

EGILS AVOTS

Brain abnormality detection
using statistical analysis of
individual structural connectivity
networks and EEG signals



EGILS AVOTS

Brain abnormality detection using statistical
analysis of individual structural connectivity
networks and EEG signals



UNIVERSITY OF TARTU

Press

Institute of Technology, Faculty of Science and Technology, University of Tartu, Estonia.

Dissertation has been accepted for the commencement of the degree of Doctor of Philosophy (PhD) in Engineering and Technology on 16.11.2023 by the Council of the Institute of Technology, University of Tartu.

Supervisors

Prof. Gholamreza Anbarjafari
University of Tartu
Tartu, Estonia

Assoc. Prof. Maie Bachmann
Tallinn University of Technology
Tallinn, Estonia

Reviewer

Assoc. Prof. Naveed Muhammad
University of Tartu
Tartu, Estonia

Opponent

Prof. Ausra Saudargiene
Lithuanian University of Health Sciences
Kaunas, Lithuania

The public defense will take place on 15.12.2023 at 12:15 in Nooruse 1-121.

The publication of this dissertation was financed by the Institute of Technology, University of Tartu.

Copyright © 2023 by Egils Avots

ISSN 2228-0855 (print)

ISBN 978-9916-27-421-7 (print)

ISSN 2806-2620 (pdf)

ISBN 978-9916-27-421-7 (pdf)

University of Tartu Press

<http://www.tyk.ee/>

To my family and friends

ABSTRACT

Brain abnormalities are characterised by anatomical or functional changes in the brain that can be present at birth or later emerge as a result of trauma, disease, or other circumstances. These variations may significantly affect a person's physical and mental health, resulting in a variety of symptoms and health issues. A few examples of brain abnormalities are variations in the size and structure of certain brain regions and variations in neural connectivity or activity. Magnetic resonance imaging, computed tomography, and methods that assess the electrical activity of the brain, such as electroencephalography, have all significantly improved the detection and characterization of brain disorders such as Alzheimer's disease and clinical depression. The detection of the two previously mentioned brain disorders is the focus of this thesis, where Alzheimer's disease is classified using magnetic resonance imaging and depression is classified using electroencephalograms.

Alzheimer's disease, a degenerative neurological condition, impacts millions of individuals worldwide. The diagnosis of Alzheimer's requires a thorough evaluation that includes cognitive tests, brain scans, and other clinical exams, which can make it difficult. With the use of cutting-edge technologies, machine learning algorithms have emerged as a viable tool to assist in the diagnosis of Alzheimer's disease. Machine learning algorithms can be used to spot patterns in brain scans that are specific to Alzheimer's disease, like changed patterns of brain activity or changes in the size and shape of specific brain regions. In this study, such changes were encoded in histograms, histogram-based local descriptors, and features extracted using pre-trained neural networks. The magnetic resonance images were obtained from the Alzheimer's Disease Neuroimaging Initiative dataset, and the trained models distinguished between cognitively normal and Alzheimer's disease patients.

A mental health condition such as clinical depression is normally diagnosed following a thorough assessment by a mental health professional; however, using machine learning techniques to assist the diagnosis procedure has shown potential. Data about depression, including patient symptoms and demographic data, can be used to train machine learning algorithms to spot trends and forecast the occurrence of depression. Moreover, depression-related brain activity can be identified using electroencephalography, and machine learning algorithms can be trained to interpret these patterns to help with diagnosis. This study shows how different linear and non-linear features extracted from electroencephalography recordings, provided by Tallinn University of Technology, were classified with an ensemble setup and used to classify patients who have had clinical depression.

Although the application of machine learning to the diagnosis of Alzheimer's and clinical depression in medical practice is still in its infancy, it has the potential to improve diagnostic procedures.

CONTENTS

List of Original Publications	17
Preface	20
1. Introduction	21
1.1. Alzheimer’s Disease	23
1.2. Clinical Depression	25
2. Related Work	26
2.1. Alzheimer’s Disease and Magnetic Resonance Imaging	26
2.2. Clinical Depression and Electroencephalography	29
3. Brain Structural and Functional Imaging	30
3.1. Magnetic Resonance Imaging	30
3.1.1. Nuclear Magnetic Resonance	30
3.1.2. Magnetic Resonance Imaging Device	33
3.2. Electroencephalography	37
3.2.1. Electroencephalogram Registration and Analysis	38
3.2.2. Human Electroencephalogram Rhythms	41
4. Machine Learning	43
4.1. Machine Learning Types	43
4.2. Feature Extraction Algorithms	44
4.2.1. Principal Component Analysis	44
4.2.2. Singular Value Decomposition	45
4.2.3. Linear Discriminant Analysis	46
4.3. Conventional Machine Learning Algorithms	47
4.3.1. Logistic Regression	47
4.3.2. Naive Bayes	47
4.3.3. K-nearest Neighbors	47
4.3.4. Decision Tree	48
4.3.5. Support Vector Machine	48
4.3.6. Classifier Ensemble	48
4.4. Deep Learning	49
4.4.1. Neurons and Activation Functions	50
4.4.2. Feedforward Neural Networks	51
4.4.3. Backpropagation Algorithm	52
4.4.4. Gradient Descent Optimization	54
4.4.5. Single Layer Perceptron	56
4.4.6. Multi Layer Perceptron	57
4.4.7. Convolutional Neural Network	57
4.4.8. Transfer Learning	61

5. Datasets	62
5.1. Alzheimer’s Disease Neuroimaging Initiative MRI Dataset	62
5.1.1. Magnetic Resonance Imaging Correction Methods	63
5.2. Tallinn University of Technology EEG Dataset	64
5.2.1. Electroencephalogram Recording Procedure	64
5.2.2. Hamilton Depression Rating Scale	64
5.2.3. Emotional State Questionnaire	64
5.2.4. Participant Selection	65
6. Volumetric Histogram Based Alzheimer’s Disease Detection	66
6.1. Introduction	66
6.2. Methodology	66
6.2.1. Dimensionality Reduction	66
6.2.2. MRI Slice Histograms	67
6.3. Results and Discussions	68
7. Alzheimer’s Classification Using Image Texture Descriptors	72
7.1. Introduction	72
7.2. Features	72
7.2.1. Histogram-based Local Descriptors	73
7.2.2. Feature Extraction Using Pre-trained Convolutional Neural Networks	74
7.3. Methodology	75
7.4. Results and Discussion	76
7.4.1. Axial Plane	76
7.4.2. Coronal Plane	81
7.4.3. Sagittal Plane	86
7.4.4. Multi-slice View	91
8. Ensemble Approach for Detection of Depression Using EEG Features	95
8.1. Introduction	95
8.2. Features	95
8.2.1. Linear Features	96
8.2.2. Non-linear Features	96
8.3. Methodology	97
8.3.1. Feature Selection	99
8.3.2. Machine Learning Algorithms	99
8.4. Results and Discussion	99
9. Conclusion	103
Bibliography	104
Acknowledgements	117

Sisukokkuvõte (Summary in Estonian)	118
Publications	119
Ensemble Approach for Detection of Depression Using EEG Features . . .	121
Volumetric Histogram Based Alzheimer’s Disease Detection Using Support Vector Machine	133
Audiovisual Emotion Recognition in Wild	159
Privacy-Constrained Biometric System for Non-Cooperative Users . . .	173
Automatic Access Control Based on Face and Hand Biometrics in A Non-Cooperative Context	191
A New Kernel Development Algorithm for Edge Detection Using Singular Value Ratios	202
Curriculum Vitae	218
Elulookirjeldus (Curriculum Vitae in Estonian)	219

LIST OF FIGURES

1. Illustration of brain with and without AD	23
2. Free induction decay	32
3. Image (a) Schematic illustration of the MRI system and (b) Configuration of the gradient coils used for spatial encoding in three dimensions	33
4. Gradient coils	34
5. Phase and frequency encoding of XY slice	35
6. MRI pulse sequence	36
7. Illustration of brain MRI. An image pixel (i,j) is represented in the 2D MRI slice and an image voxel (x,y,z) is represented in 3D space	36
8. Example of EEG acquisition setup by Marchesotti et al.	37
9. Technomed EEG cup electrodes (left) and cap with g.SCARABEO EEG electrodes from g.tec (right)	38
10. International 10-20 system for EEG	39
11. International 10-10 system for EEG	40
12. Human EEG with resting state activity – alpha rhythm. Left: EEG traces (horizontal – time in seconds; vertical – amplitudes, scale 100 μ V). Right: power spectra of shown signals (vertical line – 10 Hz). Andrii Cherninskyi, 2015	41
13. Topographic maps of age related changes in non-rapid eye movement sleep	41
14. Illustrative example of PCA (left) and LDA (right)	46
15. Representation of a NN	49
16. Neuron structure	50
17. Activation functions in ANNs that introduce non-linearity	51
18. NN with k -layers consisting of fully connected neurons	52
19. Gradient descent visualisation	54
20. Gradient descent types	54
21. Gradient descent optimization using momentum	55
22. Validation loss and converging training	55
23. Single layer perceptron with one neuron	56
24. AlexNet block diagram	60
25. An overview of data collected throughout the ADNI study	63
26. MRI slices of the axial, sagittal, and coronal planes	66
27. Flowchart of histogram based classification	67
28. Maximum and minimum accuracy per slice for multiple classifiers	69
29. CN and AD image slices 5, 10, and 15	69
30. Slice accuracy for FT, Medium Gaussian SVM, and Medium KNN	70
31. Image slices 36, 37, and 38 for CN (top) and AD (bottom row)	70
32. Accuracy for majority and weighted vote for trained model on slice 36 and 37	71

33. Accuracy for majority and weighted vote for concatenated histograms	71
34. Axial plane: Histogram feature ranking	77
35. Axial plane: Classifiers ranking for histogram descriptors	77
36. Axial plane: Pre-trained CNN feature ranking	80
37. Axial plane: Classifiers ranking for pre-trained CNN features	80
38. Axial plane: Slice histogram for histogram descriptors. Slices are ordered from top to bottom of the patient's head	80
39. Axial plane: Slice histogram for pre-trained CNN features. Slices are ordered from top to bottom of the patient's head	80
40. Axial plane: MRI slice Nr. 88 for CN (top) and AD (bottom row)	80
41. Coronal plane: Histogram descriptor feature ranking	84
42. Coronal plane: Classifiers ranking for histogram descriptors	84
43. Coronal plane: Pre-trained CNN feature ranking	84
44. Coronal plane: Classifiers ranking for pre-trained CNN features	84
45. Coronal plane: Slice histogram for histogram descriptors. Slices are ordered from the front to the back of the patient's head	84
46. Coronal plane: Slice histogram for pre-trained CNN features. Slices are ordered from the front to the back of the patient's head	84
47. Coronal plane: MRI slice Nr. 62 for CN and AD. Best selection according to histogram descriptors.	85
48. Coronal plane: MRI slice Nr. 87 for CN and AD. Second best selection according to histogram descriptors.	85
49. Coronal plane: MRI slice Nr. 101 for CN and AD. Best selection according to pre-trained CNN features.	85
50. Coronal plane: MRI slice Nr. 77 for CN and AD. Second best selection according to pre-trained CNN features.	86
51. Sagittal plane: Histogram descriptor feature ranking	87
52. Sagittal plane: Classifiers ranking for histogram descriptors	87
53. Sagittal plane: Pre-trained CNN feature ranking	89
54. Sagittal plane: Classifiers ranking for pre-trained CNN features	89
55. Sagittal plane: Slice histogram for histogram descriptors. Peak locations are at slice Nr. 91 and 136. Slices are ordered from the left to the right of the patient's head	90
56. Sagittal plane: Slice histogram for pre-trained CNN features. Peak locations are at slice Nr. 88 and 139. Slices are ordered from the left to the right of the patient's head	90
57. Sagittal plane: Slice Nr. 88 for CN and AD. Left side of the brain	90
58. Sagittal plane: Slice Nr. 139 for CN and AD. Right side of the brain	90
59. Global: Top 10% histogram descriptor ranking	91
60. Global: Top 10% pre-trained CNN feature ranking	91
61. Cross validation procedure for ensemble classifier with weight calculation	98

LIST OF TABLES

1. ADNI and OASIS results for AD, MCI, very MCI (vMCI), and HC.	28
2. AlexNet structural details	60
3. Information about sex, age, HDRS, and EST-Q test results for subjects in the dataset, where in each table cell the number before the back slash is for healthy subjects and on the right of the back slash for people who have had depression (Healthy/Depressed)	65
4. AD and CN classification performance for slices from 31 to 40	68
5. AD and CN classification performance for testing data for slices from 31 to 40 using trained model of slice 36 and 37	69
6. Histogram-based local descriptors	73
7. Pre-trained convolutional neural networks	74
8. Axial plane: Best CN/AD classification accuracy using histogram descriptors	78
9. Axial plane: Top 10% CN/AD classification accuracy using histogram descriptors	78
10. Axial plane: Best CN/AD classification accuracy using pre-trained CNNs	79
11. Axial plane: Top 10% CN/AD classification accuracy using pre-trained CNNs	79
12. Coronal plane: Best CN/AD classification accuracy using histogram descriptors	82
13. Coronal plane: Top 10% CN/AD classification accuracy using histogram descriptors	82
14. Coronal plane: Best CN/AD classification accuracy using pre-trained CNNs	83
15. Coronal plane: Top 10% CN/AD classification accuracy using pre-trained CNNs	83
16. Sagittal plane: Best CN/AD classification accuracy using histogram descriptors	87
17. Sagittal plane: Top 10% averaged CN/AD classification accuracy using histogram descriptors	88
18. Sagittal plane: Best CN/AD classification accuracy using pre-trained CNNs	88
19. Sagittal plane: Top 10% CN/AD classification accuracy using pre-trained CNNs	89
20. Concatenated feature CN/AD classification accuracy using histogram descriptors for slices: Axial-88, Coronal-62, and Sagittal-91 for top 5 global (from Figure 59) and plane (from Figures 34, 41, and 51) feature ranking	92

21. Concatenated feature CN/AD classification accuracy using histogram descriptors for slices: Axial-88, Coronal-87, and Sagittal-91 for top 5 global (from Figure 59) and plane (from Figures 34, 41, and 51) feature ranking	92
22. Concatenated feature CN/AD classification accuracy using pre-trained CNNs for slices: Axial-88, Coronal-87, and Sagittal-139 for top 5 global (from Figure 60) and plane (from Figures 36, 43, and 53) feature ranking	93
23. Concatenated feature CN/AD classification accuracy using pre-trained CNNs for slices: Axial-88, Coronal-101, and Sagittal-139 for top 5 global (from Figure 60) and plane (from Figures 36, 43, and 53) feature ranking	93
24. Concatenated feature CN/AD classification accuracy using histogram descriptors selected according to Tables 8, 12, and 16	94
25. Concatenated feature CN/AD classification accuracy using pre-trained CNNs selected according to Tables 10, 14, and 18	94
26. EEG feature classification accuracy	100
27. Classifier accuracy for EEG features selected by univariate feature ranking using F-Tests	101
28. Classifier accuracy for EEG features selected by ReliefF algorithm	101
29. Selected electrodes based on F-test and ReliefF	101
30. Concatenated EEG feature classifier accuracy	102
31. Ensemble classifier accuracy	102

LIST OF ABBREVIATIONS

Acronyms

- AD** Alzheimer's Disease. 11–14, 23, 24, 26–28, 62, 63, 66–70, 72, 75–83, 85–94
- ADNI** Alzheimer's Disease Neuroimaging Initiative. 11, 13, 22, 26–28, 62, 63, 66, 67, 72, 75
- ANN** Artificial Neural Network. 11, 27, 51, 57
- APV** Alpha Power Variability. 29, 96, 97
- CN** Cognitively Normal. 11–14, 26–28, 62, 66–70, 75, 76, 78–80, 82, 83, 85–90, 92–94
- CNN** Convolutional Neural Network. 12–14, 28, 57–60, 72, 74–77, 79–81, 83–86, 88–94
- CSF** Cerebrospinal Fluid. 26, 27
- CT** Computed Tomography. 21, 30, 76, 95
- DFA** Detrended Fluctuation Analysis. 29, 96, 97, 99
- DT** Decision Tree. 27, 48, 66, 72, 75, 77, 81, 99
- EEG** Electroencephalography. 11, 14, 21, 22, 25, 26, 29, 30, 37–42, 62, 64, 65, 95–97, 99–102
- EST-Q** Emotional State Questionnaire. 13, 64, 65, 95
- fMRI** Functional Magnetic Resonance Imaging. 21, 24, 95
- FT** Fine Tree. 11, 66, 68, 70
- HC** Healthy Controls. 13, 26–28
- HDRS** Hamilton Depression Rating Scale. 13, 64, 65, 95
- HFD** Higuchi's Fractal Dimension. 29, 96, 97, 99
- HOG** Histogram of Oriented Gradients. 27, 72, 73, 76, 86, 91, 94
- KNN** K-Nearest Neighbors Algorithm. 11, 27, 47, 68, 70, 72, 75, 77, 81, 86, 93, 99
- LAP** Local Arc Pattern. 76, 77, 81, 86
- LBP** Local Binary Pattern. 27, 72, 73
- LDA** Linear Discriminant Analysis. 11, 29, 44, 46, 72, 75, 77, 81, 86, 93, 99
- LDN** Local Directional Number Pattern. 81, 91
- LDTP** Local Directional Texture Pattern. 76
- LGP** Local Gradient Pattern. 81, 86, 92
- LR** Linear Regression. 29, 47, 66, 68

LZC Lempel Ziv Complexity. 29, 96, 97, 100

MBC Monogenic Binary Coding. 77

MCI Mild Cognitive Impairment. 13, 26–28, 62

ML Machine Learning. 21, 24–26, 43, 46–49, 61, 66, 67, 72, 75, 76, 94, 99

MLP Multi Layer Perceptron. 57

MRI Magnetic Resonance Imaging. 11, 12, 21, 22, 24, 26–28, 30–33, 36, 62, 63, 66, 67, 72–77, 80, 81, 85, 86, 91–94

NB Naive Bayes. 47, 99

NMR Nuclear Magnetic Resonance. 30, 31, 34, 36

NN Neural Network. 11, 27, 28, 43, 49–54, 56, 57, 61, 72, 76, 93, 94

OASIS Open Access Series of Imaging Studies. 13, 26–28

PC Principal Components. 44, 45

PCA Principal Component Analysis. 11, 27, 28, 44–46, 92

PET Positron Emission Tomography. 21, 24, 26, 27, 62, 63

QDA Quadratic Discriminant Analysis. 72, 75, 77, 81

RBP Relative Band Power. 95, 97, 100, 102

ReLU Rectified Linear Unit. 50, 53, 57–59

RF Random Forrester. 26, 27, 48

RGP Relative Gamma Power. 29

RNN Recurrent Neural Networks. 57, 59

ROI Region-of-Interest. 28, 73

SASI Spectral Asymmetry Index. 29, 96, 97

SGD Stochastic Gradient Descent. 53, 54

SLP Single Layer Perceptron. 56

sMRI Structural Magnetic Resonance Imaging. 21, 24, 28

SVD Singular Value Decomposition. 44–46

SVM Support Vector Machine. 11, 26, 27, 48, 66, 68–70, 72, 75, 77, 81, 86, 91–93, 99

TE Echo Time. 36

TR Repetition Time. 35, 36

VBM Voxel-based Morphometry. 26, 27

WLD Weber Local Descriptor. 76, 77, 81, 86, 91, 92

LIST OF ORIGINAL PUBLICATIONS

Publications included in the thesis

- I. **E. Avots**, K. Jermakovs, M. Bachmann, L. Päske, C. Ozcinar, and G. Anbarjafari, "Ensemble approach for detection of depression using eeg features," *Entropy*, 2022.
Author contributions: Developed the method; wrote the code and contributed to manuscript.
- II. H. Elshatoury, **E. Avots**, G. Anbarjafari, ADNI, et al., "Volumetric histogram-based alzheimer's disease detection using support vector machine," *Journal of Alzheimer's Disease*, 2019.
Author contributions: Developed the methodology for the experiment; contributed to manuscript.
- III. **E. Avots**, T. Sapiński, M. Bachmann, and D. Kamińska, "Audiovisual emotion recognition in wild," *Machine Vision and Applications*, 2019.
Author contributions: Performed simulations; wrote the code and contributed to manuscript.
- IV. M. N. Sabet Jahromi, P. Buch-Cardona, **E. Avots**, K. Nasrollahi, S. Escalera, T. B. Moeslund, G. Anbarjafari, "Privacy-Constrained Biometric System for Non-Cooperative Users," *Entropy*, 2019.
Author contributions: Conducted simulations for baseline results; wrote parts of the discussion on the experimental results.
- V. M. N. Jahromi, M. B. Bonderup, M. Asadi-Aghbolaghi, **E. Avots**, K. Nasrollahi, S. Escalera, S. Kasaei, T. B. Moeslund, and G. Anbarjafari, "Automatic access control based on face and hand biometrics in a non-cooperative context," *IEEE Winter Applications of Computer Vision Workshops (WACVW)*, 2018.
Author contributions: Conducted simulations for baseline results; wrote parts of the discussion on the experimental results.
- VI. **E. Avots**, H. S. Arslan, L. Valgma, J. Gorbova, and G. Anbarjafari, "A new kernel development algorithm for edge detection using singular value ratios," *Signal, Image and Video Processing*, 2018.
Author contributions: Conducted simulations; wrote parts of the discussion on the experimental results.

