc.org/plugins/mwiki/index.php?title=neurobureau:BurnerPipeline

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

- Compute the mean of each voxel time series and coregister mean image to corresponding s-MRI
- Write data into template space
- Use calculated masks from anatomical processing on white matter (WM) and cerebrospinal fluid (CSF) and extract its time-courses
- Use regression models on WM, CSF, motion corrected time series along with third-order polynomial
- Band-pass filter the time series so they include only frequencies implicated in resting state functional connectivity
- Use 6mm FWHM Gaussian filter to get blurred data

Output of this resting state functional data preprocessing includes preprocessed data as compressed 4D NIfTI files, mean image and mask for the data also as compressed NIfTI file and motion parameters as AFNI 1D file.

From aforementioned preprocessed data, subjects with the cleanest data were chosen for functionally defined CC200 and CC400 brain parcellation created by two-stage spatially-constrained clustering procedure and time courses were extracted by averaging the voxel time series within each labeled region. Another derivatives from preprocessed rs-fMRI data are brain parcellations and averaged regional time courses using the automated anatomical labeling (AAL), Eickhoff-Zilles (EZ), Harvard-Oxford (HO) and Talairach and Tournoux (TT) atlases. Also the regional homogenity (ReHo), functional connectivity (FC) and fractional amplitude of low-frequency fluctuations (fALFF) were derivated from preprocessed rs-fMRI for further processing.

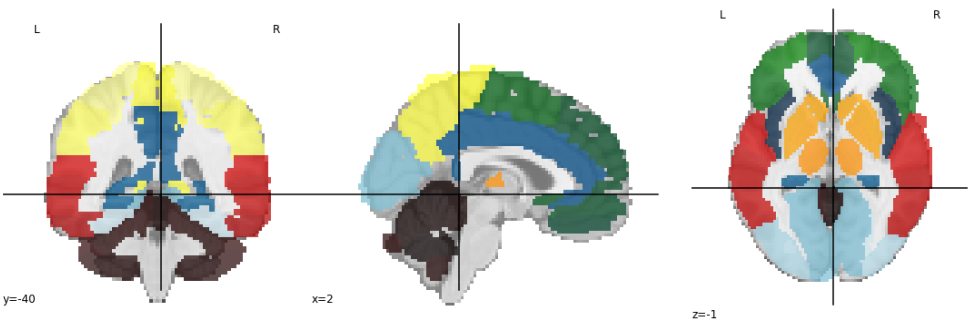


Figure 4.1: Automated anatomical labeling (AAL) brain atlas ROIs

The Athena pipeline steps of the preprocessing for structural data are as follows:

- Dataset reorentation into RPI coordination and brain extraction from surrounding tissue
- Segmentation into grey matter (GM), white matter (WM), and cerebrospinal fluid (CSF) probability maps
- WM and CSF masks creation
- Linear transformation applied on anatomical and template data followed by non-linear transformation to refine the result
- Write anatomical and GM probability data into template space
- Use 6mm FWHM Gaussian filter to get blurred GM data

Some subjects were excluded due to not containing the diagnostic label, anatomical data or having corrupted files.
In this work the emphasis is placed on anatomical parcellation of atlas assigned with AAL Toolbox. As the input serves rs-fMRI time series of all voxels accross each region warped into template space, each parcellation coregistrated and then resampled into the functional space using nearest-neighbor interpolation. Finally, the time courses were averaged within each region of interest (ROI) accross voxels signals. Each region is expressed by one local signal channel and is served as an input to the deep learning model.

- asi zminit, ze jsou napric sites velke rozdily, nektere files jinak rozlozene, merene na jinych pristrojich, nekdy otevrene a jindy zavrene oci apod - zminit nazvy sites a specificky ty, ktere byly vstupem pro muj model - rict, ze i mereni probihalo po rozdilnou dobu a ze jsem po uvazeni zvolila zarovnani na stejnou delku pro vsechny sites

Chapter 5

Implementation

In this chapter, the used technologies for the analysis, data processing and implementation are summarized. There is also a description of the designed network including a description of the individual parts. The model tuning process and hyperparameter optimization are also captured.

5.1 Technologies

For this work, Python and Keras[12] framework offering deep learning high-level API in Python were used. Keras is the most used deep learning framework built on top of Tensorflow 2.0[13]. I used the TensorFlow as a backend. TensorFlow is an open-source platform for developing and training machine learning models working with tensors.