Other published work of the author

- I. **E. Avots**, A. Vecvanags, J. Filipovs, A. Brauns, G. Skudrins, G. Done, J. Ozolins, G. Anbarjafari, and D. Jakovels, "Towards automated detection and localization of red deer cervus elaphus using passive acoustic sensors during the rut," Not in Thesis, *Remote Sensing*, 2022.
- II. A. Vecvanags, K. Aktas, I. Pavlovs, **E. Avots**, J. Filipovs, A. Brauns, G. Done, D. Jakovels, and G. Anbarjafari, "Ungulate detection and species classification from camera trap images using retinanet and faster r-cnn," Not in Thesis, *Entropy*, 2022.
- III. G. Anbarjafari, A. Anier, **E. Avots**, A. Georgadze, A. Hektor, M. Kiisk, M. Kutateladze, T. Lepp, M. Mägi, V. Pastsuk, et al., "Atmospheric ray tomography for low-z materials: Implementing new methods on a proof-of-concept tomograph," Not in Thesis, *arXiv preprint arXiv:2102.12542*, 2021.
- IV. A. Georgadze, M. Kiisk, M. Mart, **E. Avots**, and G. Anbarjafari, "Method and apparatus for detection and/or identification of materials and of articles using charged particles," Not in Thesis, *US Patent App. 16/977,293*, 2021.
- V. D. Kamińska, T. Sapiński, S. Wiak, T. Tikk, R. E. Haamer, **E. Avots**, A. Helmi, C. Ozcinar, and G. Anbarjafari, "Virtual reality and its applications in education: Survey," Not in Thesis, *Information*, 2019.
- VI. I. Ofodile, A. Helmi, A. Clapés, **E. Avots**, K. M. Peensoo, S. M. Valdma, A. Valdmann, H. Valtna-Lukner, S. Omelkov, S. Escalera, et al., "Action recognition using single-pixel time-of-flight detection," Not in Thesis, *Entropy*, 2019.
- VII. **E. Avots**, M. Madadi, S. Escalera, J. Gonzalez, X. Baro, P. Pällin, and G. Anbarjafari, "From 2d to 3d geodesic-based garment matching," Not in Thesis, *Multimedia Tools and Applications*, 2019.
- VIII. M. Daneshmand, **E. Avots**, and G. Anbarjafari, "Proportional error back-propagation (peb): Real-time automatic loop closure correction for maintaining global consistency in 3d reconstruction with minimal computational cost," Not in Thesis, *Measurement Science Review*, 2018.
- IX. A. Clapés, O. Bilici, D. Temirova, **E. Avots**, G. Anbarjafari, and S. Escalera, "From apparent to real age: Gender, age, ethnic, makeup, and expression bias analysis in real age estimation," Not in Thesis, *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2018.
- X. R. E. Haamer, K. Kulkarni, N. Imanpour, M. A. Haque, **E. Avots**, M. Breisch, K. Nasrollahi, S. Escalera, C. Ozcinar, X. Baro, et al., "Changes in facial expression as biometric: A database and benchmarks of identification," Not in Thesis, *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, 2018.
- XI. J. Guo, Z. Lei, J. Wan, **E. Avots**, N. Hajarolasvadi, B. Knyazev, A. Kuharenko, J. C S. J. Junior, X. Baro, H. Demirel, et al., "Dominant and

- complementary emotion recognition from still images of faces," Not in Thesis, *IEEE Access*, 2018.
- XII. J. Gorbova, **E. Avots**, I. Lüsi, M. Fishel, S. Escalera and G. Anbarjafari, "Integrating vision and language for first-impression personality analysis," Not in Thesis, *IEEE MultiMedia*, 2018.
- XIII. M. Daneshmand, A. Helmi, **E. Avots**, F. Noroozi, F. Alisinanoglu, H. S. Arslan, J. Gorbova, R. E. Haamer, C. Ozcinar, and G. Anbarjafari, "3d scanning: A comprehensive survey," Not in Thesis, *arXiv preprint arXiv:1801.08863*, 2018.
- XIV. **E. Avots**, M. Daneshmand, A. Traumann, S. Escalera, and G. Anbarjafari, "Automatic garment retexturing based on infrared information," Not in Thesis, *Computers & Graphics*, 2016.
- XV. A. Skorodumovs, **E. Avots**, J. Hofmanis and G. Korāts, "Hardware design of the energy efficient fall detection device," Not in Thesis, *Latvian Journal of Physics and Technical Sciences*, 2016.
- XVI. G. Korāts, J. Hofmanis, A. Skorodumovs and **E. Avots**, "Fall detection algorithm in energy efficient multistate sensor system," Not in Thesis, *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2015.

PREFACE

Dear Reader,

Before you is a doctoral thesis titled "Brain abnormality detection using statistical analysis of individual structural connectivity networks and EEG signals." It has been written to fulfil the graduation requirements of the Engineering and Technology doctoral school programme from the Faculty of Science and Technology at the Institute of Technology, University of Tartu, Estonia.

The time spent during the doctoral studies has not been easy, but I have obtained competence in project management, teaching, supervising, scientific writing, reviewing articles, and doing research. I have also discovered that adversity is a necessary component of the process, as it helps to build character and provides invaluable professional and personal lessons.

I would like to express gratitude to my supervisor, Prof. Gholamreza Anbarjafari, for his superb leadership and assistance throughout the studies. I also want to express my gratitude to Assoc. Prof. Maie Bachmann from TalTech for her assistance in providing information and direction for the clinical depression research. A crucial component of my doctoral studies was working in the iCV research lab, where I could always find support and assistance from its great members.

Finally, I want to express my gratitude to my family and friends for supporting me. Also, I want to thank you for reading this thesis. I hope you find it interesting.

Egils Avots
Tartu, 2023

1. INTRODUCTION

Any anatomical or functional abnormalities in the brain are referred to as brain anomalies. They might already exist at birth or might appear later in life as a result of an illness, injury, or other circumstances. Brain irregularities may have a significant negative impact on a person's physical and mental health, necessitating ongoing medical care. Depending on the particular ailment and the severity of the symptoms, treatment options could include drugs, surgery, therapy, and lifestyle modifications.

Brain abnormalities can affect many different parts of the brain, which can result in a wide range of symptoms and medical conditions. Typical abnormalities of the brain include structural, functional, and neurological ones.

Anomalies in the brain's structure known as structural differences include variations in the size and shape of certain brain regions. Brain deformities like hydrocephalus and cerebral asymmetry are examples of such anomalies. Techniques like Computed Tomography (CT) and Structural Magnetic Resonance Imaging (sMRI), often simply referred to as Magnetic Resonance Imaging (MRI), can be used to find structural anomalies.

Functional anomalies are variations in the neural activity, connections, or brainwave patterns that occur in various regions of the brain. Techniques like Functional Magnetic Resonance Imaging (fMRI), Positron Emission Tomography (PET), or Electroencephalography (EEG) are used to identify such aberrations.

Disorders of the nervous system, known as neurological illnesses, can cause a variety of physical and mental symptoms. Neurological disorders like epilepsy, Parkinson's disease, Alzheimer's disease, and clinical depression are just a few that can result from abnormalities in the brain.

This thesis focuses on supporting the diagnostic aspects of Alzheimer's disease and clinical depression using advanced statistical models applied to data depicting structural changes and electrical activity of the brain. This process refers to pattern recognition in medical diagnostics and can be defined as the ability to identify and discern the typical anatomical and physiological features depicted in a specific data modality such as MRI or EEG, as well as detect any deviations from these normal appearances that may suggest the presence of pathological conditions. This implies that certain criteria should be met to be proficient in pattern recognition, whether the analysis is done by a medical professional or by an automated system. In computer science, the field of pattern recognition is discussed together with topics such as data acquisition, data pre-processing, feature extraction, and Machine Learning (ML).

Even though this thesis describes potential diagnostic methods for classification of medical conditions, the information provided in this thesis is for general informational purposes only and is not intended to be a substitute for professional medical advice, diagnosis, or treatment. The author of this thesis is not a medical professional, and the content should not be considered as medical advice.

This part of the introduction will focus on a brief overview of the upcoming chapters. The first four chapters comprise the initial half of the thesis and serve as complementary chapters that offer background knowledge, while the second part adopts the format of a collection of research publications.

This thesis begins its investigation into the complex realm of brain abnormalities and their diagnosis. The first chapter introduces the reader to the main topics of interest: Alzheimer’s disease and clinical depression, establishing the groundwork for a more in-depth look into the procedures used to detect them.

In the second chapter, the narrative expands to reviewing existing scholarly work. The emphasis here is on understanding the links between Alzheimer’s disease and MRI, as well as the interaction between clinical depression and EEG.

The third chapter delves deeper into the investigation, shedding light on the key technologies in this study: MRI and EEG. This section delves into the intricacies of these medical imaging techniques, from their fundamental concepts to the devices used and data acquisition.

The thesis transitions to the fourth chapter, introducing the reader to the world of machine learning. This section covers many types of machine learning, as well as feature extraction methods, classification algorithms, and deep learning, that are essential tools for automated medical diagnosis.

The fifth chapter delves deeper into the study’s complexities, revealing additional details of the used data. The MRI dataset from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) and the EEG dataset from Tallinn University of Technology play critical roles. The investigation’s methodology, recording processes, and participant selection criteria emerge as essential components.

As the sixth chapter begins, image histograms take centre stage in the detection of Alzheimer’s disease using MRI images. This section discusses dimensionality reduction and the generation of MRI slice histograms as important components of the diagnostic toolset, guiding the reader through the obtained results.

The seventh chapter expands on the previous chapter by introducing the use of several image texture descriptors in the classification of Alzheimer’s disease using MRI images. Histogram-based local descriptors and features extracted from pre-trained convolutional neural networks serve as analytical tools in this case, allowing for classifications to be examined on several image planes.

The eighth chapter shifts the focus and transfers the reader into the topic of clinical depression. An ensemble technique, based on linear and non-linear features collected from EEG recordings, takes front stage. The reader is led through this investigation with objectivity by means of feature selection, machine learning methods, and informed explanations of results.

The thesis ends with the ninth chapter, with reflections on the obtained results of the proposed methods and a vision for future studies in the ever-changing terrain of technological advancements and their use in automated medical diagnosis.

1.1. Alzheimer's Disease

Alzheimer's Disease (AD) is a degenerative brain condition that impairs thinking, behavior, and memory. AD is the most common form of dementia, which is a term for a general decrease in cognitive function severe enough to interfere with daily activities and accounts for approximately 60-80% of all dementia cases [1]. Mostly, AD affects elderly people, particularly individuals aged over 60 [2]. The disease's name pays homage to Dr. Alois Alzheimer, who first recognized the condition in 1906. AD causes the degeneration and death of brain cells called neurons, which results in a decline in cognitive function. The illness typically starts out with minor confusion and memory loss before progressing to more serious symptoms, including trouble communicating, trouble identifying friends and family, disorientation, and mood changes [3]. Individuals may become fully dependent on others for their care as the condition progresses. AD currently has no known cure, although there are medications, treatments, and lifestyle modifications that can slow down its progression and assist in managing symptoms, but eventually patients die from the condition. Alzheimer's patients at the advanced stage need 24-hour guidance on hygiene, clothing, and eating. Movement, sitting up, and swallowing become harder due to brain changes, with the most common cause of death being aspiration pneumonia. Due to the significant effects of the disease on society, AD has provoked interest among researchers throughout the world [4], [5]. Ron et al. [6] reported that in 2006, 26.6 million people were found to suffer from AD and the researchers also predicted that there would be approximately 106.8 million patients affected by AD in 2050.

Most individuals with AD are aged 65 and above. Although younger people can receive a pre-diagnosis, the risk of AD typically increases with age. According to the Alzheimer's Association, close to 50 million people suffer from AD, or another form of dementia, around the world [7].

Figure 1 shows how the brain of an AD patient shrinks dramatically, restricting the functions of the brain [8], [9]. The disease results in the death of brain cells, which causes tissue loss and shrinkage, as shown in existing research. Studies that utilize structural imaging of the brain also evidence the more severe effects of AD on particular parts of the brain, such as the hippocampus area. Signs of shrinkage affecting the hippocampus can indicate the onset of AD. However, the standard volume of shrinkage within such cases and how this corresponds to individual patients are contested [10].



A brain without the disease. A brain with advanced Alzheimer's. How the two brains compare.

Figure 1: Illustration of brain with and without AD [8]

The process of determining whether someone has AD can be difficult [11], and it frequently calls for a thorough examination that takes the following measures:

1. Medical history and physical exam: This entails examining the medical history to rule out other possible causes of symptoms as well as analyzing the patient's symptoms, such as memory loss and changes in behavior.
2. Cognitive and neurological testing: Tests that evaluate memory, language, attention, problem-solving, and other cognitive processes, such as the famous clock drawing test [12]. These examinations can aid in locating certain regions of the brain that the illness may have damaged.
3. Laboratory tests: Blood tests to eliminate other possible reasons for memory loss, such as thyroid issues, vitamin shortages, or more specialized biomarker measurements, such as beta-amyloid 4, tau, phospho-tau, and others, may be part of this process.
4. Imaging tests: This could include brain scans such as PET or MRI scans, which can produce images of the brain and assist in identifying structural abnormalities that could be an indication of AD.

Based on the patient's symptoms and test results, a healthcare professional will determine if they have AD after conducting a complete evaluation. In some situations, a conclusive diagnosis cannot be obtained until the patient has passed away and been subjected to a postmortem analysis. It's crucial to remember that other disorders, like depression, thyroid issues, and pharmaceutical side effects, can produce symptoms that are similar to those of AD, so it's crucial to undergo a thorough evaluation to rule out other potential causes.

The physical makeup of the brain, including the size and shape of various brain regions, can be depicted in sMRI images [13], [14]. On the basis of these images, ML algorithms can be trained to recognize structural alterations in the brain that are indicative of AD, such as shrinking in the hippocampus, a key memory-related brain region. Images from a fMRI display the levels of activity in various brain regions during particular tasks. These images can be used to train ML algorithms to recognize patterns of brain activity that are indicative of AD, such as decreased activity in areas of the brain related to memory and attention. To increase the precision of the diagnosis, these ML algorithms can also be employed in conjunction with additional clinical data, such as the results of cognitive tests.

Although ML has shown promise in the diagnosis of AD, it is crucial to remember that due to the associated costs of clinical trials and certification of the developed algorithms as medical devices, they are not yet commonly employed in clinical practice. In order to increase accuracy and present a more complete picture of the patient's state, ML algorithms can only be used as an additional tool in conjunction with other diagnostic procedures performed by medical professionals.

1.2. Clinical Depression

A mental health illness known as clinical depression, commonly referred to as major depressive disorder, is characterized by persistent feelings of melancholy and hopelessness and a lack of interest in once-enjoyable activities. Physical signs of depression might include exhaustion, adjustments to eating and sleep patterns, and trouble concentrating. Depression is not merely a passing emotion of melancholy or "the blues." This serious disorder may have a negative impact on a person's everyday functioning and ability to function at work, home, and in relationships. Depression is an illness that is curable and is thought to impact millions of individuals globally. A mix of medications, psychotherapy, and lifestyle modifications like regular exercise, sound sleep habits, and a healthy diet are frequently used as part of the treatment of depression.

A mental health expert, such as a psychologist or psychiatrist, makes the diagnosis of clinical depression. The diagnostic procedure may include the following:

1. Medical and mental health history assessment: An expert in mental health will inquire about symptoms, their length, and how they affect daily living. Inquiries regarding any physical or psychiatric issues, family history, and drug or alcohol usage will also be made.
2. Psychological evaluation: To determine whether the person satisfies the criteria for depression, the mental health professional will evaluate the person's mood, thoughts, actions, and general functioning.
3. Physical examination: Physical examinations are done to rule out any underlying medical issues that might be causing the patient's symptoms.
4. Lab tests: In order to rule out any other conditions that might be causing a patient's symptoms, a doctor may do blood tests or other laboratory tests.

Research has revealed that, as compared to healthy people, those with depression display unique patterns of brain activity that may be observed using EEG [15], a non-invasive method that measures the electrical activity of the brain by applying electrodes to the scalp. The EEG data analysis then can uncover depression-related patterns, and ML algorithms can be trained to examine EEG waves or features derived from EEG and spot depression related abnormalities.

Although this combination of EEG and ML has demonstrated promise in the diagnosis of depression [16], it is crucial to keep in mind that it is still in its infancy and that further research is required to determine its efficacy. Moreover, doctors do not carry out the diagnosis of depression solely on the basis of EEG findings. Making an accurate diagnosis requires a complete evaluation by a mental health specialist, which includes a detailed medical and psychiatric history. EEG can offer useful data and ML can aid in the diagnosis of depression, but due to the intricacy of the decision-making process and the scarcity of clinical trials to validate the algorithms and the complexity of the used models, medical professionals typically do not trust autonomous decisions generated by ML algorithms.

2. RELATED WORK

This chapter will cover various methods for medical image based Alzheimer's disease classification using classical ML algorithms and deep learning. The second part expands on EEG based feature classification for depression classification.

2.1. Alzheimer's Disease and Magnetic Resonance Imaging

Computer vision and ML methods are frequently applied in studies on the diagnosis of AD. They are used in segmentation, classification, and regression procedures involving MRI. In case of AD studies, researchers use the following notation to differentiate between patient groups: Alzheimer's Disease (AD), Mild Cognitive Impairment (MCI), and Cognitively Normal (CN). Cognitively normal people are also known as Healthy Controls (HC) in some studies. This section compares and discusses the performance of earlier research findings.

The work of Zhang and Shen [17] suggests the use of two components: a multi-class feature selection and a multi-model Support Vector Machine (SVM). With the data of 45 AD, 91 MCI, and 50 HC, they conducted their experiment using MRI, PET, and Cerebrospinal Fluid (CSF) data from the ADNI dataset [18]. Within the AD vs. HC experiment, the use of MRI returned a result of 84.8% accuracy; the use of PET had 84.5% accuracy, and the use of CSF had 80.5% accuracy. Within their research, Westman et al. [19] investigated MRI and CSF biomarkers employing the data of 369 subjects, which they attained from the ADNI dataset. They involved 96 AD, 162 MCI, and 111 HC in their study. When distinguishing between AD and HC, this study returned a result of 87% for accuracy, as they employed multivariate data analysis. The combination of CSF and MRI distinguished between AD and HC most effectively. In their study, Zhou et al. [20] suggest the use of MRI data in combination with a mini-mental state examination (MMSE) to identify AD. The study involved 59 AD and 127 CN subjects, and applied FreeSurfer, which transforms the cortex into a 2D surface, to determine 55 volumetric variables. When combining MMSE with MRI, this study returned an accuracy rate of 92.4%; however, when MMSE was not employed, a result of 78.2% accuracy was achieved.

By utilizing MRI volumes, voxel-based PET signal intensities, CSF biomarkers, and categorical genetic information, Grey et al. [21] achieved 89% accuracy. The data was gathered from 147 participants, including 37 AD, 75 MCI, and 35 HC. The Random Forrest (RF) classifier was employed and determined an accuracy rate of 89%. In a study by Papakostas et al. [22], Voxel-based Morphometry (VBM) and Deformation-based morphometry (DBM) were combined to distinguish between AD and HC. The study involved 98 females, whose data were sourced from the Open Access Series of Imaging Studies (OASIS) dataset [23]. By employing the mean and standard deviation of gray matter probability voxels within each cluster feature from the VBM model of feature extraction, the study

achieved 85% accuracy. In a different approach, Zhang et al. [24] used MRI, PET, and CSF as biomarkers to identify AD. They involved 51 AD and 52 HC in their research. Combining these methods achieved 93.2% accuracy.

Nayaki and Varghese [25] used MRI scans from the ADNI dataset. Local patterns were used to encode the loss of gray matter volume and classify it using SVM. In CN, MCI, and AD classifications, an accuracy of 76% was reached with Local Binary Pattern (LBP), and the best result was achieved using Local Graph Structure (LGS) with 81% accuracy.

Ben Ahmed et al. [26] also used MRI images from the ADNI dataset and classified hippocampal visual features of CN, MCI, and AD patients. The experimental results show that the classification for MRI versus CN subjects achieves accuracies of 87% for a subset of the ADNI dataset. For MCI, they reached accuracies of 78.22% and 72.23% for MCI/CN and MCI/AD, respectively.

Sarwinda and Bustamam [27] proposed to use Histogram of Oriented Gradients (HOG) features of three orthogonal planes for the classification of AD using MRI images. The feature dimensionality was reduced using probabilistic principal component analysis (PPCA), and then RF was used as a binary classifier to classify the features. The HOG-TOP approach achieved an accuracy of 95.8% for AD/CN, 94.4% for MCI/CN, and 93.64% for AD/MCI.

Altaf et al. [28] performed AD classification using images and clinical features based on the ADNI dataset. From image feature descriptors, they tested the gray level co-occurrence matrix (GLCM), Scale Invariant Feature Transform (SIFT), HOG, and LBP. In the case of AD vs. CN, the binary classification results of SVM, ensemble, K-Nearest Neighbors Algorithm (KNN), and Decision Tree (DT) accuracy were in the range of 50% to 60%, while other combinations such as AD/MCI and CN/MCI with GLCM reached 80% accuracy. Their proposed method reached 86.7% accuracy for CN/MCI.

In a more modern approach, neural networks are used to learn features automatically. A feedforward multi-layer Neural Network (NN) was applied in the work of Mahmood et al. [29] and used to classify MRI images. The data for 230 subjects was sourced from the OASIS dataset. The model was tested on each of the participants, and the features were reduced to 150 using Principal Component Analysis (PCA) to achieve an accuracy of 89.22%. By reducing the number of features to 100, the authors achieved a lower accuracy rate of 86.47%. Alternatively, Ding et al. [30] utilized VBM and texture analysis for the identification of 54 AD and 58 CN patients from the ADNI dataset. Finally, using SVM with the RBF kernel, an accuracy of 92.86% was reached.

For example, Cao et al. [31] combined 2D and 3D images in a multi-view based multi-model (MVMM) learning framework that was evaluated on the ADNI dataset. Their proposed model could classify MCI/HC with 87.50% accuracy and MCI/AD with 83.18% accuracy. Similarly, Cohen et al. [32] used categorical data from the ADNI dataset, which also included MRI images, to distinguish between CN, MCI, and AD with an accuracy of 87.2% using an Artificial Neural Network

Table 1: ADNI and OASIS results for AD, MCI, very MCI (vMCI), and HC.

Dataset	Method	Classification	Data	Samples	Acc.
ADNI	DemNet [40]	AD/MCI/HC	sMRI	900	91.85
ADNI	ROI SRNN [37]	AD/HC	sMRI	80	94.63
ADNI	LeNet-5 CNN [41]	AD/HC	fMRI	43	96.85
ADNI	3D-CNN [36]	AD/MCI/HC	sMRI	210	97.61
ADNI	Autoencoder NN [42]	MCI/HC	fMRI	100	87.50
ADNI	SGBN [43]	MCI/HC	fMRI	120	74.83
ADNI	SGBN [43]	AD/HC	PET	103	71.47
ADNI	SVM [42]	MCI/HC	fMRI	100	82.10
OASIS	PCA+NN [29]	AD/MCI/vMCI/HC	sMRI	230	89.92
OASIS	McRBFN [39]	MCI/HC	sMRI	198	77.56
OASIS	SRAN [38]	MCI/HC	sMRI	60	91.18
OASIS	SVM [38]	MCI/HC	sMRI	60	90.57

(ANN) classifier and an accuracy of 88.3% using a 1D Convolutional Neural Network (CNN) classifier. Basaia et al. [33] automated CNN to predict the diagnosis of AD and MCI in patients who will convert to AD (c-MCI) based on a single cross-sectional brain MRI scan from the ADNI and Milan datasets. The learning algorithm could distinguish AD/CN with 99.2% accuracy and c-MCI/CN with 87.1% accuracy.