For hyperparameter optimization was used KerasTuner[14]. KerasTuner is hyperparameter tuning library that helps to choose the optimal set of hyperparameters. It has bilt-in deep search algorithms like Hyperband, Bayesian optimization and Random search. I have choosen the Hyperband tuning algorithm, that uses adaptive resource allocation and has implemented early stopping to terminate a training optimization if a monitored metric has stopped improving for choosen number of epochs.
For visualization of the tuning process was used TensorBoard[15]. TensorBoard can visualize model graph and layers along with tracking and visualizing the training losses and metrics.

[12] https://keras.io/
[13] https://www.tensorflow.org/
[14] https://keras.io/keras_tuner/
[15] https://www.tensorflow.org/tensorboard

Google Colab[16] was chosen as a free Jupyter notebook[17], that runs in the cloud and provides access to CPU, TPU and GPU units. Since it has limitations on maximum RAM and running time without interruption, more model optimization were needed.

For image processing, loading and working with data in NIfTI (most common MRI format) data format, the NiBabel[18] library was used.
In addition, for brain volume data analysis and statistics was used neuroimaging Python library Nilearn[19], than helps manipulating with NiFti files and offers itself already ADHD-200 dataset subsample with 40 subjects. Some of the reviewed papers used this library and build deep learning model based on this small subset of data.

Data ballancing, data scaling and other computations like accuracy and precision were calculated using sklearn[20], open source machine learning library for the Python. Numerical calculations were done by NumPy[21] library and for plotting and visualization, Matplotlib[22] library was chosen.

5.2 Network input and architecture

The design of my own network for the diagnostic classification of ADHD was inspired mainly by the paper[36], where I was interested in separate ROI processing and abstract feature learning using separate channels to capture particularity of distinct fMRI ROI's signals.

The proposed network consists of four main parts. The first part contains a separate CNN for each ROI. The second part of the network consists of stacked GRU layers. This is followed by the self-attention block. The last part consists of the classifier.

5.2.1 Separated CNN for each ROI

Region's signals were extracted from the whole brain rs-fMRI BOLD time courses accross all of its voxels within the same region of interest using automated anatomical labeling (AAL) atlas. After computing averaging within ROI voxel time courses for each region separately were obtained 116 distinct time series signals. Finally, each of this 116 time series for each subject pro-

[16] https://colab.research.google.com/
[17] https://jupyter.org/
[18] https://nipy.org/nibabel/
[19] https://nilearn.github.io
[20] https://scikit-learn.org/
[21] https://numpy.org/
[22] https://matplotlib.org/

vided the input for single channel of this proposed network.

The length of the extracted rs-fMRI time series varied among all input dataset sites. In order to deal with this differencies the extracted ROIs were aligned to the same length just before entering the convolution.

For the reason that the data comes from the different sites and are measured at different scales by different devices, they do not contribute equally to the fit and the learning function of the model. To avoid the possible bias, the standardization was applied.

To standardize a dataset with multiple features and time-series data, before applying machine learning model, the data was firstly reshaped in a format so that each feature could be normalized separately. Given the distribution of the data, each feature (ROI) in the dataset had the mean value subtracted, and then divided by the standard deviation of the individual feature part of the dataset. Subsequently, the data was returned in the same shape as before.

Each CNN can be viewed as a separate channel to capture an discriminative abstract representation of a given ROI signal. Each channel is processed with stacked convolution and pooling layers followed by Drop and Batch normalization layer. Neurons of the hidden layers learn the representations over the input time series. It is assumed, that all learned representations are independent of one another.

Dropout layer serves as a regularization to prevent overfitting in deep learning models by randomly changing the network architecture. We want to avoid that the learned values of the weights are too fitted to the baseline training data and thus not generalized well enough for the test data and dropout is one of the highly used techniques.[38]

As an activation function for convolution layers was chosen the Leaky Rectified Linear Unit (Leaky ReLU), which is based on ReLU with small slope for negative values that helps in dying units problem prevention. Leaky ReLU can be computed by the following equation[39]

$$LeakyReLU(x) = max(\beta * x, x) \tag{5.1}$$

where β is a small constant.

First convolution layer has number of filters equal to the number of input ROIs timesteps and kernel size of 5. Then is applied max pooling layer for down-sampling 1D temporal data and batch normalization to reduce generalization error and accelerate the training.

Second and last convolution layer has the number of filters two times higher then the previous one and has kernel size of 3.
Thereafter, the representation is downsampled by taking the maximum value over the time dimension with global max pooling layer.

All these CNN sub-networks work as feature extractor are followed by concatenation layer that takes as input a list of tensors of the same shape, except for the concatenation axis, and returns a single tensor that is the concatenation of all extracted features. This tensor serves as an input for the following LSTM Layer.

5.2.2 LSTM

Here, the multiple ROIs local features, that carry information of different brain regions, are integrated and the information about the interconnection of these regions can be extracted.

LSTM Layer set dropout fraction of the units for the linear transformation of the inputs and recurrent dropout fraction of the units for the linear transformation of the recurrent state, which serves as a regularization method for RNNs.

As an activation function is used hyperbolic tangent. Number of features is chosen as a number of units. Number of units is equal to the number of features (ROIs). After layer normalization follows second LSTM layer with less number of units.

5.2.3 Self-Attention Layer

Attention mechanism was chosen for its ability to learn which part of the data is more crucial than the other. In 2017 was introduced as a new form of attention mechanism called self-attention.[40]

Self-attention, sometimes called intra-attention, is an attention mechanism with ability to recognize the context of each data sequence time step.
This mechanism takes into account different positions in the sequence to compute its representation and give larger weights to the more important features, which is greatly handful for this task.[39]

Self-attention was computed by the following equations

$$h_{(t,t)} = \tanh(x_t^T W_t \dot{+} x_t^T, W_x \dot{+} b_t) \tag{5.2}$$

$$e_{(t,t\prime)} = \sigma(W_a h_{(t,t\prime)} \dot{+} b_a) \tag{5.3}$$

$$a_t = softmax(e_t) \tag{5.4}$$

$$l_t = \sum_{t\prime} \alpha_{(t,t\prime)} x_{t\prime} \tag{5.5}$$

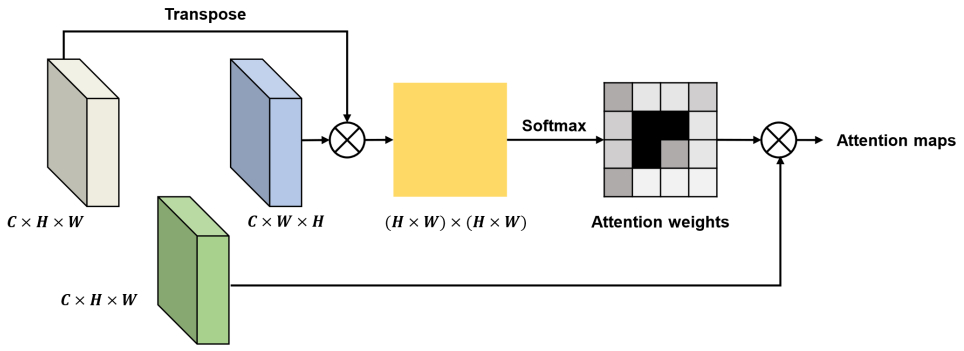
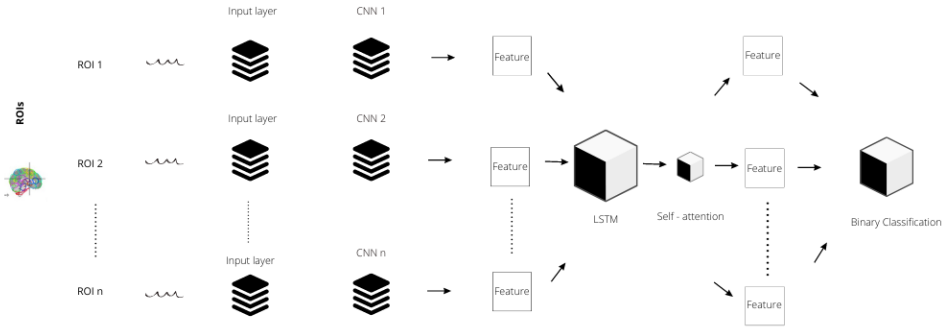


Figure 5.1: Example of self-attention block structure. Picture adapted from [3]

5.2.4 Classifier

New extracted features serve as an input for the last layer. For the output layer, sigmoid activation was used as it is suitable for binary classification problems. As loss function is used weighted binary cross-entropy to predict the binary class.
Weighted binary cross-entropy considers that samples are imbalanced.