Several NN-based AD classifiers use 3D or voxel-based input features to incorporate spatial information into their models. This is due to the fact that NN-based methods can manage more complex types of data inputs. The alternative is the Region-of-Interest (ROI) method proposed by Jack et al. [34]. This is typically referred to as the whole brain morphometric method, as described by Magnin et al. [35]. Hosseini-Asl et al. [36] chose to use a CNN on structured 3D sMRI data with a 3D convolutional autoencoder, which provided them with some of the highest accuracy values in Table 1. They evaluated both binary AD/HC and tri-state classification for AD/MCI/HC, achieving accuracies of 97.61% and 89.1%, respectively. Yang et al. [37] used voxel-level information to incorporate the entire brain as their input. This was compared to ROI methods, where the use of specific features reduced the over-fitting problem and increased the accuracy of the results from 73.81% to 94.63% accuracy. Mahanand et al. [38] addressed the issue of voxel over-fitting using PCA, which they incorporated into their final classifier, the PCA Selfadaptive Resource Allocation Network with an accuracy of 91.18%. In order to enhance results in [29] and [39], PCA and independent component analysis (ICA) were again used for dimensionality reduction and ROI selection.

Among NN-based classification methods, the ADNI dataset [18] and the OASIS dataset [23] are widely used for both training and evaluation, which facilitates comparisons between variously implemented methods. For better dataset results comparison, some examples are listed in Tables 1.

2.2. Clinical Depression and Electroencephalography

Brain activity can also be monitored with EEG while the patient is resting, performing mental tasks, or responding to stimuli. De Aguiar Neto et al. [44] have found that various linear and non-linear features, including absolute and relative band powers, are promising biomarkers with which to identify a brain affected by depression. Knott et al. [45] used a three-way multivariate analysis of variance (MANOVA) to examine absolute and relative band power in EEG signals and found that depressed patients displayed higher relative beta power compared to controls. This is the case at each electrode location, and absolute beta power is even elevated at certain electrode locations. Bachmann et al. [46] were the first to advance the use of Alpha Power Variability (APV) and Relative Gamma Power (RGP). RGP is used to identify high frequency components, and APV detects variations in the power and frequency of the alpha band. Across multiple EEG channels, there were substantial differences between depressive patients and the control groups. Consequently, the Linear Regression (LR) classification accuracy was found to be 81%. The relative asymmetry among higher and lower frequency bands was presented using a Spectral Asymmetry Index (SASI).

Among healthy and depressed patients, Hinrikus et al. [47] found that SASI values differed markedly across all channels. Already, the analysis of single EEG channels has been shown to be effective for detecting depression [46], [48]. Higuchi's Fractal Dimension (HFD) can be used to quantify a signal's fractal dimension in the time domain [49]. Bachmann et al. [50] utilized this non-linear method, which they evaluated using Students' T-tests for two-tailed distribution, to uncover differences between depressed and healthy individuals' EEG signals. Using this method, each EEG channel was found to have statistically significant discrepancies, and 94% participants within the depression-affected group were identified as depressed. In addition, the use of HFD showed that 76% of participants within the control group were non-depressive.

Lempel and Ziv's published non-linear Lempel Ziv Complexity (LZC) [51], has been used effectively to measure the complexity of EEG signals and detect different mental conditions [52], [53]. EEG signals were assessed using Detrended Fluctuation Analysis (DFA) [54] to detect long-term correlations between signals and uncovered substantial discrepancies between healthy and depressive participants [55]. Linear Discriminant Analysis (LDA) was also used and produced a classification accuracy of 70.6%. Upon combining DFA and SASI, the accuracy of LDA classification grew to 91.2% [48]. The diagnostic potential of linear (SASI, APV, and RGP) and non-linear (HFD, DFA, and LZC) features in the classification of depression has been evaluated in the work of Bachmann et al. [46]. Using APV or RGP, single channel classification that employed LR generated 81% accuracy. By combining two linear measures, namely SASI and RGP, an accuracy of 88% was achieved. In turn, by using both linear and non-linear measures, 92% accuracy was reached.

3. BRAIN STRUCTURAL AND FUNCTIONAL IMAGING

This chapter will give a brief introduction to the technology of MRI and the main concepts about how the MRI image is constructed. The second part will describe EEG acquisition methods and common human brain wave notations.

3.1. Magnetic Resonance Imaging

Using the nuclear magnetic resonance phenomenon, the technique for acquiring tomographic medical pictures for the examination of interior organs and tissues is called Magnetic Resonance Imaging (MRI) [56]. A medical use of Nuclear Magnetic Resonance (NMR) is MRI. While CT is better at displaying bone features, MRI is better at displaying soft tissue. On MRI, as opposed to CT, nerves, muscles, ligaments, and tendons may be seen much more clearly. The MRI technique is also essential for assessing the spinal cord and brain. MRI can distinguish between white and gray matter in the brain. MRI is successfully used in the diagnosis of inflammatory, infectious, and oncological diseases, as well as in the study of joints, all parts of the spine, mammary glands, the heart, abdominal organs, the small pelvis, and blood vessels. This is due to the high accuracy and clarity of the images obtained.

The development of digital technology that allows for image processing, computer modeling of surgical operations, and getting functional information, such as mapping the cerebral cortex, is the main advancement in contemporary radiation diagnostics. MRI is frequently used to confirm diagnoses in patients who have already undergone examinations with X-ray, ultrasound, CT, and scintigraphy, sometimes known as a gamma scan.

3.1.1. Nuclear Magnetic Resonance

NMR is realized in nuclei with non-zero spins. The most interesting for medicine are the nuclei of hydrogen (^1H), carbon (^{13}C), sodium (^{23}Na), and phosphorus (^{31}P), since they are all present in the human body, and the most (63%) hydrogen atoms are found in fat and water. For these reasons, modern magnetic resonance tomographs are most often "tuned" to hydrogen nuclei - protons. In the absence of an external field, the spins and magnetic moments of the protons are oriented randomly. If a proton is placed in an external magnetic field, then its magnetic moment will be either codirectional or oppositely directed to the magnetic field, and in the second case, its energy will be higher. A spin particle placed in a magnetic field of strength B can absorb a photon with a frequency ν , which depends on its gyromagnetic ratio γ .

$$\nu = \gamma B \tag{3.1}$$

For hydrogen, $\gamma = 42.58 \text{ MHz/T}$. A particle can undergo a transition between two energy states by absorbing a photon. A particle at the lower energy level absorbs a photon at the upper energy level. The energy of a given photon must exactly match the difference between these two states. The energy of a photon, E , is related to its frequency, ν , through the Planck constant ($h = 6.626 \cdot 10^{-34} \text{ J} \cdot \text{s}$).

$$E = h\nu \quad (3.2)$$

In NMR, the quantity ν is called the resonance or Larmor frequency. $\nu = \gamma B$ and $E = h\nu$, therefore, in order to cause a transition between two spin states, a photon must have the energy

$$E = h\gamma B \quad (3.3)$$

When the photon energy matches the difference between the two spin states, energy is absorbed. The strength of the constant magnetic field and the frequency of the radio frequency magnetic field must strictly match each other (resonance). In NMR experiments, the photon frequency corresponds to the radio frequency range. In clinical MRI, to display hydrogen, ν is typically between 15 and 80 MHz.

At room temperature, the number of protons with spins at the lower energy level slightly exceeds their number at the upper level. The signal in NMR spectroscopy is proportional to the difference in the level populations. The number of excess protons is proportional to B_0 . This difference in a field of 0.5T is only 3 parts per million; in a field of 1.5T - 9 parts per million. However, the total number of excess protons in 0.02 ml of water in a 1.5T field is 6.02×10^{15} . The higher the magnetic field strength, the better the image.

In equilibrium, the net magnetization vector is parallel to the direction of the applied magnetic field B_0 and is called the equilibrium magnetization M_0 . In this state, the Z -component of the magnetization M_Z is equal to M_0 . M_Z is also called longitudinal magnetization. In this case, there is no transverse (M_X or M_Y) magnetization. By sending an radio frequency pulse at the Larmor frequency, it is possible to rotate the net magnetization vector in a plane perpendicular to the Z axis, in this case, the X - Y plane.

Upon termination of the radio frequency pulse, the net magnetization vector will recover along the Z -axis, emitting radio frequency waves. The time constant describing how the M_Z returns to equilibrium is called the spin-lattice relaxation time (T_1).

$$M_Z = M_0(1 - e^{-t/T_1}) \quad (3.4)$$

T_1 relaxation occurs in a volume containing protons. However, the bonds between protons in molecules are not the same. These links are different for each tissue. One 1H atom can be very strongly bonded, as in adipose tissue (body fat), while another atom can have a weaker bond, for example, in water. Strongly

bound protons release energy much faster than weakly bound protons. Each tissue releases energy at a different rate, which is why MRI has such good contrast resolution.

The T_1 relaxation time, also known as the spin-lattice relaxation time, is a measure of how quickly the net magnetization vector (NMV) recovers to its ground state in the direction of B_0 . The return of excited nuclei from the high energy state to the low energy, or ground state, is associated with a loss of energy to the surrounding nuclei.

T_2 relaxation describes processes in the X - Y plane, and T_1 relaxation describes processes occurring in the Z direction. T_2 relaxation is the process by which the transverse components of magnetization (M_{XY}) decay or dephase. As described by Felix Bloch (1946), T_2 relaxation is considered to follow first order kinetics, resulting in a simple exponential decay with a time constant T_2 . Thus T_2 is the time required for the transverse magnetization to fall to approximately 37% ($1/e$) of its initial value. Synonyms for T_2 relaxation are transverse and spin-spin relaxation.

Immediately after exposure to the radio frequency pulse, the net magnetization vector (now called transverse magnetization) begins to rotate in the X - Y plane about the Z axis. All vectors have the same direction because they are in phase. However, they do not retain this state. The net magnetization vector begins to phase out (out of phase) due to the fact that each spin packet experiences a magnetic field slightly different from the magnetic field experienced by other packets and rotates with its own Larmor frequency. At first, the number of dephased vectors will be small but rapidly increase until the moment when the phase coherence disappears; there will not be a single vector that coincides in direction with another. The total magnetization in the XY plane tends to zero, and then the longitudinal magnetization increases until M_0 is along Z .

The time constant describing the behavior of the transverse magnetization, M_{XY} , is called the spin-spin relaxation time, T_2 which describes the interactions between protons in their immediate environment (molecules). T_2 relaxation is a decaying process, meaning high phase coherence at the beginning of the process, but rapidly decreasing until the coherence disappears completely at the end. The signal is strong at the beginning, but quickly weakens due to T_2 relaxation. The signal shown in Figure 2 is called the decay of magnetic induction, or free induction decay (FID).

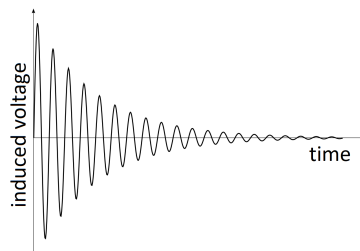


Figure 2: Free induction decay

$$M_{XY} = M_{XY0}e^{-t/T_2} \quad (3.5)$$

The phase shift rate is different for each tissue type. Dephasing in adipose tissue is faster than in water. In addition, T_2 relaxation proceeds much faster than T_1 relaxation. T_2 relaxation occurs in tens of milliseconds, while T_1 relaxation can take up to seconds.

3.1.2. Magnetic Resonance Imaging Device

MRI is a medical imaging technique used in radiology to form pictures of the anatomy and physiological processes of the body. The MRI device includes a magnet, gradient coils, and radio frequency coils (see Figure 3).

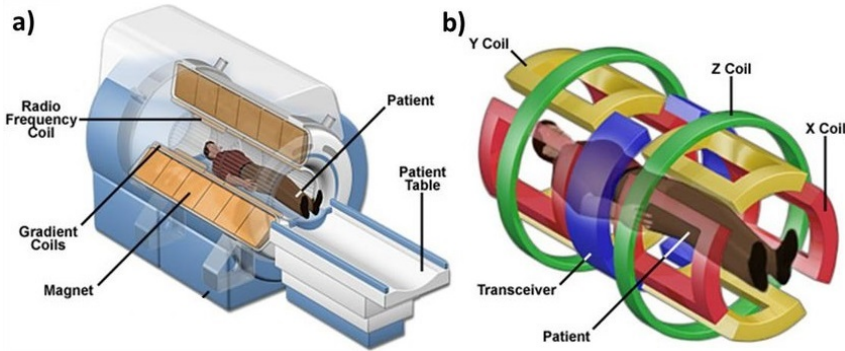


Figure 3: Image (a) Schematic illustration of the MRI system and (b) Configuration of the gradient coils used for spatial encoding in three dimensions [57]

Permanent Magnet. The quality and speed of image acquisition depend on the strength of the field created by powerful magnets in MRI scanners. Modern MRIs use either permanent or superconducting magnets. Permanent magnets are cheap and easy to use, but they do not allow the creation of magnetic fields with an intensity greater than 0.7 T. Most of the MRI machines are models with superconducting magnets (0.5–1.5 T). Ultra-strong field tomographs above 3.0 T are very expensive to operate. With MRI tomographs with a field below 1 T, it is impossible to make high-quality tomography of internal organs since the power of such devices is too low to obtain high-resolution images; therefore, only examinations of the head, spine, and joints can be carried out.

Gradient Coils. Gradient coils are located inside the magnet. Gradient coils allow the creation of additional magnetic fields superimposed on the main magnetic field B_0 . There are three sets of coils. Each set can create a magnetic field in a specific direction: X, Y or Z. For example, when current enters the Z gradient, a uniform linear field change is created in the Z direction along the long axis of the body. In the center of the magnet, the field has a strength B_0 , and the resonant frequency is equal to ν_0 , but at a distance ΔZ the field changes by the amount ΔB (see Figure 4), and accordingly, the resonance frequency also changes.

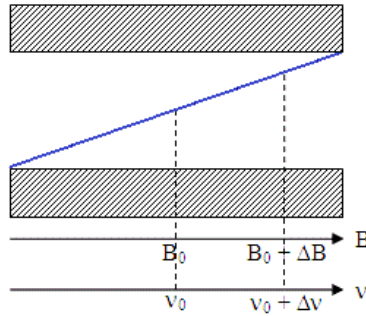


Figure 4: Gradient coils

By adding a gradient magnetic disturbance to the overall uniform magnetic field, localization of the NMR signal is provided. The action of the gradient, which provides the choice of the cut, provides selective excitation of protons precisely in the desired region. The speed, signal-to-noise ratio, and resolution of the tomograph depend on the power and speed of the coils.

Radio Frequency Coils. The radio frequency coils create a B_1 field that rotates the net magnetization in a pulse train. They also register transverse magnetization as it progresses in the XY plane. Radio frequency coils come in three main categories: transmit and receive, receive only, and transmit only. The radio frequency coils serve as emitters of B_1 fields and receivers of radio frequency energy from the object under study.

Signal Coding. When a patient is in a uniform magnetic field B_0 , all protons from head to toe align along B_0 . They all rotate at the Larmor frequency. If radio frequency excitation pulse is generated to translate the magnetization vector into the X - Y plane, all the protons react, and a response signal occurs, but there is no localization of the signal source.

Slice Encoding Gradient. When the Z -gradient is on, an additional magnetic field G_Z is generated in this direction, superimposed on B_0 . A stronger field means a higher Larmor frequency. Along the entire slope of the gradient, the field B is different, and, therefore, the protons rotate at different frequencies. If an radio frequency pulse with a frequency of $\nu + \Delta\nu$ is generated, only protons in a thin cut will react because they are the only ones rotating at the same frequency. The response signal will only be from protons from this cut. Thus, the signal source is localized along the Z axis. The protons in this cut rotate with the same frequency and have the same phase. The slice contains a huge number of protons, and the localization of the sources along the X and Y axes is unknown. Therefore, further coding is required to accurately determine the direct source of the signal.

Phase Encoding Gradient. To further encode the protons, the G_Y gradient is switched on for a very short time. During this time, an additional gradient magnetic field is generated in the Y direction. In this case, the protons will have slightly different rotational speeds. They no longer rotate in phase. The phase difference will accumulate. When the G_Y gradient is off, the protons in the slice will rotate at the same frequency but have a different phase.

Frequency Encoding Gradient. For left-right encoding, a third G_X gradient is included. The protons on the left side rotate at a lower frequency than those on the right. They accumulate an additional phase shift due to differences in frequencies, but the already acquired phase difference obtained by encoding the gradient phase in the previous step is retained.

Thus, magnetic field gradients are used to localize the source of the signals that are received by the coil:

- The G_Z gradient selects the axial slice.
- The G_Y gradient creates rows with different phases.
- The G_X gradient produces columns at different frequencies.

Phase encoding is performed for one line only in one step. To scan a whole slice, the phase and frequency encoding processes must be repeated several times. This creates small volumes called voxels that are encoded by a unique combination of frequency and phase (see Figure 5).

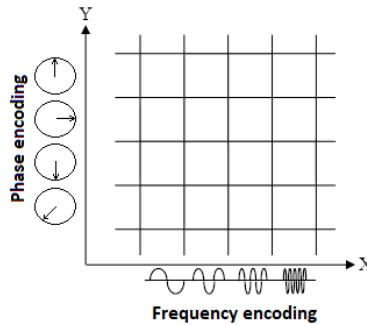


Figure 5: Phase and frequency encoding of XY slice

The number of protons in each voxel determines the amplitude of the radio frequency wave. The received signal, coming from different areas of the body, contains a complex combination of frequencies, phases, and amplitudes.

Pulse Sequences. Figure 6 shows a diagram of the pulse sequence. After applying a 90° excitation pulse, the net magnetization is in the $X - Y$ plane. Phase displacement due to T_2 relaxation begins immediately. It is because of this dephasing that the signal drops sharply. To maintain phase coherence and provide the best signal, a 180° pulse is applied shortly after the 90° radio frequency pulse. The 180° pulse causes the spins to phase out. When all spins are in phase, the signal becomes high again, and the image quality is much higher. The time between repetitions of a sequence is called Repetition Time (TR). With each successive sequence, the magnitude of the phase encoding gradient changes.

First, the slice selective gradient (1) (G_{PE}) is turned on. A 90° radio frequency pulse is applied at the same time (2). Then, a phase encoding gradient (3) (G_{PE}) is turned on to perform the first phase encoding step. G_{SS} (4) is turned on again during the 180° rephasing pulse (5), so the same protons that were excited by

the 90° pulse are affected. After that, a frequency-encoding (6) (G_{PE}) is applied, during which a signal (7) is received. The complete process must be repeated several times. TR is the time between two 90° excitation pulses, and Echo Time (TE) is the time between the 90° excitation pulse and the echo.

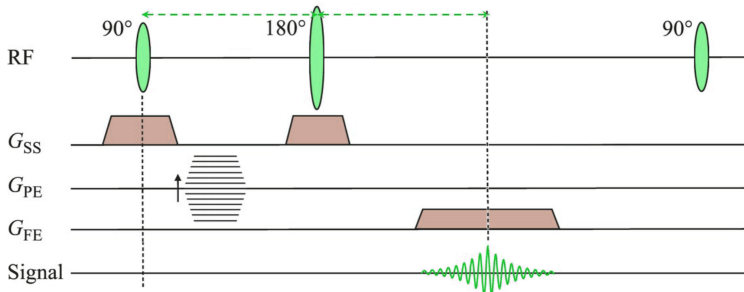


Figure 6: MRI pulse sequence

During NMR scanning, two relaxation processes T_1 and T_2 occur simultaneously, where $T_1 \gg T_2$. The image contrast is highly dependent on these processes and on how fully each of them manifests itself at the selected scan parameters, TR and TE.

Scan Output. After the scanning procedure is done, the MRI images can be represented as 2D slices, similar to X-ray images, or in 3D using specialized software, as shown in Figure 7. The scans are represented as 3D voxel cubes and accompanied by metadata with parameters and corrections of the MRI scan.

When examining the scans, one must note the conventions used to understand which is the right and left side of the brain. *Radiological* convention assumes that the patient is approached from feet first, and *neurological* convention assumes that the observer stands at the patient's head and is looking top down at the brain as if performing brain surgery.

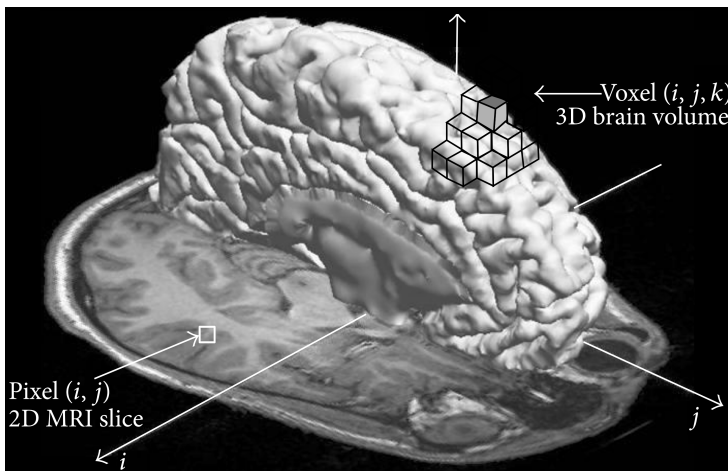


Figure 7: Illustration of brain MRI. An image pixel (i,j) is represented in the 2D MRI slice and an image voxel (x,y,z) is represented in 3D space [58]

3.2. Electroencephalography

Currently, one of the most informative methods for studying the human brain from the standpoint of its holistic system activity is electroencephalography. This method is based on recording the electrical activity of neurons by placing noninvasive electrodes on the surface of the scalp — an electroencephalogram (EEG). EEG signals are complex curves consisting of many frequency components.

The EEG method, using modern mathematical methods of data processing and analysis, remains one of the most common for studying the human brain. EEG gives the possibility of qualitative and quantitative analysis of the functional state of the brain and its reactions under the action of stimuli and when performing various activities. In clinical practice, EEG is used to diagnose a number of mental and neurological diseases and, above all, epilepsy. The recording setup usually consists of EEG electrodes, a signal amplifier, and a computer capable of signal processing. To reduce outside inferences, the recordings are often done in shielded rooms, as shown in Figure 8.

Currently, it is generally accepted that EEG waves are mainly the result of the summation of postsynaptic potentials: excitatory postsynaptic potential (EPSP) and inhibitory postsynaptic potential (IPSP). Also, some contribution is made by the electrical activity of glial cells.



Figure 8: Example of EEG acquisition setup by Marchesotti et al. [59]

3.2.1. Electroencephalogram Registration and Analysis

EEG is recorded by placing electrodes on the skin surface, which are connected with the help of conductors to the panel of the biopotential amplifier, or electroencephalograph. Electrodes that are applied to the skin should have a low transition resistance of no more than 3-5 kOhm, a small degree of polarization, and high corrosion resistance. The electrodes are placed in a special helmet mesh or a helmet with sets of already mounted electrodes, as shown in Figure 9. Currently, there are two ways to register an EEG: monopolar and bipolar. In bipolar leads, the difference potentials are measured between two electrically active areas of the brain, where both electrodes are on the scalp. With a monopolar lead, the potential difference is recorded between electrically active and electrically neutral locations such as the earlobe or nose bridge. The most widespread are two international systems for the location of electrodes: 10-20% and 10-10%.



Figure 9: Technomed EEG cup electrodes (left) and cap with g.SCARABEO EEG electrodes from g.tec (right)

International 10-20% System. The International 10-20 System shown in Figure 10 is a worldwide recognized method of positioning electrodes on the scalp during an EEG study. It got its name due to the fact that the distance from any electrode to another is defined as 10% or 20% of individually measured head dimensions. This method was developed to support standardized testing methods that enable the compilation, reproduction, efficient analysis, and comparison of the results of a subject's study (clinical or research) using the scientific method. The 10-20 system is based on the relationship between the location of the electrode and the main area of the brain, in particular the cerebral cortex. The 10-20 system is recommended by the International Federation of Electroencephalography and Clinical Neurophysiology. This system is suitable for most patients and is also convenient in terms of time and effort. This is the most commonly used system in EEG studies. The names of the electrodes include the first letter of the Latin name of the area on which the electrode is placed and the number indicating the side and location of the electrode within this area: pre-frontal (Fp), frontal (F), temporal (T), parietal (P), occipital (O), central (C), midline (Z).

The electrodes are placed in the following locations:

- Fp1, Fp2 - prefrontal (prefrontal),
- F3, F4 - frontal (frontal),
- Fz - mid-frontal,
- C3, C4 - central (central),
- Cz - central vertex,
- P3, P4 - parietal (parietal),
- Pz - central parietal,
- F7, F8 - anterior temporal,
- T3, T4 - middle temporal (temporal),
- T5, T6 - posterior temporal,
- O1, O2 - occipital (occipital),
- A1, A2 - ear.

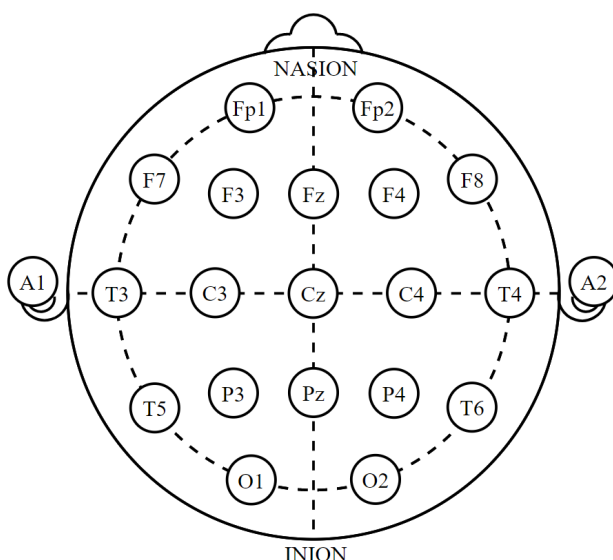


Figure 10: International 10-20 system for EEG

International 10-10% System. In the 10-10 electrode arrangement system shown in Figure 11, which is a modification of the "10-20" system, the number of leads is increased; this system provides for the installation of additional electrodes, displaced with respect to the position of the electrodes in the 10-20 system forward indicated by (') or backward ("); for example, lead C1' in the "10-10" system is anterior to lead C1 in the system 10-20. The reason for using this system is the need for a high density of electrodes, for example, to study areas of particular interest or create high resolution topographic brain mapping.

Artifacts. During a recording, the EEG signal can be influenced by external interference called artifacts that can be separated into physiological and technical categories. Physiological artifacts refer to eye movements, muscle contractions,

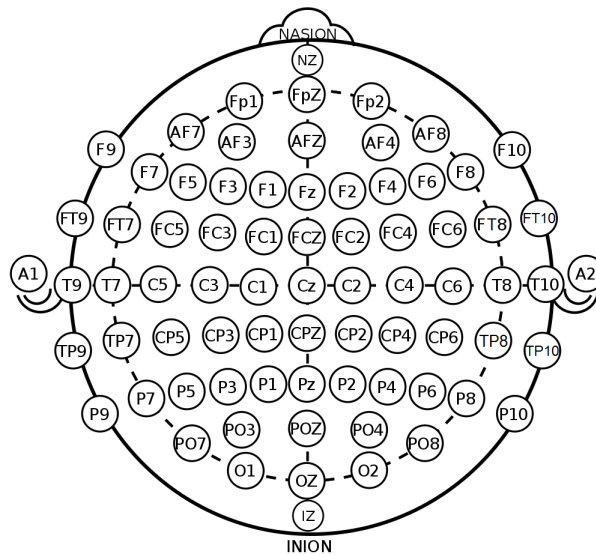


Figure 11: International 10-10 system for EEG

swallowing movements, cardiograms, etc. Technical artifacts include network interference with a frequency of 50 Hz due to the presence of electromagnetic fields generated by the electrical network in the room, as well as artifacts associated with the movement of wires and loose electrodes.

To obtain an artifact-free EEG recording, it is necessary that the subject be in a relaxed sitting position in a comfortable chair during the experiment. The amount of external light and sound stimulation should be minimized. Artifacts that still appear in the EEG are typically removed using special software during the subsequent processing of the EEG recording.

Analysis. Besides the visual method of EEG analysis, which is mainly used when conducting clinical examinations of patients, various other methods use computer analysis of EEG, and one of them is spectral analysis, which allows one to mathematically isolate and investigate EEG frequency characteristics. To do so, a fast Fourier transformation (FFT) is used on the original signal, which is represented as a sum of sinusoids of various frequencies and amplitudes. The results are displayed in graph format for each EEG channel, with amplitude peaks in the appropriate frequency ranges. Thus, the spectral analysis allows to measure the amplitude or power of the investigated frequency range and compare the severity of EEG rhythms (see example in Figure 12).

According to spectral analysis data, one can build topographic maps of the severity of frequency ranges, also known as topographic brain mapping (Quantitative EEG, QEEG), as shown in Figure 13, which provides a map of the electrical activity of the brain. It is done by processing the recorded EEG activity from a multi-electrode recording and may be used to relate the connectivity and functionality of the brain through imaging. The mapping between electrodes helps to find functionally integrated relationships between spatially separated brain regions.

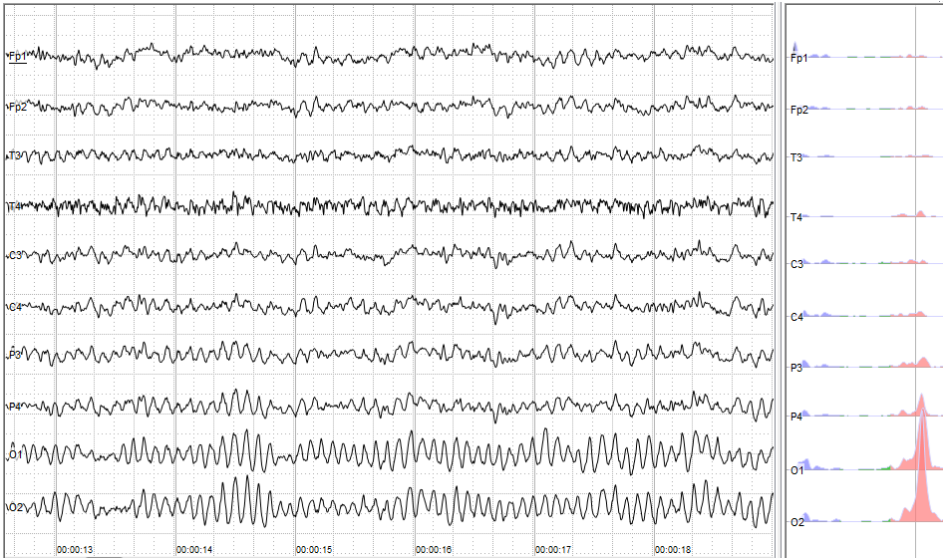


Figure 12: Human EEG with resting state activity – alpha rhythm. Left: EEG traces (horizontal – time in seconds; vertical – amplitudes, scale 100 uV). Right: power spectra of shown signals (vertical line – 10 Hz). Andrii Cherninskyi, 2015

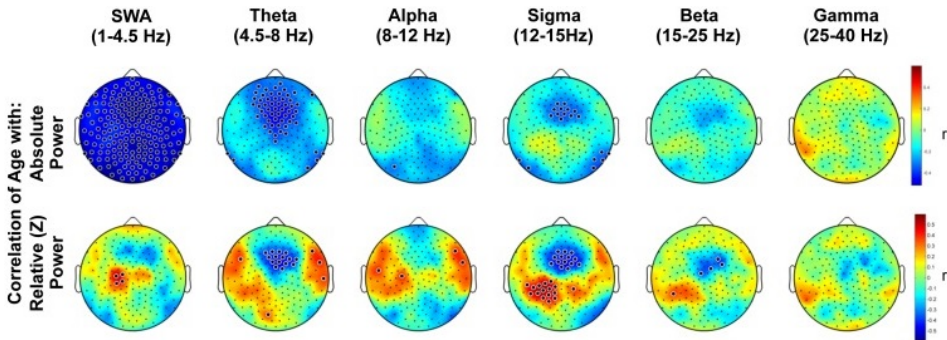


Figure 13: Topographic maps of age related changes in non-rapid eye movement sleep [60]

3.2.2. Human Electroencephalogram Rhythms

EEG activity, representing waves of approximately one constant frequency, is called rhythm, and the most pronounced EEG rhythm is called dominant. If the amplitude of the waves is rhythmic activity that gradually increases and then decreases, it is said that rhythmic wave activity is spindle-shaped. Reducing the amplitudes of potential oscillations without changing the frequency response is called rhythm depression. A process expressed in the formation of regular, ordered rhythmic activity and an increase in the amplitude of oscillations is denoted by rhythm synchronization. Violation of the wave rhythm process with the replacement of wave activity with less regular activity of a different frequency and lower amplitude is called desynchronization.

It is widely accepted to divide the main EEG rhythms into the following categories:

Delta rhythm - The delta rhythm includes EEG waves at frequencies ranging from 0.5 to 3.5 Hz. The EEG of a healthy adult delta rhythm can only be present in very small amounts - its amplitude does not exceed 40 μV , usually about 20 μV . The delta rhythm is the main rhythm of slow-wave sleep (SWS) when its amplitude reaches 300 μV or more.

Theta rhythm - According to various studies, the boundaries of theta rhythm slightly differ, but on average, oscillations with a frequency from 4 to 8 Hz are considered to be in theta rhythms. Normally, the wave amplitude does not exceed 40 μV . An increase in the theta rhythm index can indicate pathology or reflect certain functional conditions associated either with a decrease in the level of brain activity or, on the contrary, with concentration of attention and cognitive and emotional activation.

Alpha rhythm - The alpha rhythm is well modulated, high-amplitude rhythmic activity with a frequency in the range from 7.5 to 13 Hz and an average amplitude 60-80 μV . The alpha rhythm is best expressed in healthy adult subjects in a state of calm wakefulness with closed eyes, mainly in the parieto-occipital regions of the brain. Depression of the alpha rhythm indicates a general activation of the cerebral cortex. Eye opening or mental activity is usually accompanied by a depression of the alpha rhythm. An important feature of alpha activity is its functional asymmetry under different cognitive and emotional stressors.

Mu rhythm - Sometimes outwardly similar to an alpha rhythm and corresponding to it in frequency, it appears as brief bursts of 7 to 11 Hz activity. Unlike the alpha rhythm, the mu-rhythm is most expressed in the area of the precentral gyrus, and it reacts weakly to the opening of the eyes. The depression of this rhythm is observed when the subject performs movements.

Beta rhythm - Frequency of 14 to 35 Hz and amplitude up to 15 μV . It is recorded best of all in the frontal and temporal regions. Synchronized and desynchronized beta activity can show different types of emotions and cognitive processes. Synchronization of the beta rhythm in the frontal regions is associated with attention processes.

Gamma rhythm - Generally, the range of gamma rhythm refers to EEG fluctuations with a frequency of 30 to 70 Hz. The gamma rhythm manifests itself in humans and animals in a wide variety of situations and is associated with such phenomena as wakefulness, perception, attention, and consciousness.

After obtaining and pre-processing the EEG signals they are analysed by methods which involve the examination of various aspects of EEG data to identify abnormal patterns associated with specific neurological conditions, such methods might include: visual inspection, frequency analysis, time-frequency analysis, event-related potentials, source localization, connection analysis, quantitative EEG, pattern recognition, and machine learning.

4. MACHINE LEARNING

The medical data modalities used in this thesis are treated from the perspective of data analysis, with the aim of finding patterns in the data. Pattern recognition is a data analysis process that uses machine learning (ML) which is a field of science where algorithms are used to predict an outcome based on given input data [61]–[64]. In a classical ML setup, input data is prepared in a structured manner and presented to the ML model in the form of feature vectors. From the perspective of the ML model, it does not matter what the data represents, for example, flower petal lengths as in the famous Fisher’s Iris dataset, locations of human face feature points, or sensor data from a smart electrical grid. What matters is the consistency of the available data and the existence of a statistical difference between the represented states, classes or patterns.

This chapter will focus on a few classical feature extraction algorithms, provide a short overview of well-known classical ML algorithms, and expand on the present-day topic of deep learning.

4.1. Machine Learning Types

It is possible to apply ML to raw data, for example, audio samples or images, but a clear difference has to be made between the classical ML approach and the use of neural networks (NN). The classical approach usually starts with raw data that is preprocessed, for example, by removing outliers or normalizing it. What typically follows is a feature extraction step where the extracted features are fed to a ML model to solve the particular problem. On the other hand, with enough data, NN can learn the features of the underlying data and also perform classification or regression tasks. It is also possible to use a pre-trained NN for a different task or problem, this method is known as transfer learning.

There are three primarily types of ML:

- *Unsupervised learning*: Unsupervised learning is a type of self-organized learning that helps to find previously unknown patterns or relationships in a data set without labels. It is also referred to as clustering, which can be used for sorting databases, data mining, and creating labels.
- *Supervised learning*: An important part of supervised learning is a dataset with labels. The correct answers are known beforehand, and the data is split into training and testing sets. The task of a supervised learning algorithm is to learn from the training set and predict discrete values from the test set or a target value in regression problems.
- *Reinforcement learning*: In reinforcement learning, an agent tries to optimize its rewards based on its interaction with the environment. The focus is on finding a balance between exploration of an unknown environment and exploitation of acquired knowledge. The agent’s behavior is modeled on

Markov decision process and actions are determined by policy and value functions. Examples of reinforcement learning include self-driving robots, agents in games, trading bots in financial markets, etc.

4.2. Feature Extraction Algorithms

Feature extraction is usually performed on raw or preprocessed data. Based on the nature of the data (digital images, time series, or categorical data), there exist field-specific feature descriptors. Handcrafted features can be created from simpler statistical metrics such as mean, median, variance, histograms, entropy, information gain, etc. Nevertheless, there exist standard approaches, such as PCA, SVD and LDA, that help select the most relevant features and reduce the number of dimensions in feature vectors.

4.2.1. Principal Component Analysis

PCA is a dimension reduction method that is often used to reduce the dimensions of large datasets by transforming long feature vectors into smaller ones [65], [66]. By selecting the most relevant Principal Components (PC), it is possible to retain most of the information. The aim of PCA is to transform the feature space to maximize variance in each principal component. PC are new variables that are constructed as linear combinations of the initial variables. These combinations are done in such a way that the new variables are uncorrelated; this is an important property of PCA and is referred to as the orthogonality (independence) of the PC.

Let there be a matrix of variables X with dimensions $(I \times J)$, where I is the number of samples (rows), and J is the number of independent variables (columns), which are $J > 1$. The PCA uses new formal variables t_a ($a = 1, \dots, A$), which are a linear combination of the original variables x_j ($j = 1, \dots, J$)

$$t_a = p_{a1}x_1 + \dots + p_{aJ}x_J \quad (4.1)$$

Using the new variables, the matrix X is decomposed into the product of two matrices T and P

$$X = TP^t + E = \sum_{a=1}^A t_a p_a^t + E \quad (4.2)$$

The matrix T is called the scores matrix. Its dimension is $(I \times A)$, the matrix P is called the loadings matrix. Its dimension is $(J \times A)$. E is the matrix of residuals, dimension $(I \times J)$.

The new variables t_a are called Principal Components, hence, the method itself is called Principal Component Analysis. The number of columns, t_a in matrix T and p_a in matrix P , is A , which is called the number of PC. This value is less than the number of variables J and the number of samples I .

There are several methods for PCA calculation, but the four methods discussed by Wu et al. [67] are considered as classical: Non-linear Iterative Partial Least

Squares (NIPALS), POWER, Singular Value Decomposition (SVD) and eigen decomposition (EVD). The following description is for the NIPALS recursive algorithm, which calculates one component at each step. First, the original matrix X is transformed (i.e., centered) into the matrix E_0 , $a = 0$, followed by listed steps:

1. Choose an initial vector t .
- 2.

$$p^t = \frac{t^t E_a}{t^t t} \quad (4.3)$$

- 3.

$$p = \frac{p}{(p^t p)^{1/2}} \quad (4.4)$$

- 4.

$$t = \frac{E_a p}{p^t p} \quad (4.5)$$

5. Check convergence.

After calculating the next (a -th) component, put $t_a = t$ and $p_a = p$. To obtain the next component, it is necessary to calculate the remainders $E_{a+1} = E_a - t p^t$ and apply the same algorithm to them, replacing the index a with $a + 1$.

$$E_{a+1} = E_a - t p^t \quad (4.6)$$

After the PC space has been constructed, the new samples X_{new} can be determined by their score matrices T_{new} . In the PCA method, this done like this:

$$T_{new} = X_{new} P \quad (4.7)$$

PCA is closely related to another value decomposition technique, called SVD.

4.2.2. Singular Value Decomposition

SVD is a convenient technique when working with matrices, shows the geometric structure of the matrix and allows for visualization of the available data. This method is used in solving a variety of problems, from approximation by the least squares method and solving systems of equations to compression and image recognition. Since calculating the rank of a matrix is a task that occurs very often, SVD can be called a fairly popular method.

For any real $n \times n$ matrix A , there are two real orthogonal $n \times n$ matrices U and V such that

$$U^T A V = \Lambda \quad (4.8)$$

Also, U and V can be chosen so that the diagonal elements of Λ have the form

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_r > \lambda_{r+1} = \dots = \lambda_n = 0 \quad (4.9)$$

where r is the rank of the matrix A . In particular, if A is invertible, then

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n > 0 \quad (4.10)$$

The index r of the element λ_r is the actual dimension of the matrix A 's own space. The columns of the matrices U and V are called left and right singular vectors, respectively, and the values of the diagonal of the matrix Λ are called singular values. Let A be an $m \times n$ -matrix and let it be associated with a linear operator, also denoted by A . The SVD formula $A = U\Lambda V^T$ can be reformulated in geometric terms. A linear operator that is mapping elements of space \mathbb{R}^m into elements of space \mathbb{R}^n can be represented as sequentially performed linear operations of rotation, stretching, and, again, rotation. The number of non-zero elements on the diagonal of the matrix Λ is the actual dimension of the matrix A . Therefore, the components of the SVD clearly show geometric changes when the linear operator A represents the set of vectors from one vector space to another.

4.2.3. Linear Discriminant Analysis

Discriminant analysis is a method to find a linear combination of features that assumes that class features have different Gaussian distributions. To train a LDA based classifier, the fitting function estimates the parameters of a Gaussian distribution for each class in such a way as to minimize the variance and maximize the distance between the means of the two classes. To predict a class, the data is projected into LDA feature space, and the trained model finds the class with the smallest misclassification cost [68]–[70]. An example of PCA and LDA comparison is shown in Figure 14.

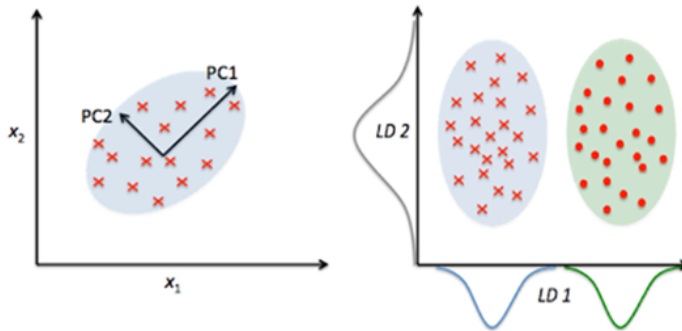


Figure 14: Illustrative example of PCA (left) and LDA (right) [71]

LDA is a generalization of Fisher's linear discriminant, a method in statistics, pattern recognition, and ML that is used to find a linear combination of features that describes or separates two or more classes or events. In the case of the problem of constructing decision functions for recognizing two classes of objects w_1 and w_2 according to some description $X = (x_1, x_2, \dots, x_L)$, the linear classifier turns out to be a convenient method that allows the construction of a decision function using different optimization criteria, including the Fisher criterion.

Besides using LDA feature space for component selection, it is mostly utilized as a classifier. However, there are more advanced algorithms available. The following sections will give an overview of dedicated classification algorithms.

4.3. Conventional Machine Learning Algorithms

Conventional ML techniques and deep learning each have their own unique set of qualities as well as advantages and disadvantages. ML algorithms that have been around for a while are typically easier to interpret; they are also scalable to structured data and need fewer computer resources. Deep learning is particularly effective when it comes to learning hierarchical representations from raw data, dealing with unstructured data, and obtaining state-of-the-art performance on certain tasks. The choice between classical ML and deep learning depends on the problem at hand, the readily available data, the level of interpretability that is required, and the available computational resources. Both methods are essential to the development of artificial intelligence and have the potential to work together in a complementary manner to address a wide variety of challenges [72]–[74].

Both of these approaches are often used for supervised learning, where the main task of the ML algorithm is to learn a solution to the given problem, where the model is general enough to correctly classify or predict a target value for unseen data. The performance of the trained model highly depends on the quality of the input data. If the data is not structured correctly, there is noise in the dataset, or there are incorrect labels, it will reflect in the prediction results. The following sections give a brief description of the most well-known conventional ML algorithms used in classification problems.

4.3.1. Logistic Regression

LR is a commonly used technique in binary classification problems. It usually makes use of the sigmoid function. This function can map the values of the used features into a probability value ranging between 0 and 1. Classification is done using a certain threshold to label the data point as one of two classes.