5.3 Model tunning and hyperparameter optimization

During the training and validation phases, Keras Tuner[41] was used to optimize some hyperparameters. The hyperparameter search was done using the Hyperband algorithm[42] implemented in Keras Tuner. Hyperband tuner was designed as an optimized version of random search tuner which uses early-stopping to accelerate the hyperparameter tuning process.

Searched hyperparameters include dropout percentages for each layer separately, number of hidden layers, number of layer neurones (specifically in GRU corresponds to the amount of information remembered between time steps), learning rate Both LSTM and GRU were tried.

In final dense layer was used sigmoid activation function that calculates loss for the binary cross-entropy which gives out binary output.

The learning rate (lrate) is a floating-point value between 1e-4 and 1e-1, chosen logarithmically (not linearly). The L2 regularization value is chosen from a set of five predefined values (0.0, 1e-1, 1e-2, 1e-3, and 1e-4).

The number of hidden nodes (num_hidden) is an integer chosen from the range 32 to 256 in increments of 32. These values are then used in the model-building code as normal.

5.3.1 Training Optimization Algorithm

As the optimization algorithm was chosen one of the stochastic gradient descent methods, adaptive moment estimation (Adam). The Adam optimization algorithm was first introduced in the paper Adam: A Method for Stochastic Optimization[43] by Diederik P. Kingma and Jimmy Ba. Algorithm is based on adaptive estimation of first-order moment, is computationally efficient and convenient for cases when we have a lot of data or parameters, that is my case. Additionaly, it requires less memory and is well suited for noisy and spare gradients for non-stationary objectives and problems.[43]

To optimize the model the original dataset was splitted into small batches as it may offer a regularizing effect due to the noise that minibatches add to the learning process. This technique is called minibatch stochastic method and for the model training were used four sizes of batches 8, 16, 32, 64. As size 8 was too small and the runtime was very high and size 64 was too high, with less than linear returns, I finally decided to use size 32 that leaded to the best performance.[14]

5.3.2 Class Balancing

In the ADHD-200 dataset are disproportionately more healthy controls (HC) than patients diagnosed with ADHD. This disproportion also varies across data of different sites. This may bias the model in favour of one category over the other.
Many articles used imbalanced datasets or imbalanced dataset's subsets and only a few of them took balancing of classes into account.

To increase the clinical plausibility of the results I also opted for balancing of classes by one of the techniques. Since this is a binary classification I opted for weight balancing of classes to give equal importance for both classes on gradient updates and directly specified the weights for each of the target classes. Each sample contributes to the loss proportionally to its class weight, underrepresented class loss is multiplied with frequency of overrepresented class and vice versa.

Chapter 6

Results

The accuracy metric was used to measure the performance of the classification model for ADHD diagnosis. For its evaluation was used leave-one-site-out cross-validation, i.e. using each single data site as a test set while training on all the others and then comparing the model performance individually. The attached graph compares the results of my model S-CNN-LSTM-Attention model with the results based on the ADHD-200 Worldwide Competition. It is important to note that not every model used all data sites from the competition.

I used datasets from the NYU, Peking, CCI and NI sites. The results of my model are comparable to those with which they were compared. Thus, the model meets the claim of sufficient accuracy. Evaluation through leave-one-site-out cross-validation confirms the performance of the implemented model.

	NYU	Peking	KKI	NI	Average accuracy
Dey et al. (2014)	None	58.82	54.5	48	60.93
FCNet	58.5	62.7	None	60.0	60.4
DeepFMRI (2020)	73.1	62.7	None	67.9	67.9
3D-CNN	None	62.9	72.8	None	67.8
SC-CNN-Attention	60.4	65.2	77.7	75.3	68.6
S-CNN-LSTM-Attention	53.2	74.58	74.7	72.92	68.85