4.3.2. Naive Bayes

A Naive Bayes (NB) classifier, which is based on Bayes' theorem, assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature [75], [76]. It is often used as a baseline classifier for comparison with other ML algorithms; nevertheless, its accuracy can be improved when combined with kernel density estimation.

4.3.3. K-nearest Neighbors

The KNN algorithm uses feature similarity to predict the values of new samples based on how closely they match the points in the training set. The similarity is usually measured as a distance between points, where the Euclidean distance is one of the most popular choices [76]–[79]. The choice of 'K' determines the number of neighbors considered, and the algorithm is non-parametric, making minimal assumptions about the underlying data distribution.

4.3.4. Decision Tree

DT classifiers are structured similarly to a tree structure, where instances are classified according to their feature values. DTs classify instances by sorting them down the tree from the root to some leaf node, which provides the classification of the instance. An instance is classified by starting at the root node of the tree, testing the attribute specified by this node, then moving down the tree branch corresponding to the value of the attribute [80]–[82]. An extension of DT is RF [83] classifier which fits a number of DT classifiers on sub-samples of the dataset and uses averaging to improve the prediction accuracy.

4.3.5. Support Vector Machine

SVM is an algorithm in ML used to mostly solve classification problems, although support vector regression (SVR) can be used to solve regression problems. A SVM analyzes input data and recognizes patterns in a multi-dimensional feature space called the hyper-plane. An SVM model represents data as points in space, mapped so that the samples of separate categories are clearly separable. New samples are then mapped into that same space and predicted to belong to a category based on the previously established hyper-planes [69], [70], [84]–[86]. The error is being minimized in the SVM algorithm by maximizing the margin between the features and the hyperplane separating the two classes [87], [88]. Non-linear separation, for example, the XOR problem, can be solved by non-linear kernel functions such as a radial basis function [87].

4.3.6. Classifier Ensemble

An ensemble of classifiers is a method where several classifier decisions are combined in a manner that allows voting for the classification of new examples. Given that misclassified predictions are uncorrelated, the ensemble can correct for errors made by any individual classifier, leading to better overall accuracy [89]–[92]. Some of the main algorithms used in the ensemble method are bagging, boosting, and stacked generalization [93].

Bagging predictors [94] divide the dataset into sets of training data and train the same classifier on those sets; the predictions of each classifier are combined, and the final prediction is based on majority voting [93], which is a process of combining all classifiers and making predictions based on the majority result of all the classifiers [95]–[97].

In the boosting method [98], majority voting is done on a number of weak classifiers, and the combined predictions of those weak classifiers result in a strong classifier. For example, the AdaBoost [93] method works by training a weak classifier on subsets of the training data, whose weight is included in the final classifier. The output is the weighted vote of the predicted classes from the combined weak classifiers.

4.4. Deep Learning

Deep learning is a potent subfield of ML that has had a revolutionary impact in a variety of industries. The inspiration for developing NNs comes from the human brain and it is made up of stacked layers of artificial neurons that are connected to one another, as shown in Figure 15. These layers each carry out computations and pass information on to the next layer. The neurons are using activation functions, which are essential components found inside NNs that permit sophisticated calculations by introducing non-linearity. The most well-known activation functions are the sigmoid, ReLU, and softmax.

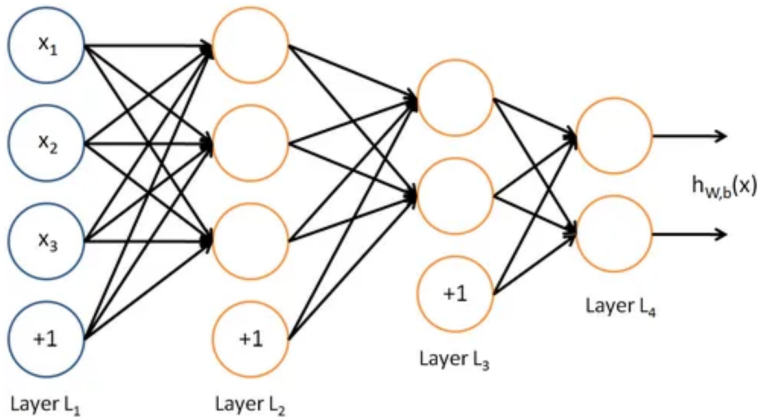


Figure 15: Representation of a NN [99]

To understand NNs, there are several concepts that should be covered. The first concept is feedforward NNs, which are the simplest and most frequently employed architecture in deep learning. Their structure consists of information moving in a single direction from input to output levels without any linkages for feedback.

The next concept is the backpropagation algorithm [100], which is an essential method for the training of NNs. Backpropagation entails computing gradients and iteratively adjusting network parameters in order to reduce the amount of variance that exists between the anticipated outputs and those that are actually produced. Through this iterative process, networks are able to enhance their performance by learning the NN weights.

The work horse of learning weights and biases is the optimization technique known as gradient descent, which is an essential part of the process of training NNs. There are various optimization algorithms, such as gradient descent, stochastic gradient descent, Adam, and others and several variations for feeding the data to the network, such as batch and mini-batch.

The goal of this overview is to provide insight into the fundamental ideas behind NNs, including activation functions, feedforward architectures, and the algorithms that are responsible for their training.

4.4.1. Neurons and Activation Functions

A NN is made up of many different components, the most important of which are the neurons and the activation functions. Neurons are responsible for doing computations on the incoming data, while activation functions are responsible for introducing non-linearities in order to capture complicated relationships. When it comes to successful learning and obtaining the desired level of performance, the proper selection of activation functions is absolutely necessary.

Each neuron receives data as input, processes that data according to its own set of rules, and then creates an output. A weighted sum of the inputs is performed during the calculation that takes place inside of a neuron, which is then followed by an activation function that introduces non-linearity (see Figure 16).

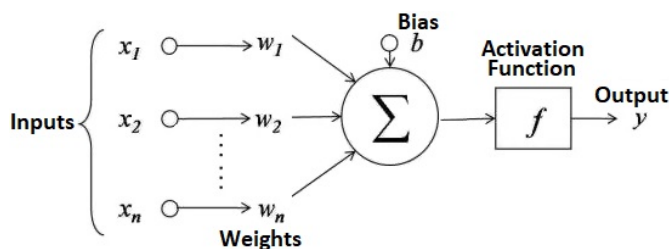


Figure 16: Neuron structure [99]

The weights and biases that are associated with neurons define the strength of the connections that neurons have with other neurons. While the weights function as a multiplier that applies to the inputs, the biases are additional constants that are added to the weighted total. In order to achieve the best possible results from the network, these parameters are refined through the training process.

In order to calculate the weighted sum, each input is multiplied by the weight that is associated with it, and then the results are added together. A neuron's weighted sum, denoted by the symbol z , can be expressed mathematically as the dot product of the input vector, denoted by x , and the weight vector, denoted by w , along with the bias, denoted by b : $z = (w \cdot x) + b$.

The following are examples of often used activation functions:

- Sigmoid [101]: The sigmoid function assigns a value between 0 and 1 to the weighted total, and it does this by mapping the sum. It features a smooth S-shaped curve and might be helpful when solving problems involving binary classification or when modeling probability.
- Hyperbolic Tangent (tanh) [102]: The hyperbolic tangent function is comparable to the sigmoid function, except it maps the weighted sum to a value that falls in the range of -1 and 1 and is centered around zero.
- Rectified Linear Unit (ReLU) [103]: When a negative input is received, the output of the ReLU is set to zero, and when a positive input is received, the output remains linearly proportionate to the positive input. ReLU has seen broad adoption due to its capacity to ease the vanishing gradient problem.

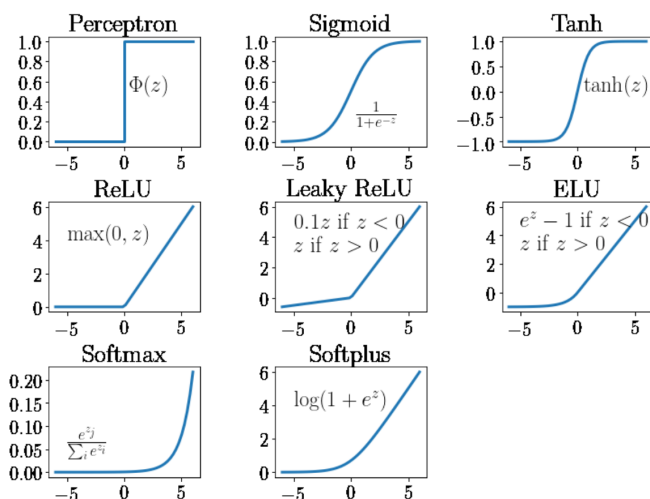


Figure 17: Activation functions in ANNs that introduce non-linearity [104]

Other common activation functions are shown in Figure 17, such as Perceptron, Leaky ReLU, Exponential Linear Unit (ELU), Softmax and Softplus, which are employed in multi-class classification.

4.4.2. Feedforward Neural Networks

Feedforward NNs [73] are the backbone of deep learning and have been successfully applied to a wide variety of tasks, including image classification, speech recognition, sentiment analysis, and the detection of fraud, among many other applications. Because of their capacity to learn hierarchical representations and model complicated relationships, they are a versatile tool that may be applied across a variety of areas.

Feedforward NN consists of an input layer, one or more hidden layers, and an output layer (see Figure 18). The network has as many neurons in the input layer as input variables (n) and, for classification, as many output neurons as there are classes in the data (m). A neuron is connected to all neurons in the two adjacent layers via a weighted connection (w).

The input layer is the one that is responsible for taking in the initial data; the hidden layers are the ones that are responsible for performing computations and extracting features; and the output layer is the one that is responsible for producing the final result. The NN gets its name from the fact that information flows in a forward direction, which is why the network is called a feedforward NN. The data travels through the network in a layer-by-layer fashion. The output of one layer is used as the input for the next layer until it reaches the final output layer.

In a feedforward NN, every neuron in one layer is connected to every neuron in the layer below it. The weight of the connections between neurons is also determined by the layer below them. These connections each have their own associated weights, which serve to define the connection's overall strength. The

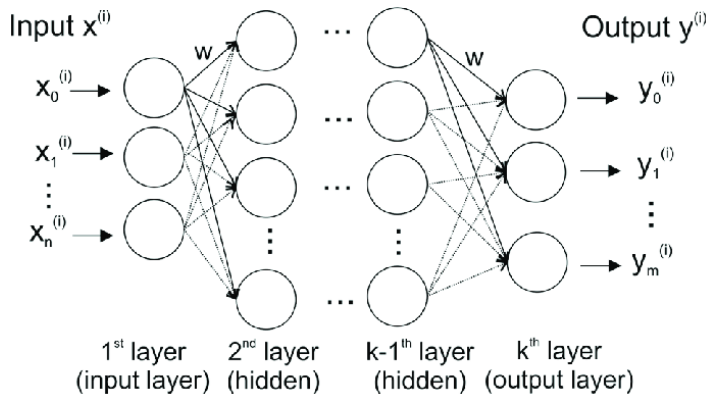


Figure 18: NN with k -layers consisting of fully connected neurons [105]

network will make adjustments to these weights while it is being trained so that it can learn and generalize patterns from the input data. The computations performed by individual neurons inside a feedforward NN are greatly aided by the utilization of activation functions. They bring in non-linearities, which enable the network to model complex interactions and make decisions that are non-linear.

The process of training a feedforward NN is an iterative process that combines forward propagation and backpropagation. While the network is performing forward propagation, it is also making predictions based on the training data. Backpropagation [100], on the other hand, is used to calculate the gradients of the network's parameters with respect to a particular loss function. After that, gradient descent optimization is utilized to revise the weights and biases of the network using the calculated gradients.

Loss functions [106] are used to quantify the amount of deviation that exists between the outputs that were predicted by the network and the actual labels or targets. They not only serve as a guide for the training process but also as a measurement of how well the network is working. The mean squared error (MSE) is a common loss function that is used for regression problems, and the categorical cross-entropy is typically used for classification problems. During the training process, feedforward neural networks are prone to overfitting, which is when they memorize the training data rather than generalizing to data they have not seen before. To prevent this from happening, regularization strategies [107] are used, including L1 and L2 regularization, dropout, and early halting.

4.4.3. Backpropagation Algorithm

In order to train a NN, a method is needed to assess the difference between the expected outputs of the network and the actual labels or targets being sought. This requirement is satisfied by backpropagation, which provides a methodical and effective means of training deep models and enables a NN to learn new knowledge from previously seen information by making adjustments to its weights and biases in accordance with the gradients of a predetermined loss function.

The backpropagation algorithm uses the chain rule to calculate the gradients layer by layer. Gradient descent is another method for calculating gradients. It does so by passing on the error from the output layer to the layers that came before it, providing information regarding the manner in which each parameter contributes to the overall loss. After that, gradient descent optimization is used to make adjustments to the network's weights and biases in accordance with these gradients, with the goal of gradually enhancing the prediction accuracy of the NN.

The backpropagation process starts with a forward pass, which is when the input data is transferred through the network in order to generate predictions. After computing the weighted total of its inputs, each neuron then applies an activation function before sending its output on to the subsequent layer. The loss function provides a numerical representation of the degree to which the projected outputs deviate from the actual labels or objectives. The error is computed by making a comparison between the network's predictions and the actual data using the loss function that was selected. During the backward pass, the error is sent through the network in the opposite direction of the forward pass, beginning at the output layer. The chain rule is used to construct gradients for each parameter, and these gradients indicate the direction and magnitude of change that must occur in order to minimize loss. After that, the obtained gradients, are utilized to revise the weights and biases.

The gradients of each neuron's output are computed as the backpropagation process is carried out. The activation function that is used has an effect on the shape that these gradients take, which in turn can have an effect on the learning process. Some activation functions, like the sigmoid function, are affected by the vanishing gradient problem, which makes it difficult for deep NNs to learn new information. This problem can be remedied by utilizing alternative activation functions such as ReLU or Leaky ReLU.

Backpropagation frequently makes use of an optimization process known as Stochastic Gradient Descent (SGD). Instead of calculating the gradients and updating the weights for the full dataset all at once, SGD chooses at random a part of the data to execute the computations on. This subset of data is referred to as a batch, where batch size is an important hyperparameter in training NNs and is often determined by available hardware.

The backpropagation algorithm makes forward and backward passes on the training data in an iterative manner until a stopping requirement is satisfied, such as reaching a maximum number of iterations or achieving adequate performance. Once the stopping criterion is met, the iterative training process is complete. Through the use of this iterative procedure, the aim of the network is to gradually converge towards optimal weights and biases.

Overfitting is a common problem while training NNs; therefore, regularization strategies, such as L1 and L2 regularization, can be added to the backpropagation algorithm. These methods involve the addition of a regularization term to the loss function, which modifies the gradients and facilitates more robust learning.

4.4.4. Gradient Descent Optimization

During the process of training a NN, the gradient descent optimization method is a basic approach that is used to update the weights and biases of the NN based on the gradients that are generated during the process of backpropagation (see visualisation in Figure 19). Using optimization strategies such as learning rate modification and adaptive methods, as well as versions of gradient descent such as batch gradient descent and SGD, gradient descent optimization makes it possible for a network to learn and converge on the best possible solution.

Optimization. Optimization in NNs refers to the process of locating the optimal collection of weights and biases that minimize the loss function of the network. In NNs, one of the most commonly used optimization algorithms is called gradient descent. Its purpose is to iteratively update the parameters so that they point in the direction of the steepest fall. The fundamental principle underlying gradient descent is to update the parameters in the direction that is opposite to the gradients of the loss function. During each iteration of the batch gradient descent, the gradients are computed and the parameters are updated using the full training dataset. Due to the necessity that all of the samples need to be processed at the same time, it might be computationally expensive for large datasets, despite the fact that it is accurate.

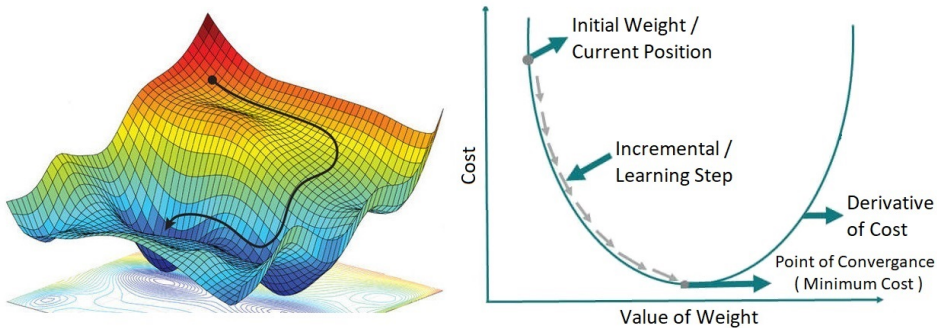


Figure 19: Gradient descent visualisation [108]

Stochastic Gradient Descent. SGD overcomes the computational inefficiencies of batch gradient descent by randomly selecting a single training sample or a mini-batch of data to compute the gradients and update the parameters (see Figure 20). This strategy, which involves adding more randomness to the mix, has the potential to result in more rapid convergence as well as improved generalization.

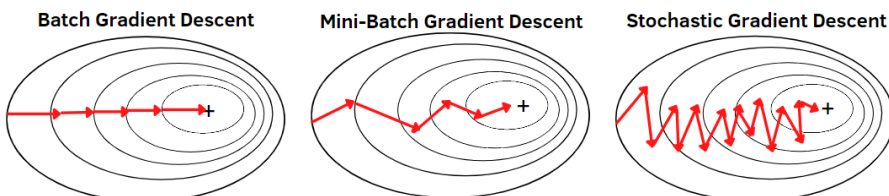


Figure 20: Gradient descent types [109]

Learning Rate Scheduling. The learning rate determines the size of the steps that are made when the parameters are updated. When the learning rate is low, this can lead to slow convergence, whereas when the learning rate is large, this can induce unstable convergence or overshooting. It is absolutely necessary to fine-tune the rate of learning in order to locate the optimal equilibrium. Stabilizing convergence and fine-tuning the learning process can be accomplished through the use of strategies such as slowing down the rate of learning over time (learning rate decay) or adjusting it based on the amount of loss improvement.

Momentum Optimization. Momentum optimization presents the idea of "momentum" in order to speed up the gradient descent process (see Figure 21). Even when there are only a few gradients or there is a lot of noise in the gradients, it remembers the gradients from the past and uses them to keep traveling in the right direction. This can result in faster convergence as well as an improved ability to handle local minima. Furthermore, adaptive learning rate methods, such as AdaGrad, RMSprop, and Adam, dynamically alter the learning rate based on the observed gradients. These techniques seek to strike a compromise between rapid initial learning and more nuanced modifications as training develops, with the end goal of offering optimization that is both efficient and successful.

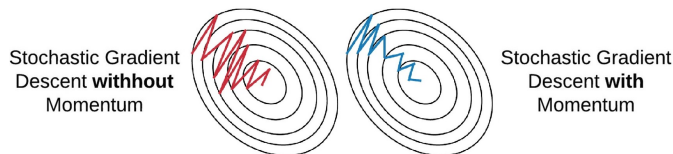


Figure 21: Gradient descent optimization using momentum [110]

Convergence. Convergence in optimization refers to the moment at which the performance of the network reaches a plateau and additional updates to the parameters offer limited progress. The criteria for stopping optimization at this point are known as the stopping criteria. When it comes to determining when the training process should be terminated, stopping conditions, such as reaching a maximum number of iterations or noticing a slight change in the loss function, are quite helpful. Figure 22 shows an example of an optimum stopping condition that is based on diverging training and validation loss.

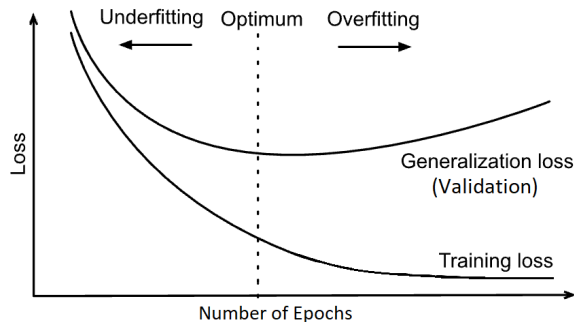


Figure 22: Validation loss and converging training loss [111]

4.4.5. Single Layer Perceptron

One of the basic components of NN is the Single Layer Perceptron (SLP), which consists of a single layer of artificial neurons. Each neuron takes in inputs, gives those inputs weights, and then generates an output. The output is created by multiplying the inputs by their respective weights, then running the weighted sum through an activation function [112]–[114].

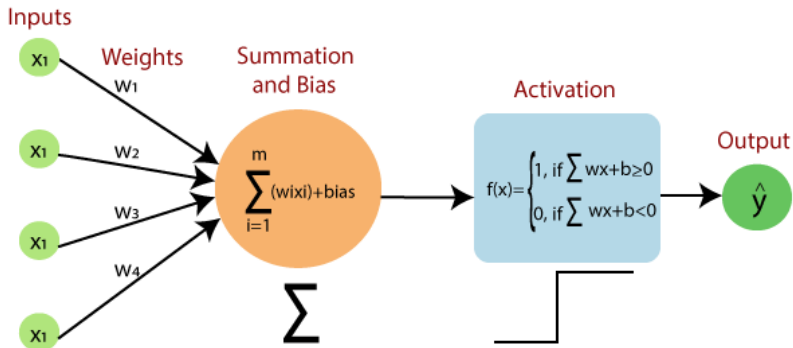


Figure 23: Single layer perceptron with one neuron [115]

Perceptron consists of the following parts (see Figure 23):

- *Input*: Numerical values from a dataset.
- *Weight*: The strength of the connection between units. The higher the weight, the more influence an input has on the neuron.
- *Bias*: An additional parameter to modify the output along with the weighted sum of the input with respect to other neurons.
- *Net sum*: Total sum of inputs multiplied by weights.
- *Activation function*: A rule for activating or not activating the neuron.

To reduce the classification error, the SLP is trained by modifying the weights based on training data. Weight updates are performed using the perceptron learning rule, referred to as the delta rule or the Widrow-Hoff rule. By comparing the perceptron’s projected output with the desired output, the learning rule iteratively changes the weights. No weight update is carried out if the prognosis is accurate. However, if the prediction is wrong, the weights are changed to get the prediction more in line with the intended result.

Numerous restrictions on the SLP prevent it from solving complex issues. Its inability to learn non-linear decisions is a serious drawback. The perceptron can only identify linearly separable data since it employs a linear activation function. Therefore, the perceptron will not reach a solution if the data points cannot be separated linearly. SLPs cannot address more complicated problems that call for hierarchical representations.