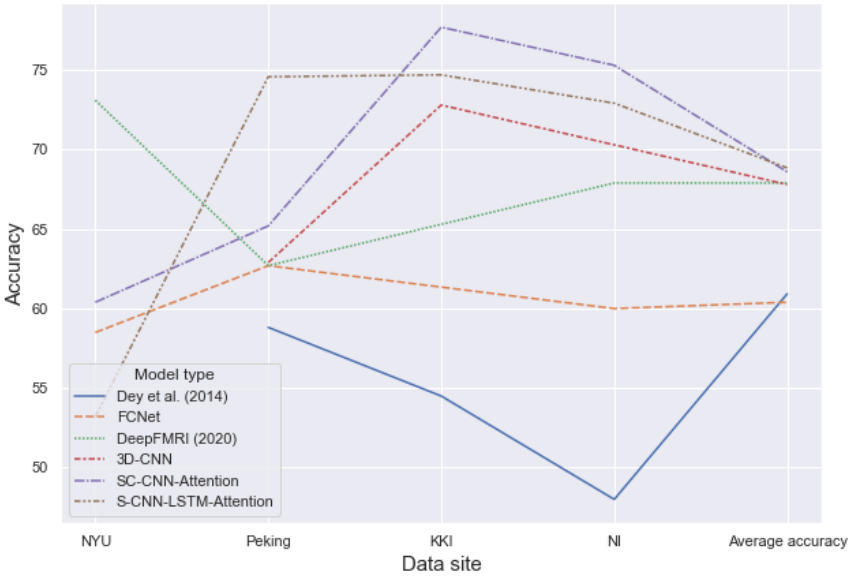


Figure 6.1: Comparison of the accuracies of selected models

Conclusion

In her thesis entitled Imaging-Based Diagnostic Classification of ADHD, the author aimed to investigate the existing techniques for the diagnosis of ADHD with special reference to functional magnetic resonance imaging. The author's interest in the topic stemmed from the fact that ADHD is one of the most common neurodevelopment disorders in children and adolescents. Yet, in many European countries, it is still a disorder whose diagnosis is inadequate. ADHD, which persists into adulthood, has a number of psychological complications and is often associated with anxiety and depressive disorders and learning disabilities. Proper diagnosis of the disorder thus becomes crucial in preventing and dealing with the complexities associated with ADHD.

The author of this thesis assumes that the current diagnostic practice of ADHD depends mainly on the clinical evaluation of behavioral symptoms with respect to the responses of patients and concerned persons. Meanwhile, the technique of ADHD diagnosis varies based on the methods used (clinical interviews, observations, etc.) and the different sources (teachers, parents and other persons in contact with the person with the disorder). This variation in diagnostic techniques can lead to misdiagnosis. For these reasons, the author discussed in her thesis the existing classification models of ADHD that can be considered objective.

In the thesis, the author explains the method of examining the issue of ADHD, its diagnosis, the description of her own classification model and the results of the comparison. In the first chapter, the author explains the basics of recurrent neural network as she assumes the reader's knowledge of machine learning. In the following chapter, the author describes the basic methods used for imaging the human brain. She then moves on to summarize the current state of knowledge in classification techniques for ADHD diagnosis, in which the dataset from the ADHD-200 Worldwide Competition was used.

In the fourth chapter, devoted to data description, the author presents the dataset from the aforementioned competition and explains the preprocessing and data extraction methods used. Chapter five describes the design and implementation of the author's model for ADHD diagnostic classification. Chapter six provides a clear and concise explanation of the results of the author's model and a comparison with the results of current existing ADHD diagnosis techniques.

Based on the knowledge gained, the author of the thesis proposed her own model for ADHD classification diagnosis through the use of functional magnetic resonance imaging and deep neural networks. She then compared the results of her model with classification models using data from the ADHD-200 Worldwide Competition. The author concludes that her results in her thesis are comparable to those of other classification models.

Bibliography

[1] Fang, X.; Xu, M.; Xu, S.; aj.: A deep learning framework for predicting cyber attacks rates. *EURASIP Journal on Information Security*, ročník 2019, 05 2019, doi:10.1186/s13635-019-0090-6.

[2] Greff, K.; Srivastava, R. K.; Koutník, J.; aj.: LSTM: A Search Space Odyssey. *CoRR*, ročník abs/1503.04069, 2015, 1503.04069. Dostupné z: http://arxiv.org/abs/1503.04069

[3] Sun, C.; Ai, Y.; Wang, S.; aj.: A Two-Branch Network for Weakly Supervised Object Localization. *Electronics*, ročník 9, 06 2020: str. 955, doi:10.3390/electronics9060955.

[4] Faraone, S. V.; Banaschewski, T.; Coghill, D.; aj.: The World Federation of ADHD International Consensus Statement: 208 Evidence-based conclusions about the disorder. *Neuroscience & Biobehavioral Reviews*, ročník 128, 2021: s. 789–818, ISSN 0149-7634, doi: https://doi.org/10.1016/j.neubiorev.2021.01.022. Dostupné z: https://www.sciencedirect.com/science/article/pii/S014976342100049X

[5] ABLE, S. L.; JOHNSTON, J. A.; ADLER, L. A.; aj.: Functional and psychosocial impairment in adults with undiagnosed ADHD. *Psychological Medicine*, ročník 37, č. 1, 2007: str. 97–107, doi:10.1017/S0033291706008713.

[6] Katzman, M.; Bilkey, T.; Chokka, P.; aj.: Adult ADHD and comorbid disorders: Clinical implications of a dimensional approach. *BMC Psychiatry*, ročník 17, 08 2017: str. 302, doi:10.1186/s12888-017-1463-3.

[7] Willcutt, E. G.: The Prevalence of DSM-IV Attention-Deficit/Hyperactivity Disorder: A Meta-Analytic Review. *Neurotherapeutics*, ročník 9, 2012: s. 490–499.

[8] Wolraich, M. L.; Hagan, J., Joseph F.; Allan, C.; aj.: Clinical Practice Guideline for the Diagnosis, Evaluation, and Treatment of Attention-Deficit/Hyperactivity Disorder in Children and Adolescents. *Pediatrics*, ročník 144, č. 4, 10 2019, ISSN 0031-4005, doi:10.1542/peds.2019-2528, e20192528, https://publications.aap.org/pediatrics/article-pdf/144/4/e20192528/1078222/peds_20192528.pdf. Dostupné z: https://doi.org/10.1542/peds.2019-2528

[9] Emser, T.; Johnston, B.; Steele, J.; aj.: Assessing ADHD symptoms in children and adults: Evaluating the role of objective measures. *Behavioral and Brain Functions*, ročník 14, 12 2018, doi:10.1186/s12993-018-0143-x.

[10] Thome, J.; Ehlis, A.-C.; Fallgatter, A.; aj.: Biomarkers for attention-deficit/hyperactivity disorder (ADHD). A consensus report of the WFSBP task force on biological markers and the World Federation of ADHD. *The world journal of biological psychiatry : the official journal of the World Federation of Societies of Biological Psychiatry*, ročník 13, 07 2012: s. 379–400, doi:10.3109/15622975.2012.690535.

[11] Consortium, T.: The ADHD-200 Consortium: A Model to Advance the Translational Potential of Neuroimaging in Clinical Neuroscience. *Frontiers in Systems Neuroscience*, ročník 6, 09 2012, doi:10.3389/fnsys.2012.00062.

[12] Brown, M.; Sidhu, G.; Greiner, R.; aj.: ADHD-200 Global Competition: Diagnosing ADHD using personal characteristic data can outperform resting state fMRI measurements. *Frontiers in systems neuroscience*, ročník 6, 09 2012: str. 69, doi:10.3389/fnsys.2012.00069.

[13] Bellec, P.; Chu, C.; Chouinard-Decorte, F.; aj.: The Neuro Bureau ADHD-200 Preprocessed repository. *NeuroImage*, ročník 144, 2017: s. 275–286, ISSN 1053-8119, doi:https://doi.org/10.1016/j.neuroimage.2016.06.034, data Sharing Part II. Dostupné z: https://www.sciencedirect.com/science/article/pii/S105381191630283X

[14] Goodfellow, I.; Bengio, Y.; Courville, A.: *Deep Learning*. MIT Press, 2016, http://www.deeplearningbook.org.