4.4.6. Multi Layer Perceptron

The Multi Layer Perceptron (MLP) [116] is a potent and popular technique for supervised learning tasks in the field of ANNs. MLPs are feedforward NNs made up of many interconnected layers of neurons. Due to their capacity to learn intricate patterns and relationships in data, MLPs have found use in a variety of fields, such as image identification, natural language processing, and financial forecasting.

An MLP architecture is made up form an input layer, one or more hidden layers, and an output layer. Multiple neurons make up each layer, which are linked together via weighted connections. Each neuron in the input layer corresponds to a particular feature, and the input layer receives the features or inputs of the dataset. The hidden layers process the data from the input layer by using activation functions to spot intricate patterns. Finally, based on the data that has passed through the hidden layers, the output layer generates the desired predictions.

In an MLP, forward propagation includes feeding input data into the network to provide predictions. Each neuron in the hidden layers and the output layer performs a weighted sum of each neuron's inputs before applying an activation function. The weighted sum is calculated by multiplying the input numbers by the appropriate weights. The network gains non-linearity from the activation function, which enables it to simulate intricate data interactions. The sigmoid function, the hyperbolic tangent (tanh) function, and the ReLU function are often used activation functions in MLPs.

In the backpropagation algorithm, which consists of forward propagation and backward propagation operations, the projected outputs are generated as the input data is sent through the network during forward propagation. Gradients are calculated in backward propagation layer by layer, starting from the output layer and moving back towards the input layer. Then, using an optimization process like gradient descent or one of its variants, the weight adjustments are carried out depending on these gradients.

Many changes and improvements have been suggested over time to increase MLPs functionality and performance. Several prominent versions include:

- Convolutional neural networks (CNNs)
- Recurrent neural networks (RNNs)
- Networks with long short-term memory (LSTM)

4.4.7. Convolutional Neural Network

CNNs have brought about a revolution in the field of computer vision and have established themselves as the method of choice for many computer vision-related problems, including object detection, picture classification, and image segmentation [117]. CNNs are a specific sort of NN built for processing grid-like input, such as images and sequential data. CNNs can be thought of as a hybrid between traditional NNs and recurrent NNs. Due to the fact that traditional NNs are fully connected, which results in a huge number of parameters and inefficient learning,

it can be difficult for these networks to interpret grid-like input in an effective manner. CNNs get over this issue by making use of shared weights and local connectivity, which enables them to capture spatial and temporal patterns in data and hierarchical representations. As CNNs are an important part of the thesis, the topic is described in more detail, and an in depth example case of a revolutionary CNN is provided, namely AlexNet.

The structure of a CNN is made up of several layers, the most important of which are the convolutional layers, the pooling layers, and the fully connected layers. By applying convolution operations, the convolutional layers are in charge of extracting local features. The pooling layers are in charge of downsampling the feature maps, and the fully connected layers are in charge of classification or regression based on the extracted features.

In order to determine the local characteristics of the input data, convolutional operations require the application of filters (kernels). These filters are small matrices that are moved over the input to perform computations such as element-wise multiplications and summations. The final product is a feature map that depicts important patterns that were discovered within the data.

CNNs take advantage of local connectivity, which means that each neuron in a layer is connected to a small part of the layer below it. Shared weights are also utilized in CNNs. Because of this local connectedness, the number of parameters in the network is reduced, which enables the network to learn particular spatial or temporal patterns.

The architecture of a typical CNN has multiple layers that are stacked on top of one another in a sequential fashion. These layers are as follows:

- **Input Layer:** The input layer is the layer that takes in the image to be layered and then sends it on to the following layers.
- **Convolutional Layers:** Convolutional layers collect local characteristics from the input by executing element-wise multiplications and then aggregating the results of those multiplications. These layers are responsible for capturing low-level characteristics, including edges, and textures.
- **Pooling Layers:** Pooling layers help to reduce the computational complexity of the feature maps by reducing the spatial dimensions of the feature maps. This helps to capture the most important information while also minimizing the amount of data that needs to be processed.
- **Activation Functions:** Non-linear activation functions, such as ReLU, bring non-linearities into the network. These non-linearities provide the network with the ability to learn complicated correlations of the input data.
- **Fully Connected Layers:** Fully connected layers allow the network to learn high-level representations and generate predictions by connecting every neuron in one layer to every neuron in the succeeding layer.
- **Output Layer:** The output layer is responsible for providing the probability or class predictions that are based on the learned features.

Deep CNN designs, such as AlexNet [118], VGGNet [119], [120], GoogLeNet [121], and ResNet [122], have had an incredible amount of success in a variety of computer vision problems. In these architectures, there are often a lot of convolutional and pooling layers stacked on top of each other. This lets the network learn hierarchical representations and pick up on features that get more complicated.

Here are a few examples of CNNs in computer vision tasks:

- **Object Recognition:** CNNs excel at distinguishing and classifying objects, enabling applications such as object detection, and facial recognition.
- **Fine-Grained Classification:** Because CNNs are able to discern between tiny changes within the same object category, they are useful for doing tasks such as the classification of species in the context of wildlife monitoring.
- **Image Captioning:** CNNs and RNNs are used in conjunction with one another to generate descriptive captions for images, which increases both the accessibility and the level of comprehension of visual content.
- **Medical Imaging:** CNNs are utilized in the process of evaluating medical images for a variety of tasks, including the detection of tumors, the classification of diseases, and the segmentation of organs.

CNNs, in spite of their widespread use and success, continue to struggle with problems such as overfitting, restricted interpretability, and the requirement of vast quantities of labeled data. In order to solve these issues and improve CNNs' capabilities, researchers are actively investigating a variety of methodologies, such as regularization methods, attention mechanisms, interpretability approaches, and leveraging models trained on large datasets using transfer learning.

CNN Case Study - AlexNet. The science of computer vision was completely transformed by the groundbreaking CNN architecture known as AlexNet [118], created by Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton and won the 2012 ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [123], which led to a breakthrough in image classification.

Several important aspects that were incorporated into AlexNet's architecture made it successful. The following are AlexNet's main elements (see network architecture in Figure 24 and structural details in Table 2):

1. **Convolutional Layers:** AlexNet has five convolutional layers, with a max-pooling operation coming after each layer. These layers identify low to high level visual elements by convolveing input images with learnable filters.
2. **ReLU:** ReLU activation functions are added to the network after each convolutional layer to make it less linear. ReLU activation facilitates improved representation learning by assisting the network in modeling complex relationships within the data.
3. **Local Response Normalization:** Following the first and second convolutional layers, AlexNet uses Local Response Normalization (LRN) layers. By averaging output from neighboring channels, LRN improves generalization and the network's capacity to distinguish between various features.

4. Dropout: Dropout minimizes overfitting and enhances generalization by setting a portion of the neuron outputs to zero at random during training. Dropout is used after AlexNet's completely connected layers.
5. Fully Connected Layers: Three fully connected layers and a softmax activation for class probabilities make up AlexNet's final structure. Based on the extracted features, these layers carry out classification.

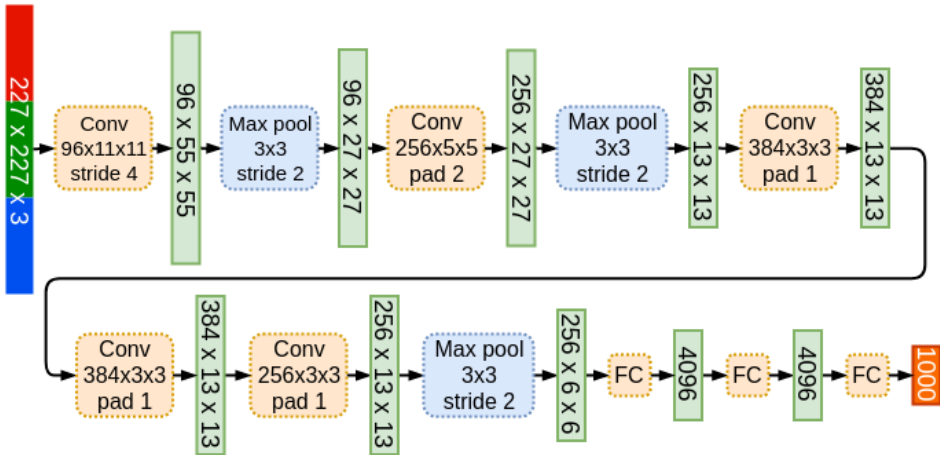


Figure 24: AlexNet block diagram [118]

Table 2: AlexNet structural details [118]

AlexNet Network - Structural Details													
Input			Output			Layer	Stride	Pad	Kernel size		in	out	# of Param
227	227	3	55	55	96	conv1	4	0	11	11	3	96	34944
55	55	96	27	27	96	maxpool1	2	0	3	3	96	96	0
27	27	96	27	27	256	conv2	1	2	5	5	96	256	614656
27	27	256	13	13	256	maxpool2	2	0	3	3	256	256	0
13	13	256	13	13	384	conv3	1	1	3	3	256	384	885120
13	13	384	13	13	384	conv4	1	1	3	3	384	384	1327488
13	13	384	13	13	256	conv5	1	1	3	3	384	256	884992
13	13	256	6	6	256	maxpool5	2	0	3	3	256	256	0
						fc6			1	1	9216	4096	37752832
						fc7			1	1	4096	4096	16781312
						fc8			1	1	4096	1000	4097000
Total												62,378,344	

AlexNet had a significant impact on the field of computer vision. It highlighted the value of CNNs and showed how deep learning can be effective in image classification applications. Due to the success of the ILSVRC 2012 competition, CNNs gained popularity for solving computer vision tasks, and the design served as an inspiration for numerous future CNN architectures.

4.4.8. Transfer Learning

Transfer learning with NNs is a technique that uses a pre-trained model that was trained on large-scale datasets as a starting point for a new task in another activity or area. The pre-trained models are able to achieve good performance despite having insufficient training data because they leverage the features that they have learned. Approaches that are frequently utilized in transfer learning include fine-tuning, freezing certain layers, and making use of pre-trained models as feature extractors. Transfer learning provides a number of benefits for a variety of ML tasks, including the following:

- **Quicker Training and Convergence:** Transfer learning greatly cuts down on the amount of time spent training and significantly speeds up convergence on the target task. This is accomplished by initializing the model using pre-trained weights and architectures.
- **Improved Generalization:** Transfer learning helps models generalize better by exploiting the knowledge that is learned from a larger dataset.
- **Overcoming Data Limitations:** Transfer learning permits the application of pre-existing knowledge, which successfully addresses the issue of data limitations in situations when labeled data is difficult to come by or requires significant financial investment to acquire.
- **Powerful Feature Representations:** As pre-trained models are often trained on enormous datasets, they have learned useful feature representations. Transfer learning enables the model to make use of previously learned features, which represent high-level representations and general patterns.

Applications of transfer learning include the following:

- **Feature Extraction:** In this technique, by removing the output layer, the pre-trained model serves the purpose of a fixed feature extractor, and the extracted features are used as input to other ML algorithms.
- **Shallow tuning:** The first layers of the model, which are responsible for learning low-level features, are kept, while the later layers of the model, which are particular to the source task, are swapped out for layers that are relevant to the target task. The weights of the pre-trained model are preserved (frozen), and the only weights that are modified during training on the target task are those of the newly added layers.
- **Deep tuning:** This step entails tailoring the pre-trained model to the task at hand by changing the output layers and revising the weights of the entire model. The earlier layers of the model can be fine-tuned with a slower learning rate, while the latter layers of the model are updated more aggressively to adapt to the target task.

Transfer learning is a prominent option in the field of medical imaging since medical datasets are frequently limited in size due to privacy issues, the cost of data collection, and the difficulty of collecting labelled data.

5. DATASETS

This chapter will provide more details about the two datasets used in this thesis. The first dataset is publicly available data from ADNI and is used for the MRI studies, which was obtained by submitting a request to ADNI. The second, not publicly available dataset, containing EEG recordings, was obtained from Tallinn University of Technology under contractual agreement.

5.1. Alzheimer’s Disease Neuroimaging Initiative MRI Dataset

Data used in this thesis were obtained from the Alzheimer’s Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). The ADNI was launched in 2003 as a public-private partnership, led by Principal Investigator Michael W. Weiner, MD. The primary goal of ADNI has been to test whether serial magnetic resonance imaging (MRI), positron emission tomography (PET), other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer’s disease (AD).

The ADNI database is held within the Image and Data Archive (IDA) and forms part of the Laboratory of Neuroimaging (LONI) [124]. The LONI gathers studies and clinical trials in the field of neuroscience and shares this data for research purposes. IDA fosters collaboration among neuroscientists around the world by managing data collection. ADNI utilizes biomarkers to produce clinical images and track patients’ cognitive health, and it supports the early diagnosis of AD by sharing data with researchers worldwide. Biomarkers can be employed to determine the progression of a disease. ADNI aims to facilitate research in aid of the prevention, treatment, and early detection of AD and to provide neuroscientists across the globe with data to support their research.

Within ADNI study [124], the process of MRI data collection took place between 2004 and 2016 and comprised four stages: ADNI-1, GO, -2, and -3. Participants either continued through each phase of the ADNI study and underwent further examination, or new participants were added to facilitate the investigation into the progression of AD. Various biomarkers were quantified throughout each phase. Measurements were taken during clinical trials, and functional imaging techniques were used; early signs of the disease were examined, and patients’ rates of decline were predicted.

In the course of the ADNI study [125], four different stages of AD’s progression were assessed: Cognitively Normal (CN), Significant Memory Concern (SMC), Mild Cognitive Impairment (MCI), and Alzheimer’s Disease (AD). Participants at the CN stage acted as control subjects and did not present with symptoms of dementia. Participants in the SMC category were added during ADNI 2, and they presented with minor memory issues that had the potential to worsen. Participants in the MCI category experienced self-reported or clinically identified

memory issues but maintained normal daily activities. At this stage, they did not present with signs of dementia. Finally, participants in the AD category were identified as affected by dementia.

5.1.1. Magnetic Resonance Imaging Correction Methods

Each of the magnetization-prepared rapid gradient-echo (MP-RAGE) images sourced from the database are accompanied by details about the pre-processing of each MRI image. In accordance with the information provided on the ADNI website, the MRI images used within this study have been corrected using the following methods [126]:

1. Gradwarp is a term for a system-specific way to fix the way gradient non-linearity messes up the shape of an image. Depending on the gradient model, the extent to which an image is distorted varies. It is likely that most users would rather use images that have undergone this correction process within their analyses.
2. B1 non-uniformity describes a correction procedure that utilizes B1 calibration scans to correct non-uniformity within image intensity. This non-uniformity occurs when radio frequency transmission is performed with a uniform body of coil and reception is performed with a coil of greater irregularity.
3. N3 describes an algorithm used to sharpen histogram peaks. It is applied to every image after grad warp and B1 correction have been carried out. As a result of the dielectric effect at 3T, N3 lowers the intensity of non-uniformity. In addition, 1.5T scans undergo N3 processing to limit the incidence of residual intensity non-uniformity.

Demographic data relating to participants' age, gender, and diagnostic category are also held within the ADNI dataset. Each participant within the study was aged 55-90, and originated from the United States and Canada [127]. Participants initially underwent a series of tests that are to be repeated each year. These include clinical evaluations, neuropsychological tests, genetic testing, lumbar punctures, and MRI and PET scans. Screening is carried out yearly according to the following intervals (see Figure 25): Screening, Baseline, months 3, 6, 12, 18, 24, 36, and 48. Participants are also to be provided with an annual check-up.

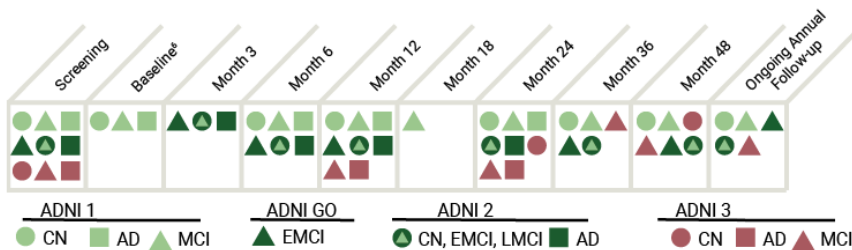


Figure 25: An overview of data collected throughout the ADNI study [125]

5.2. Tallinn University of Technology EEG Dataset

5.2.1. Electroencephalogram Recording Procedure

Tallinn University of Technology (TalTech) was responsible for collating the EEG recordings and questionnaire responses from consenting participants. This was conducted in accordance with the Declaration of Helsinki, and all procedures were formally approved by the Tallinn Medical Research Ethics Committee.

To collect EEG recordings, Cadwell Easy II EEG (Kennewick, WA, USA) measurement equipment was employed, and 18 channels (reference Cz) were placed on the participant's head in line with the international 10–20 system. Participants were required to adopt a relaxed position with their eyes closed as recordings were taken. Additional processing was applied to EEG signals with a frequency of 3–48 Hz. The EEG recording was divided into 10-second segments, and the first 30 artifact-free segments (a total of 5 minutes) were inspected by an EEG specialist in order to assess the segment quality.

5.2.2. Hamilton Depression Rating Scale

In spite of the Hamilton Depression Rating Scale (HDRS) use being criticized in clinical practice, HDRS is the most extensively applied clinician-administered depression assessment scale. Within this study, HDRS was employed to gather further information from which healthy participants could be selected. Participants with the lowest HDRS scores were chosen in instances where multiple healthy participants could act as age and gender matched candidates for a depressive subject. Among healthy participants, the mean HDRS score was 3.1; HDRS scores between 0–7 denoted the absence of depression; and, among the depressive subjects, a mean HDRS score of 9.3 indicated mild depression.

5.2.3. Emotional State Questionnaire

The Emotional State Questionnaire (EST-Q) [128] was designed for lecturers in the psychiatric clinic at the University of Tartu, Estonia. This self-assessed questionnaire detects depression and anxiety disorders and their corresponding symptoms using 28 statements. The questionnaire evaluates patients' wellbeing over a month and involves three basic scales and three additional scales. The basic scales consist of depression (DEP), general anxiety (AUR), and panic agoraphobia subscales (PAF). The additional scales encompass social anxiety (SAR), asthenia (AST) and insomnia subscales, which were not used. The extent of a participant's emotional symptoms can be assessed using the scale's total score. In the present study, the EST-Q was used to determine the control group. Excluding two healthy subjects, every participant's subscale values were lower than the threshold for each condition. For the two aforementioned subjects, their asthenia subscales were higher than 6. If the scale value is higher than the threshold, the participant is affected by the condition. All threshold values are presented in Table 3.

Table 3: Information about sex, age, HDRS, and EST-Q test results for subjects in the dataset, where in each table cell the number before the back slash is for healthy subjects and on the right of the back slash for people who have had depression (Healthy/Depressed)

Sex	Age	HAM-D >7	EST-Q				
			DEP >11	AUR >11	PAF >6	SAR >3	AST >6
M	25/24	7/8	5/13	7/0	0/0	0/2	7/5
F	33/34	1/9	1/5	3/12	0/0	0/0	2/5
F	35/35	6/21	7/16	8/11	2/0	1/2	4/10
M	36/35	1/4	4/4	8/3	0/1	1/0	3/0
F	35/37	3/13	4/8	6/11	0/0	1/0	2/13
F	39/38	0/9	3/4	7/17	1/2	0/0	3/11
M	40/43	1/6	3/8	2/9	0/0	0/0	2/7
F	46/46	5/12	6/8	3/2	0/0	0/1	1/3
F	48/48	0/8	2/4	8/7	1/0	0/2	5/6
M	60/53	7/3	5/11	8/8	0/0	1/0	9/8

5.2.4. Participant Selection

Data was collected from the EEG signals of 55 participants who made regular visits to their occupational health doctor. Data from 14 females and 6 males (20 in total), aged 24-60, were analyzed in further detail. In the control group, none of the participants had previously been diagnosed with depression. Contrastingly, half of the remaining participants had been affected by clinical depression at some point in their lives. For clarity, these participants are referred to as depressed subjects in the EEG study described in chapter eight. Given their low HDRS and EST-Q scores, the healthy control group was chosen by ensuring that no symptoms of depression or other mental illnesses were present (see Table 3). The age of the healthy control group was aligned as closely as possible with the age of the depressed subjects, and for each matched participant pair, the gender was the same.

6. VOLUMETRIC HISTOGRAM BASED ALZHEIMER'S DISEASE DETECTION

6.1. Introduction

Without a cure, AD is becoming an increasingly common disease. There is currently a far-reaching engagement taking place with issues impacting early diagnosis with the intention of enhancing the quality of life of patients suffering from AD. Current research on the early diagnosis of Alzheimer's disease has made use of current advancements in computational technology and ML approaches. It is possible that the application of pattern recognition and ML will assist medical professionals in making an earlier diagnosis based on biomarkers that suggest the development of Alzheimer's disease. This study makes use of the ADNI dataset, which contains MRI brain scans and which can be analyzed statistically using ML procedures. Within the context of MRI brain scans, the purpose of this investigation is to evaluate the application of a variety of ML methodologies in the classification of AD and CN. This work may promote the improvement of classification processes and facilitate an earlier diagnosis of AD utilizing computer vision and ML by employing supervised learning. In addition, medical experts may use these instruments while determining whether or not a person is showing signs of developing AD. Throughout the course of this research, a number of different ML based algorithms, such as the SVM, LR, DT, and Fine Tree (FT), are evaluated and compared. The representation of MRI slices in the form of a histogram is also used as a method for managing the dimensionality of the data.

6.2. Methodology

Three-dimensional MRI brain scans have been used within this study and can be observed from the axial, sagittal, and coronal planes, as shown in Figure 26.

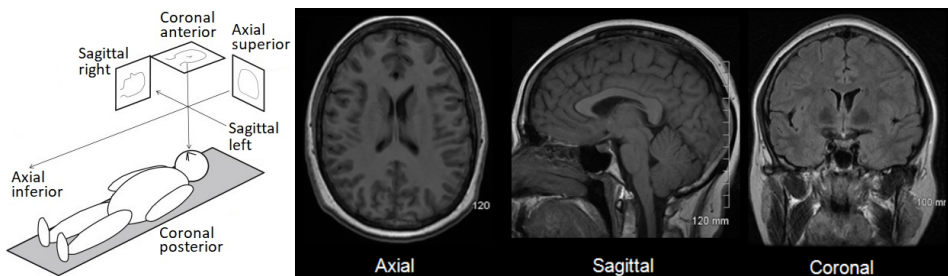


Figure 26: MRI slices of the axial, sagittal, and coronal planes [129]

6.2.1. Dimensionality Reduction

ML typically involves a large datasets and can be problematic because of the high dimensionality of the data therein. Therefore, dimensionality reduction is

applied in order to simplify, quicken, and allow large data samples to be analyzed. Feature selection and extraction can be employed during dimensionality reduction [17]. By applying this process, computation time and required storage space can be reduced. However, using dimensionality reduction can result in the loss of important information within data.