[15] Graves, A.: Sequence Transduction with Recurrent Neural Networks. *CoRR*, ročník abs/1211.3711, 2012, 1211.3711. Dostupné z: http://arxiv.org/abs/1211.3711

[16] Lipton, Z. C.: A Critical Review of Recurrent Neural Networks for Sequence Learning. *CoRR*, ročník abs/1506.00019, 2015, 1506.00019. Dostupné z: http://arxiv.org/abs/1506.00019

[17] Schmidt, R. M.: Recurrent Neural Networks (RNNs): A gentle Introduction and Overview. 2019, doi:10.48550/ARXIV.1912.05911. Dostupné z: https://arxiv.org/abs/1912.05911

[18] Philipp, G.; Song, D.; Carbonell, J. G.: Gradients explode - Deep Networks are shallow - ResNet explained. *CoRR*, ročník abs/1712.05577, 2017, 1712.05577. Dostupné z: http://arxiv.org/abs/1712.05577

[19] Hochreiter, S.; Schmidhuber, J.: Long Short-Term Memory. *Neural Computation*, ročník 9, č. 8, 11 1997: s. 1735–1780, ISSN 0899-7667, doi:10.1162/neco.1997.9.8.1735, https://direct.mit.edu/neco/article-pdf/9/8/1735/813796/neco.1997.9.8.1735.pdf. Dostupné z: https://doi.org/10.1162/neco.1997.9.8.1735

[20] Xue, G.; Chen, C.; Lu, Z.-L.; aj.: Brain Imaging Techniques and Their Applications in Decision-Making Research. *Xin li xue bao. Acta psychologica Sinica*, ročník 42, 02 2010: s. 120–137, doi:10.3724/SP.J.1041.2010.00120.

[21] Noggle, C. A.; Davis, A. S.: *Advances in Neuroimaging*. Cham: Springer International Publishing, 2021, ISBN 978-3-030-59162-5, s. 107–137, doi:10.1007/978-3-030-59162-5_5. Dostupné z: https://doi.org/10.1007/978-3-030-59162-5_5

[22] Mier, W.; Mier, D.: Advantages in functional imaging of the brain. *Frontiers in human neuroscience*, ročník 9, 05 2015: str. 249, doi:10.3389/fnhum.2015.00249.

[23] Currie, S.; Hoggard, N.; Craven, I. J.; aj.: Understanding MRI: basic MR physics for physicians. *Postgraduate Medical Journal*, ročník 89, č. 1050, 2013: s. 209–223, ISSN 0032-5473, doi:10.1136/postgradmedj-2012-131342, https://pmj.bmj.com/content/89/1050/209.full.pdf. Dostupné z: https://pmj.bmj.com/content/89/1050/209

[24] Tekely, P.: Malcolm H. Levitt. Spin dynamics: basics of nuclear magnetic resonance. John Wiley & Sons, Chichester, UK, 2001, 686 pp. Price £34.95. ISBN 0-471-48921-2. *Magnetic Resonance in Chemistry*, ročník 40, č. 12, 2002: s. 800–800, doi:https://doi.org/10.1002/mrc.1092, https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/pdf/10.1002/mrc.1092. Dostupné z: https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/abs/10.1002/mrc.1092

[25] Nassirpour, S.; Chang, P.; Henning, A.: High and ultra-high resolution metabolite mapping of the human brain using 1H FID MRSI

at 9.4T. *NeuroImage*, ročník 168, 2018: s. 211–221, ISSN 1053-8119, doi:https://doi.org/10.1016/j.neuroimage.2016.12.065, neuroimaging with Ultra-high Field MRI: Present and Future. Dostupné z: https://www.sciencedirect.com/science/article/pii/S1053811916307923

[26] Logothetis, N. K.: What we can do and what we cannot do with fMRI. *Nature*, ročník 453, 2008: s. 869–878.

[27] Ryu, K.; Baek, H.; Skare, S.; aj.: Clinical Experience of 1-Minute Brain MRI Using a Multicontrast EPI Sequence in a Different Scan Environment. *AJNR. American journal of neuroradiology*, ročník 41, č. 3, March 2020: str. 424—429, ISSN 0195-6108, doi:10.3174/ajnr.a6427. Dostupné z: https://europepmc.org/articles/PMC7077884

[28] Biswal, B.; Zerrin Yetkin, F.; Haughton, V. M.; aj.: Functional connectivity in the motor cortex of resting human brain using echo-planar mri. *Magnetic Resonance in Medicine*, ročník 34, č. 4, 1995: s. 537–541, doi:https://doi.org/10.1002/mrm.1910340409, https://onlinelibrary.wiley.com/doi/pdf/10.1002/mrm.1910340409. Dostupné z: https://onlinelibrary.wiley.com/doi/abs/10.1002/mrm.1910340409

[29] Lv, H.; Wang, Z.; Tong, E.; aj.: Resting-State Functional MRI: Everything That Nonexperts Have Always Wanted to Know. *American Journal of Neuroradiology*, ročník 39, 01 2018, doi:10.3174/ajnr.A5527.