Data sourced from the ADNI dataset has a significantly high dimensionality and needs to be processed. Consequently, in this study, multiple dimensionality reduction processes must be utilized in order to process and analyze the required data. Participants involved in the study, whether AD or CN, each had a corresponding .nii file within the ADNI dataset. The MRI images for each participant depict a voxel grid that comprises the three-dimensional structure of the brain. Each participant’s brain was represented by approximately 5 million voxels in the form of 192x192x160. Given this, the involvement of multiple participants produces vast, time-consuming, and computationally expensive data. Consequently, dimensionality reduction must be applied in order to carry out this research.

6.2.2. MRI Slice Histograms

Histograms represent continuous data by depicting the distribution frequency [130], [131]. Distribution is portrayed using a bar plot, and the data are separated into bins [130], [132], [133]. According to one’s requirements, the number of bins can vary. By applying a histogram, the dimensionality of the database could be reduced. The intensity value of pixels within an 8-bit image range between 0 and 255, therefore the number of bins was chosen according to the features included in this study: 256 features, which corresponded to 256 bins.

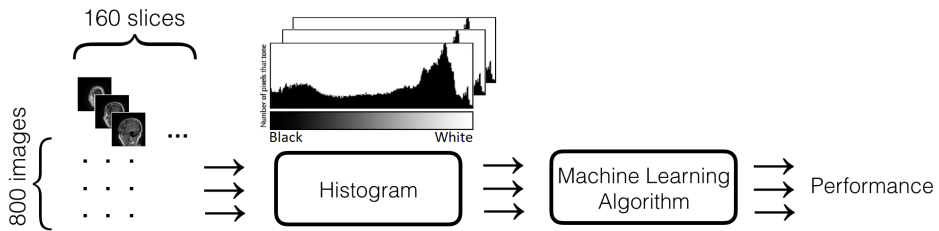


Figure 27: Flowchart of histogram based classification

In Figure 27, the stages involved in the use of histogram-based classification are depicted using a flowchart. The participants comprised 400 AD and 400 CN, from whom 800 images were gathered. The images sourced for each participant consisted of 160 slices. A histogram was extracted from the sagittal plane slice of all 800 images, and the performance accuracy of each batch was quantified using several ML algorithms. This process was repeated for the 160 slices of the sagittal plane, where slice 1 corresponds to the most left slice and slice 160 to the most right slice of the brain. Before histogram extraction, each slice was normalized and converted to the [0, 255] value range.

6.3. Results and Discussions

Three-dimensional images of participants' sagittal planes were sourced from a total of 800 images that incorporated 400 AD and 400 CN participants. A total of 256 features representing 8-bit image intensity value count were used for one slice, which corresponded with the number of bins applied to produce histograms of the images. The performance accuracy of all 160 slices was determined, and several algorithms were compared. These included FT, LR, Linear SVM, Quadratic SVM, Cubic SVM, Fine Gaussian SVM, Medium Gaussian SVM, Coarse Gaussian SVM, Fine KNN, Medium KNN, and ensemble Boosted Trees. The dataset was divided according to a 75:25 (train:test) split ratio for the training and test sets, with a balanced CN and AD sample count in each set.

Slices 31 to 40 are compared in Table 4, and emboldened values indicate the highest accuracy achieved by each classifier. Slices 36 and 37 achieved the highest accuracy, according to 8 out of 11 classifiers.

Table 4: AD and CN classification performance for slices from 31 to 40

Methods	Slice accuracy %									
	31	32	33	34	35	36	37	38	39	40
Fine Tree	58.2	56.0	54.0	54.4	55.0	51.2	52.4	54.4	53.8	51.7
Logistic Regression	48.8	48.8	46.6	50.1	52.4	54.1	51.7	52.1	49.5	52.4
Linear SVM	54.1	57.0	53.6	56.9	59.1	61.1	57.9	59.0	57.1	57.8
Quadratic SVM	59.8	60.9	55.6	60.5	58.6	61.8	62.0	57.8	60.0	59.2
Cubic SVM	58.6	57.5	58.8	61.0	59.0	63.5	62.1	58.9	57.8	59.9
Fine Gaussian SVM	54.0	55.0	55.0	56.1	55.1	52.5	56.1	57.2	56.0	52.2
Median Gaussian SVM	59.2	60.9	59.6	61.3	61.6	62.3	64.0	61.4	62.3	61.3
Coarse Gaussian SVM	54.8	56.8	56.4	56.2	57.6	58.9	58.0	57.0	56.6	56.0
Fine KNN	57.2	56.6	57.2	57.9	58.1	59.1	61.4	57.5	58.4	57.0
Medium KNN	57.1	57.6	58.0	55.5	57.8	62.1	58.6	59.2	59.8	55.6
Ensamble Boosted Trees	60.0	61.5	57.1	58.1	61.5	58.6	58.8	58.0	57.4	56.8

Each of the 160 slices maximum and minimum classification accuracy is plotted in Figure 28, as determined by multiple classifiers. Evidently, the slices closer to the beginning and the end have reduced accuracy and are therefore unsuitable for the process of identifying AD or CN. This is due to the fact that the brain is not entirely visible within these slices and only gradually appears. In Figure 29, slices 5, 10, and 15 of two participants are shown. This demonstrates the inaccuracy of the early slices, as the brain is not visible within the images. This is identical in slices taken near the end of the range.

According to Table 4, Medium Gaussian SVM achieved the highest result in the majority of slices. In Figure 30, the performance of three classifiers (as taken from the data presented in Table 4) is compared. The accuracy of the FT, Medium Gaussian SVM, and Medium KNN is shown for each of the 160 slices. As demonstrated, Medium Gaussian SVM returns the highest level of accuracy for the majority of points.

As mentioned, Table 4 shows that slices 36 and 37 achieved the highest rate of

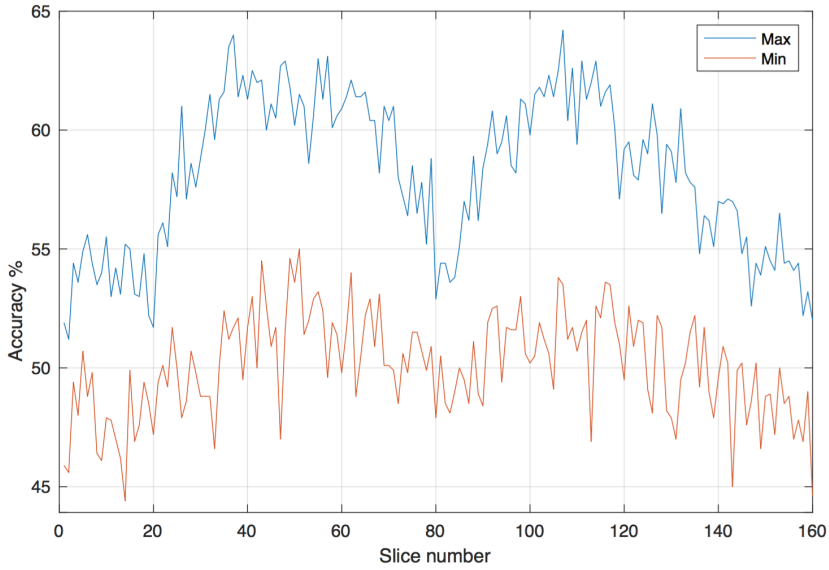


Figure 28: Maximum and minimum accuracy per slice for multiple classifiers

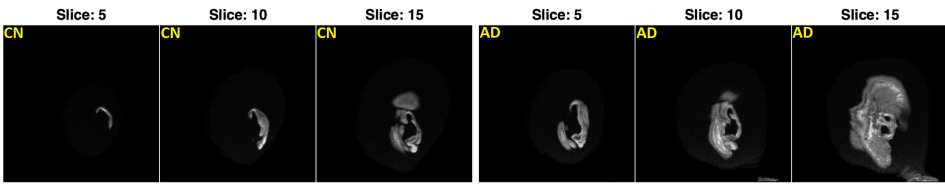


Figure 29: CN and AD image slices 5, 10, and 15

accuracy in classifying AD and CN. Therefore, these slices are considered further, and Medium Gaussian SVM, as applied in the trained model for slices 36 and 37, was used to test every other slice. The results of this are shown in Table 5. Upon testing the data of all 160 slices in accordance with the trained model of slice 36 and the Medium Gaussian SVM classifier, an accuracy rate of 70.75% was achieved for slice 35. This was the highest result. Respectively, for the trained model of slice 37, an accuracy of 70.87% was achieved for slice 36 as testing data.

Table 5: AD and CN classification performance for testing data for slices from 31 to 40 using trained model of slice 36 and 37

Median Gaussian SVM	Slice accuracy %									
	31	32	33	34	35	36	37	38	39	40
Trained on slice 36	60.50	62.13	65.50	68.38	70.75		69.25	69.00	67.75	65.75
Trained on slice 37	58.13	61.63	63.25	65.13	67.50	70.88		69.75	68.75	68.13

As slices 35 and 36 showed the highest performance from the trained models of slices 36 and 37, respectively, for Medium Gaussian SVM, 22 classifiers were used, and both majority and weighted voting were employed. Data from slice 36 was used to train a model, and predictions were made according to the data returned by slice 35. In light of the predictions made by 22 classifiers, majority voting was applied, and a rate of 74.12% accuracy was achieved. For weighted

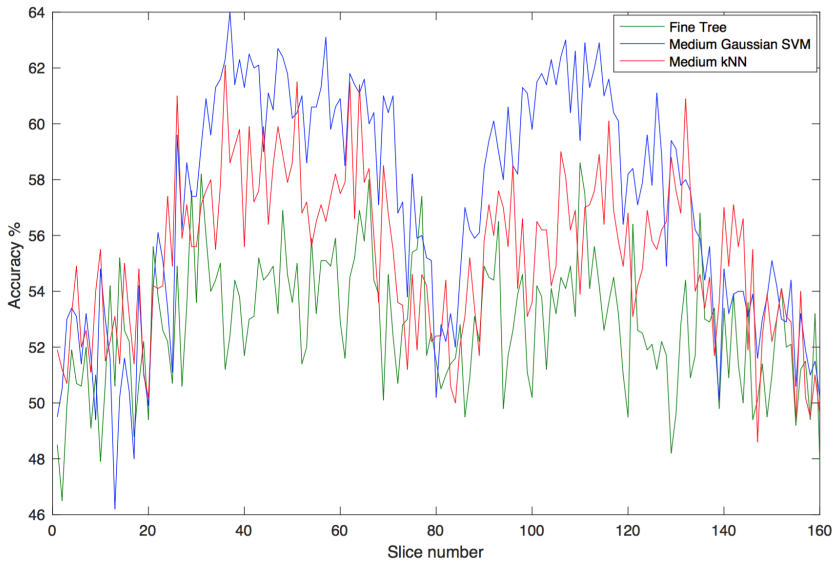


Figure 30: Slice accuracy for FT, Medium Gaussian SVM, and Medium KNN

voting, an accuracy rate of 74.75% was achieved. Subsequently, another model was constructed using training data from slice 37 and test data from slice 36. For majority voting, an accuracy of 74.62% was achieved; for weighted voting, the result was 75.87%. Figure 33 presents a comparison of the results ascertained from the majority vote and weighted vote, as produced by the test data of slices 35 and 36. The training data produced by slices 36, 37, and 38 (slice examples shown in Figure 32) presented the highest accuracy in their classification of AD and CN and were therefore investigated further.

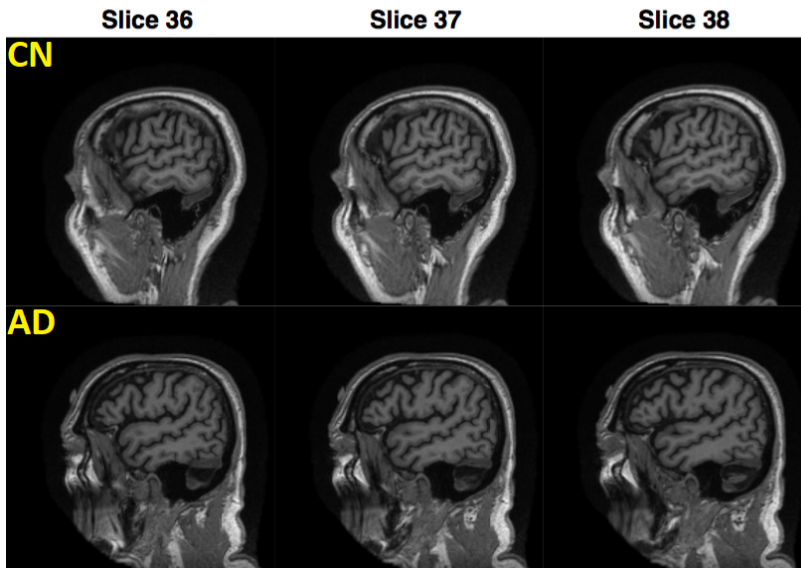


Figure 31: Image slices 36, 37, and 38 for CN (top) and AD (bottom row)

A matrix of 800 rows and 512 columns (which corresponds to the number of features for two slices) was produced using histograms of slices 36 and 37 for all 800 images. Every classifier was trained, and all 512 features were used. The model associated with each classifier was saved and used to test the data. Each prediction underwent majority voting, and these results were compared with the data labels. For slices 36 and 37 (see Figure 32), the majority vote resulted in an accuracy of 69.5%. This process was repeated to link slices 36, 37, and 38. For these slices, 768 features were involved, and the majority vote returned an accuracy of 64.5%. Each prediction also underwent weighted voting according to the accuracy of the trained model, and these results were compared. For slices 36 and 37, the weighted vote resulted in an accuracy of 67.5%. Meanwhile, an accuracy of 63.5% was achieved for slices 36, 37, and 38. A comparison of the results produced by the majority vote and the weighted vote—for slices 36 and 37, and slices 36, 37, and 38—is shown in Figure 33.

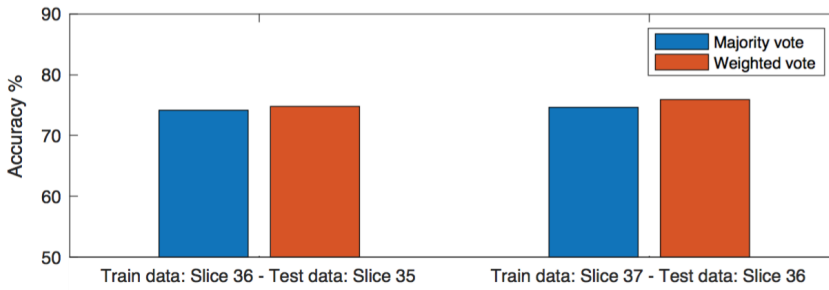


Figure 32: Accuracy for majority and weighted vote for trained model on slice 36 and 37

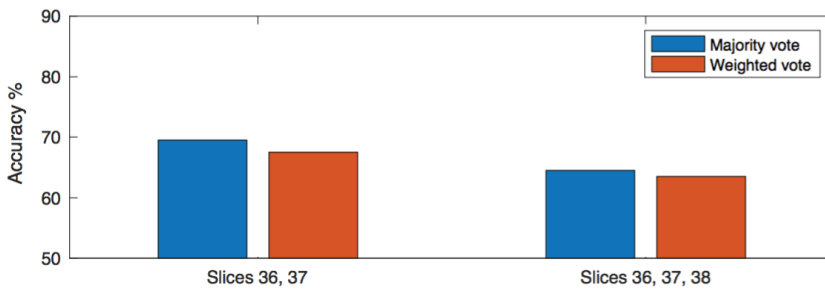


Figure 33: Accuracy for majority and weighted vote for concatenated histograms

7. ALZHEIMER'S CLASSIFICATION USING IMAGE TEXTURE DESCRIPTORS

7.1. Introduction

With the advancements in artificial intelligence and computing power, it is anticipated that new insights will be obtained regarding the diagnosis and treatment of various medical conditions. In relation to Alzheimer's Disease (AD) classification, the previous chapter focused on histograms constructed from 2D MRI slices representing the pixel value statistics to describe patterns in the brain. However, the concept of histograms has been applied in various algorithms such as LBP and HOG which are very popular histogram-based feature extractors and are often used in medical imaging. In the case of histogram descriptors, similarly to using statistical features or classical histograms, the derived features are combined with a variety of classification techniques to make predictions about the data at hand.

The histogram descriptors are referred to as handcrafted feature extractors, meaning that they are manually created methods. A more modern approach is to skip the feature engineering part and let the ML algorithm learn feature representations directly from the data. With enough data, ML methods such as NNs can be trained from scratch; another option is to use pre-trained NNs as feature extractors or retrain the pre-trained NNs with a new set of data, which is known as transfer learning.

This study uses MRI images from ADNI dataset and investigates histogram-based local descriptors as well as pre-trained NN feature extractions of popular CNN architectures. To keep the number of hyperparameters low and the results comparable, features from both feature extraction methods were classified using the same binary classification algorithms: SVM, LDA, QDA, DT, and KNN. Therefore, the main focus of this research is to investigate and show comparisons between histogram descriptors and pre-trained CNN feature extraction for the classification of AD. This study also uses the classification accuracy results to obtain insight into which MRI planes and slices contain the most pertinent data for AD classification. Besides the individually used slices, combinations of axial, coronal, and sagittal planes with top performing features are also investigated.

7.2. Features

Pattern recognition in medical imaging is the ability to recognise normal anatomical and physiological appearances on an image, as well as differences in appearance that may suggest pathology. To describe these differences, histogram descriptors are a very popular choice for local feature representation. On the other hand, methods such as CNNs learn hierarchical representations for the contents of the inspected images; therefore, with these two feature extraction methods, local and hierarchical feature representations are investigated in this study.

7.2.1. Histogram-based Local Descriptors

Local descriptors based on histograms have proven to be effective tools for feature extraction and image analysis [134]–[137]. These descriptors provide a condensed representation of image features with the capacity to recognize regional patterns and spatial information. Histogram-based local descriptors for MRI aid in image segmentation, registration, and classification by revealing local intensity variations and textural patterns.

Histogram-based local descriptors examine the local distribution of voxel intensities in a 2D MRI slice. They are frequently estimated within local neighborhoods or Regions-of-Interest (ROI). Changes in intensity, texture patterns, and spatial interactions between adjacent voxels are crucial characteristics.

Several advantages of histogram descriptors are summarized below:

- Insensitive to global image transformations such as translation, rotation, and scaling; they are resistant to viewpoint and image variations.
- Computationally efficient as they can be computed independently and concurrently.
- Adaptability to varying illumination conditions increases their resilience.
- Compact, easily stored, and suitable for large-scale image databases.

As mentioned before, the most widely used histogram-based local descriptors in medical imaging are LBP [138] and HOG. Besides the mentioned histogram descriptors, 14 additional descriptors were evaluated for brain MRI classification, bringing the total to 16 descriptors listed in Table 6. The credit for method implementation in MATLAB goes to Turan and Lam [136] who investigated local descriptor performance for facial expression recognition.

Table 6: Histogram-based local descriptors

Nr.	Abbreviation	Descriptor Name	Dim.
1	LBP [139], [140]	Local Binary Pattern	59
2	HOG [141]	Histogram of Oriented Gradients	6084
3	PHOG [142], [143]	Pyramid of HOG	168
4	LPQ [144]–[146]	Local Phase Quantization	256
5	LMP [147]	Local Monotonic Pattern	256
6	LTrP [148], [149]	Local Transitional Pattern	256
7	MBC [150], [151]	Monogenic Binary Coding	3072
8	LGIP [152]	Local Gradient Increasing Pattern	37
9	LDN [153]	Local Directional Number Pattern	56
10	GDP [154]–[156]	Gradient Directional Pattern	256
11	WLD [157], [158]	Weber Local Descriptor	32
12	LAP [159]	Local Arc Pattern	272
13	LGP [160]	Local Gradient Pattern	7
14	MBP [161]	Median Binary Pattern	256
15	LDTP [162]	Local Directional Texture Pattern	72
16	IWBC [163]	Improved Weber Binary Coding	2048

7.2.2. Feature Extraction Using Pre-trained Convolutional Neural Networks

CNNs are powerful computer vision tools that have been trained to extract meaningful features from images using large datasets like ImageNet [123], [164], [165]. These networks have learned to identify and classify objects in a variety of categories, making them useful for a wide range of tasks. Using the transfer learning technique, it is possible to use the knowledge and representations learned by these pre-trained CNNs as a starting point for new image classification or recognition tasks.

The central concept underlying the use of CNNs is that they can autonomously discover hierarchical image representations [166]. Earlier network layers capture low-level characteristics like edges and textures, whereas later network layers derive more complex and abstract characteristics like shapes and objects. These learned features can then be used to depict images in a compact and distinguishable manner, enabling efficient feature extraction or classification.

Due to the scarcity of data, the use of pre-trained CNNs and transfer learning is crucial in medical imaging. This study demonstrates how feature extraction using pre-trained CNNs can be used to classify brain MRI slices, potentially obviating the need for manually crafted feature descriptors. Table 7 lists 18 popular CNN models that have been tested in this study. The listed CNNs have been trained with the ImageNet dataset [167] to classify images into 1000 object categories, and the models were evaluated using MATLAB.

Table 7: Pre-trained convolutional neural networks

	Network	Depth	Param. (Millions)	Output dimensions
1	AlexNet [118]	8	62.3	4096
2	VGG16 [119]	16	138	4096
3	VGG19 [119]	19	144	4096
4	GoogLeNet [121]	22	7.0	1024
5	Inception-v3 [168]	48	23.9	2048
6	ResNet18 [169]	18	11.7	512
7	ResNet50 [169]	50	25.6	2048
8	ResNet101 [169]	101	44.6	2048
9	SqueezeNet [170]	18	1.24	1000
10	Inception-ResNet-v2 [171]	164	55.9	1536
11	Xception [172]	71	22.9	2048
12	DarkNet19 [173]	19	20.8	1000
13	DarkNet53 [173]	53	41.6	1024
14	ShuffleNet [174]	50	1.4	544
15	NASNet Mobile [175]	-	5.3	1056
16	NASNet Large [175]	-	88.9	4032
17	MobileNet-v2 [176]	53	3.5	1280
18	EfficientNet [177]	82	5.3	1280

7.3. Methodology

It is possible to use all MRI slices along each plane, but such an approach requires high end computer hardware that is able to process long feature vectors. Using one slice from one of the planes is the bare minimum approach. Thus, to compare the characteristics of the histogram descriptors and the pre-trained CNN features, slices from each plane were examined, namely axial, coronal, and sagittal.