[30] Lv, H.; Wang, Z.; Tong, E.; aj.: Resting-State Functional MRI: Everything That Nonexperts Have Always Wanted to Know. *American Journal of Neuroradiology*, ročník 39, č. 8, 2018: s. 1390–1399, ISSN 0195-6108, doi:10.3174/ajnr.A5527, http://www.ajnr.org/content/39/8/1390.full.pdf. Dostupné z: http://www.ajnr.org/content/39/8/1390

[31] Helms, G.: Segmentation of human brain using structural MRI. *Magma (New York, N.Y.)*, ročník 29, 01 2016, doi:10.1007/s10334-015-0518-z.

[32] Zou, L.; Zheng, J.; Miao, C.; aj.: 3D CNN Based Automatic Diagnosis of Attention Deficit Hyperactivity Disorder Using Functional and Structural MRI. *IEEE Access*, ročník 5, 2017: s. 23626–23636, doi:10.1109/ACCESS.2017.2762703.

[33] Riaz, A.; Asad, M.; al Arif, S. M. M. R.; aj.: FCNet: A Convolutional Neural Network for Calculating Functional Connectivity from Functional MRI. 09 2017: s. 70–78, doi:10.1007/978-3-319-67159-8_9.

[34] Riaz, A.; Asad, M.; Al-Arif, S. M. M. R.; aj.: Deep fMRI: AN end-to-end deep network for classification of fMRI data. *2018 IEEE 15th*

International Symposium on Biomedical Imaging (ISBI 2018), 2018: s. 1419–1422.

[35] Mao, Z.; Su, Y.; Xu, G.; aj.: Spatio-temporal deep learning method for ADHD fMRI classification. *Information Sciences*, ročník 499, 2019: s. 1–11, ISSN 0020-0255, doi:https://doi.org/10.1016/j.ins.2019.05.043. Dostupné z: https://www.sciencedirect.com/science/article/pii/S0020025519304475

[36] Zhang, T.; Li, C.; Li, P.; aj.: Separated Channel Attention Convolutional Neural Network (SC-CNN-Attention) to Identify ADHD in Multi-Site Rs-fMRI Dataset. *Entropy*, ročník 22, č. 8, 2020, ISSN 1099-4300, doi:10.3390/e22080893. Dostupné z: https://www.mdpi.com/1099-4300/22/8/893

[37] S, S.; L, L.; A, J.; aj.: The need to approximate the use-case in clinical machine learning. *GigaScience*, ročník 6, 05 2017: s. 1–9, doi:10.1093/gigascience/gix019.

[38] Srivastava, N.; Hinton, G.; Krizhevsky, A.; aj.: Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, ročník 15, 06 2014: s. 1929–1958.

[39] Xu, B.; Wang, N.; Chen, T.; aj.: Empirical Evaluation of Rectified Activations in Convolutional Network. *CoRR*, ročník abs/1505.00853, 2015, 1505.00853. Dostupné z: http://arxiv.org/abs/1505.00853

[40] Vaswani, A.; Shazeer, N.; Parmar, N.; aj.: Attention Is All You Need. 2017, doi:10.48550/ARXIV.1706.03762. Dostupné z: https://arxiv.org/abs/1706.03762

[41] O'Malley, T.; Bursztein, E.; Long, J.; aj.: Keras Tuner. https://github.com/keras-team/keras-tuner, 2019.

[42] Li, L.; Jamieson, K.; DeSalvo, G.; aj.: Hyperband: A Novel Bandit-Based Approach to Hyperparameter Optimization. 2016, doi:10.48550/ARXIV.1603.06560. Dostupné z: https://arxiv.org/abs/1603.06560

[43] Kingma, D.; Ba, J.: Adam: A Method for Stochastic Optimization. *International Conference on Learning Representations*, 12 2014.

Appendix A

Contents of CD

- `Readme.txt` ..... the file with CD contents description
- `Notebooks` ..... the directory of source codes
- `thesis` ..... the LaTeX source code files of the thesis
  - `thesis.pdf` ..... the Diploma thesis in PDF format