The training dataset consisted of a subset of the ADNI dataset comprised of first-year participants scanned with a 1.5T MRI scanner. The MRI voxel grid sizes of the scans ranged between 192x192x160 and 256x256x184 voxels. To make use of all of the sizes, the MRI scans were resized to 224x224x224, where 224x224 was the most common image height and width size for the selected CNNs. The sides of MRI scans typically do not contain information about the brain; therefore, after resizing, slices numbered from 53 to 172 (in total, 120 slices) were selected for evaluation in each plane. Each slice was normalized in accordance with its minimum and maximum value and scaled to the [0,255] range to create the 8-bit unsigned integer data format, which is a data format for grayscale images in computer vision. Due to the tested CNNs varying input dimensions, the input images (slices) were resized to match the necessary dimensions.

The MRI voxel values in NIFTI1 [178] data format can range up to 64-bit signed integers or 128-bit float data types; therefore, during the normalization procedure, the detailed quantization information is lost when the value range is brought to the digital image [0,255] unsigned integer value range.

The samples in the dataset were divided according to an 80:20 (train:test) split, with a total of 266 subjects. The CN and AD class distribution was perfectly balanced, with each class having 133 subjects. Each subject had two recordings, bringing the total size of the dataset to 532 MRI recordings. Within the ADNI dataset each subject has two or more MRI recordings. Therefore, to create a balanced dataset, each subject's first and last recordings were chosen based on the date of acquisition in ADNI-1 dataset. The MRI data is split based on subject IDs to ensure proper allocation of training and testing sets.

As MRI NIFTI data format takes several seconds to load, to speed up the data processing, the slice selection and feature extraction stages were conducted in two separate steps. First, a data structure was created for 120 slices in each plane, creating 360 data structures each containing 532 images. Afterwards, each plane was processed individually to extract features and obtain classification results.

Following feature extraction, the samples were assigned CN or AD labels according to ADNI metadata and prepared for classification. The selected classifiers are a common option for generating baseline results and can be found in the majority of ML libraries, software packages, and applications. In this study, the following 7 convolutional ML classifiers were tested: Linear Support Vector Machine (SVM), Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), Decision Tree (DT), and K-nearest Neighbour where $k = 3, 6, 9$ (KNN).

7.4. Results and Discussion

As mentioned before, each MRI plane was evaluated individually. The number of models tested for histogram-based features for a MRI plane are $120 \times 16 \times 7 = 13440$, for a total of 40320 models. Similarly, for the pre-trained CNN based features, the models tested for a MRI plane are $120 \times 18 \times 7 = 15120$, for a total of 45360 models. This brings the total model count for both feature types to 85680 trained models. Taking the large model count into consideration, only the most relevant results are shown in this section.

The main question this study seeks to answer is whether brute-force ML model testing can reveal useful patterns in the brain MRI data. Detailed model classification results are shown in two ways: the best results in each feature+model combinations and the average of the top 10% of the results, which for histogram descriptors include 1344 trained models and for pre-trained NN feature extraction, 1512 models for each plane. The best results can be used as a baseline estimate that other more advanced methods should surpass, and the average of the top 10% shows the consistency of the top performing features, classifiers, and MRI plane slices. The provided tables and figures with histograms also help to discover the best-performing features, classifiers, and their combinations.

When classifying the MRI slices, it is expected that classifiers will provide results with around 50% accuracy for head regions that are not related to detecting AD, such as the coronal slices of the face, the axial slice of the neck, and the sagittal slices of the ears. Similarly, there will be plane, slice, feature, and classifier combinations that simply fail to learn any distinguishable patterns. Out of the slices that contain relevant features, the slices with the highest accuracy should point out which brain regions are suitable for CN/AD classification.

7.4.1. Axial Plane

When considering axial slices (see Figure 34), the top four histogram-based features are Weber Local Descriptor (WLD), Local Arc Pattern (LAP), Local Directional Texture Pattern (LDTP) and HOG. According to related work, WLD has been used for intracranial hemorrhage CT classification [179] and as features for cerebral microbleed detection [180]. In computer vision, LAP is used in the field of facial expression analysis, but there are very few studies of this feature in medical imaging; for example, one study used LAP for the detection of tuberculosis using chest radiographs [181]. In the field of medicine, the acronym LAP also stands for peripheral lymphadenopathy, which is used to describe the conditions in which lymph nodes become abnormal in size, consistency, and number. LDTP, which should not be confused with Local Variations of True Phase that use the same acronym, can be found in brain tumor classification [182], mammographic breast cancer classification [183] and liver tumor segmentation [184]. The last of the top four features is HOG, which is a significantly more popular choice in medical imaging [185].

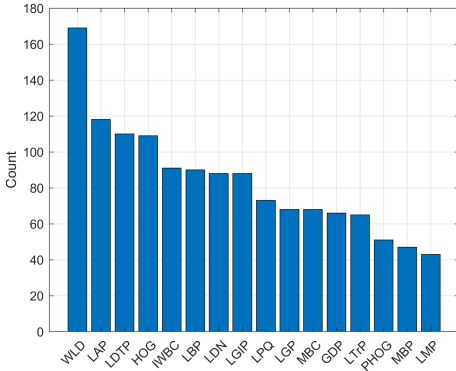


Figure 34: Axial plane: Histogram feature ranking

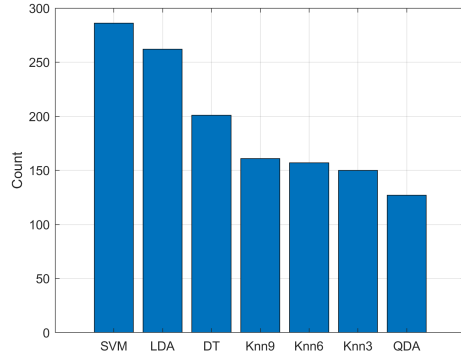


Figure 35: Axial plane: Classifiers ranking for histogram descriptors

In general, less emphasis is placed on the classifiers, as they have already been actively used in MRI based Alzheimer’s classification [186] and are only used to establish a baseline result. Advanced techniques like boosted trees, ensembles, or multi-layer perceptrons can further improve the results. Nevertheless, out of the tested classical classifiers, SVM and LDA showed the best performance for axial plane (see Figure 35).

In Table 8, which shows the highest classification results for the axial MRI slice category, it can be seen that most of the results are in the 60% to 70% range, with some combinations exceeding 70% accuracy, for example, LDA+KNN6, WLD+SVM, LAP+SVM, MBC+SVM and WLD+LDA. In the averaged top 10% classification result in Table 9, a slight drop in accuracy is noticeable, but the trend is similar, with WLD+SVM having the best performance, followed by WLD+LDA.

Considering feature extraction for axial MRI slices using pre-trained CNNs, the best choices are AlexNet, DarkNet53, SqueezeNet, and GoogLeNet (see Figure 36). A detailed examination from Table 10 shows that AlexNet has good results with several classifiers, such as SVM, QDA and DT, and DarkNet53 and SqueezeNet combined with SVM and LDA can be considered as fallback choices. For the averaged axial top 10% classification results (see Table 11), AlexNet provides the best results in combination with LDA, QDA and DT, and SqueezeNet achieved the best results for the remaining classifiers. Similarly, as with histogram descriptors, SVM and LDA were the top performing classifiers (see Figure 37).

The most interesting results come from the individual slice representation. The peak results from histogram descriptors (see Figure 38) are visible at slice number 88, and the pre-trained networks create a Gaussian looking curve around a close by location (see Figure 39). An example of the corresponding MRI axial slice is shown in Figure 40, which corresponds to an axial view slicing through *Corpus Callosum* or slightly above it. Different patterns of cortical atrophy help to distinguish between dementias such as AD and *Corpus callosum* mean surface area is significantly smaller in Alzheimer patients compared to normal controls [187].

Table 8: Axial plane: Best CN/AD classification accuracy using histogram descriptors

Feat. \Classif.	SVM	LDA	QDL	DT	KNN3	KNN6	KNN9
GDP	62.04	64.81	62.04	62.96	61.11	61.11	58.33
HOG	70.37	68.52	50.93	62.96	62.04	61.11	62.04
IWBC	68.52	65.74	50.00	64.81	62.04	63.89	61.11
LAP	70.37	69.44	60.19	64.81	63.89	62.96	64.81
LBP	62.96	62.96	59.26	61.11	62.04	62.96	64.81
LDN	67.59	68.52	62.96	67.59	64.81	74.07	65.74
LDTP	64.81	62.04	62.96	60.19	63.89	63.89	64.81
LGIP	66.67	60.19	64.81	62.04	62.04	61.11	62.04
LGP	63.89	63.89	62.96	66.67	61.11	61.11	60.19
LMP	62.96	63.89	59.26	66.67	61.11	62.04	62.04
LPQ	66.67	67.59	61.11	66.67	62.04	66.67	62.04
LTrP	62.96	64.81	62.96	66.67	62.04	62.04	66.67
MBC	69.44	62.04	50.00	61.11	61.11	62.04	60.19
MBP	62.96	61.11	59.26	62.96	63.89	64.81	63.89
PHOG	67.59	62.04	67.59	65.74	59.26	58.33	61.11
WLD	72.22	69.44	66.67	65.74	62.04	63.89	62.96

Table 9: Axial plane: Top 10% CN/AD classification accuracy using histogram descriptors

Feat. \Classif.	SVM	LDA	QDL	DT	KNN3	KNN6	KNN9
GDP	59.57	60.26	58.10	58.49	57.41	57.33	55.71
HOG	65.51	65.97	50.08	60.57	58.10	57.02	58.56
IWBC	61.88	61.73	50.00	61.81	56.87	58.87	57.87
LAP	60.26	59.49	57.87	61.19	60.73	60.42	60.80
LBP	59.65	59.41	57.25	58.02	59.26	59.34	60.11
LDN	58.41	59.72	58.95	60.96	60.65	60.96	60.49
LDTP	59.57	58.26	59.18	59.26	60.42	60.65	60.80
LGIP	59.41	57.72	61.88	59.88	57.56	57.79	59.65
LGP	59.03	60.42	59.88	61.19	58.64	57.87	57.95
LMP	59.49	60.34	57.10	59.65	57.64	58.49	57.02
LPQ	63.12	62.50	57.02	59.65	57.33	59.57	58.33
LTrP	58.87	60.26	59.03	60.65	58.72	59.18	61.03
MBC	63.73	60.88	50.00	58.87	57.72	56.33	57.87
MBP	58.33	57.64	57.18	59.80	59.41	59.41	59.72
PHOG	60.57	60.57	59.88	60.19	55.09	56.10	55.79
WLD	67.67	66.28	62.11	61.57	58.87	60.73	60.19

Table 10: Axial plane: Best CN/AD classification accuracy using pre-trained CNNs

Feat.\Classif.	SVM	LDA	QDL	DT	KNN3	KNN6	KNN9
AlexNet	74.07	70.37	62.96	72.22	64.81	65.74	67.59
DarkNet19	69.44	64.81	62.04	62.96	62.96	64.81	61.11
DarkNet53	73.15	72.22	58.33	68.52	64.81	65.74	64.81
EfficientNet	63.89	60.19	60.19	63.89	60.19	59.26	60.19
GoogLeNet	69.44	64.81	60.19	69.44	64.81	64.81	66.67
Inception-ResNet-v2	67.59	66.67	57.41	62.96	62.04	61.11	62.96
Inception-v3	69.44	68.52	58.33	64.81	66.67	62.04	64.81
MobileNet-v2	66.67	68.52	59.26	68.52	64.81	64.81	67.59
NASNet Large	68.52	68.52	55.56	63.89	65.74	62.04	62.96
NASNet Mobile	69.44	65.74	62.04	65.74	66.67	64.81	64.81
ResNet101	68.52	68.52	59.26	61.11	61.11	61.11	65.74
ResNet18	67.59	62.04	62.96	62.04	59.26	60.19	60.19
ResNet50	67.59	65.74	59.26	63.89	65.74	59.26	60.19
ShuffleNet	67.59	65.74	58.33	66.67	65.74	64.81	69.44
SqueezeNet	71.30	71.30	59.26	67.59	65.74	73.15	69.44
VGG16	65.74	65.74	62.04	62.04	65.74	63.89	65.74
VGG19	71.30	66.67	59.26	63.89	63.89	66.67	66.67
Xception	69.44	68.52	55.56	62.04	62.04	64.81	60.19

Table 11: Axial plane: Top 10% CN/AD classification accuracy using pre-trained CNNs

Feat.\Classif.	SVM	LDA	QDL	DT	KNN3	KNN6	KNN9
AlexNet	66.74	67.13	59.03	63.89	62.81	61.42	64.12
DarkNet19	64.66	60.57	57.64	59.95	59.18	58.41	58.80
DarkNet53	67.98	62.42	55.86	63.50	62.65	61.57	62.73
EfficientNet	58.80	58.26	57.10	56.94	56.64	55.94	55.32
GoogLeNet	65.43	62.58	58.02	63.89	61.73	61.34	63.27
Inception-ResNet-v2	64.35	62.96	54.78	60.11	58.26	58.41	58.18
Inception-v3	66.67	62.96	52.39	61.42	60.88	60.73	61.57
MobileNet-v2	64.89	63.50	55.79	61.57	61.19	61.57	61.50
NASNet Large	63.97	63.97	54.09	60.19	61.57	59.88	60.49
NASNet Mobile	64.97	61.57	58.33	61.34	61.11	59.41	60.65
ResNet101	65.43	63.43	56.02	59.34	58.64	58.56	59.41
ResNet18	60.11	58.87	58.49	57.87	57.79	56.79	56.48
ResNet50	65.97	63.19	55.25	59.41	59.80	57.79	56.79
ShuffleNet	65.35	61.57	56.40	61.19	62.19	62.65	62.81
SqueezeNet	68.44	62.81	55.48	61.88	63.89	63.04	65.20
VGG16	62.96	61.11	58.56	58.41	61.65	60.49	60.96
VGG19	67.44	63.81	56.79	60.96	61.03	61.88	61.27
Xception	65.97	64.04	53.09	59.65	58.87	58.87	59.03

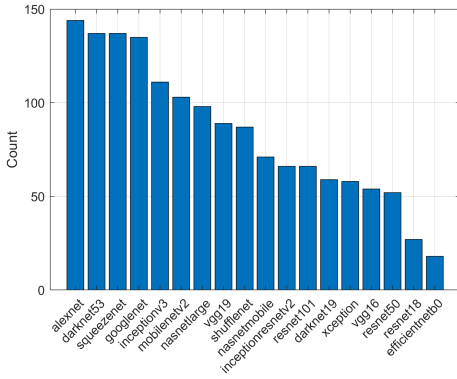


Figure 36: Axial plane: Pre-trained CNN feature ranking

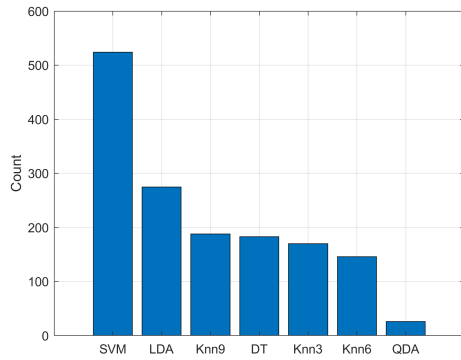


Figure 37: Axial plane: Classifiers ranking for pre-trained CNN features

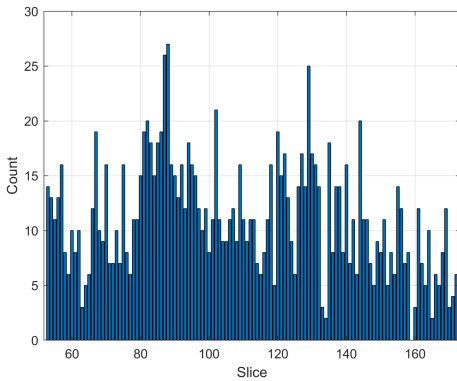


Figure 38: Axial plane: Slice histogram for histogram descriptors. Slices are ordered from top to bottom of the patient's head

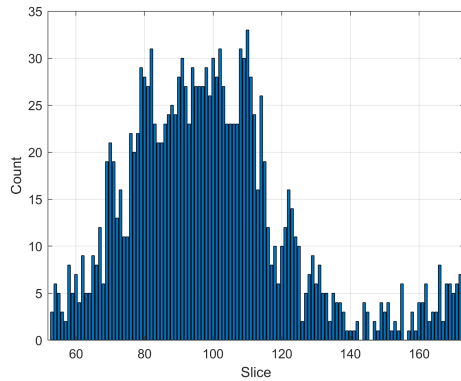


Figure 39: Axial plane: Slice histogram for pre-trained CNN features. Slices are ordered from top to bottom of the patient's head

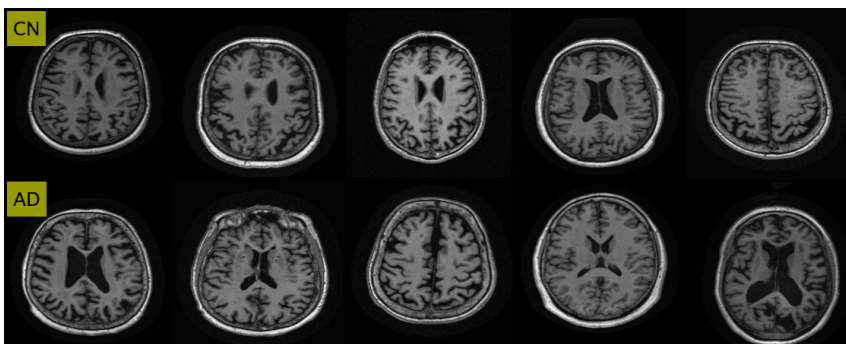


Figure 40: Axial plane: MRI slice Nr. 88 for CN (top) and AD (bottom row)

7.4.2. Coronal Plane

In the model testing for the coronal plane, the top performing histogram based features are WLD, LAP and LGP (see Table 12). The same features retain their positions, but only LAP and LGP change places in the averaged top 10% results (Table 13). The best model was achieved with WLD+SVM and WLD+LDA reaching 79.63% accuracy, and from the top 10% averaged results, WLD+SVM accuracy dropped to 71.68%.

The best results in the coronal plane were achieved using WLD, LGP and LDN histogram descriptors (Figure 41), which stays consistent with the results in Tables 12 and 13. It has to be noted that the use of LGP is more suitable for the coronal plane but not as effective in the axial plane. Similarly, as with axial plane classifier ranking, SVM and LDA were the best performing classifiers, followed by DT, QDA and KNN3 with similar top performing model counts (Figure 42).

When considering the pre-trained CNNs, AlexNet takes the lead again (Figure 43). Even in studies conducted in 2022, the use of AlexNet and transfer learning in brain pathology and AD detection is a viable option [188], [189]. The second and third places are occupied by DarkNet53 and GoogLeNet; both network architectures have been used in brain tumor detection [190], [191], as well as Alzheimer's studies [192], [193]. Regarding classifiers, SVM significantly outperforms the rest of the tested classifiers, as seen in Table 44.

In a more detailed examination, AlexNet+SVM and RasNet50+SVM achieve 75.93% accuracy, and AlexNet in combination with LDA and DT has the second best accuracy of 73.15% (Table 14). For the top 10% averaged results (Table 15), AlexNet has the highest classification for three classifiers (LDA, DT, and KNN9), but the best result of 70.06% accuracy is achieved by Inception-v3, which has been used in MRI based cancer classification [194], brain tumor segmentation [195] and Alzheimer's risk level detection [196], to mention a few.

When comparing coronal plane slices with the highest model count among the top 10% of the models, both feature variations provide different answers. According to histogram descriptors, the best slices for classification in the coronal plane are slices numbers 62 and 87 (see Figure 47 and 48), which correspond to the frontal lobe region, and pre-trained CNNs offer slice locations at numbers 101 and 77 (see Figure 49 and 50) as the best option for classification, where the frontal and temporal lobes are visible. For example, frontotemporal lobar degeneration (FTLD) and AD are the two most common forms of neurodegenerative dementia [197]. As AD progresses, damage to the frontal lobes may cause a person to have trouble making decisions, planning, or organizing [198]. Another peak value at slice number 125 in Figure 45 according to histogram descriptors can be associated with damage in the right parietal lobe, which causes trouble judging distances in three dimensions, for example, in stairway navigation. According to Figure 46, slice at number 78 and 101 at frontal lobe are good candidates for selection.

Table 12: Coronal plane: Best CN/AD classification accuracy using histogram descriptors

Feat.\Classif.	SVM	LDA	QDL	DT	KNN3	KNN6	KNN9
GDP	62.04	65.74	62.04	62.96	62.96	61.11	64.81
HOG	63.89	68.52	50.00	62.96	59.26	55.56	53.70
IWBC	64.81	71.30	50.00	65.74	64.81	65.74	62.96
LAP	69.44	63.89	66.67	71.30	63.89	61.11	68.52
LBP	64.81	65.74	60.19	61.11	62.04	62.04	64.81
LDN	66.67	69.44	65.74	64.81	66.67	63.89	59.26
LDTP	63.89	65.74	65.74	70.37	66.67	62.96	60.19
LGIP	70.37	73.15	64.81	63.89	62.96	65.74	62.04
LGP	73.15	75.93	71.30	65.74	63.89	64.81	65.74
LMP	64.81	62.96	61.11	64.81	66.67	62.04	62.04
LPQ	69.44	65.74	64.81	62.96	62.96	63.89	62.04
LTrP	62.96	62.04	59.26	62.04	62.04	62.04	66.67
MBC	74.07	70.37	53.70	63.89	59.26	58.33	57.41
MBP	67.59	65.74	60.19	65.74	62.96	62.96	65.74
PHOG	69.44	66.67	64.81	65.74	63.89	57.41	63.89
WLD	79.63	79.63	68.52	66.67	65.74	62.96	68.52

Table 13: Coronal plane: Top 10% CN/AD classification accuracy using histogram descriptors

Feat.\Classif.	SVM	LDA	QDL	DT	KNN3	KNN6	KNN9
GDP	58.80	58.87	57.64	58.41	59.72	58.56	61.19
HOG	61.88	64.58	50.00	59.65	55.48	53.24	52.70
IWBC	61.42	63.04	50.00	59.57	60.49	60.65	59.18
LAP	60.65	60.26	59.41	62.27	59.95	58.33	59.49
LBP	59.65	60.80	57.02	58.80	58.72	57.33	59.18
LDN	63.27	64.27	59.49	61.11	61.03	60.03	57.87
LDTP	59.95	60.96	61.03	60.49	59.34	58.56	56.87
LGIP	65.74	65.20	61.34	61.42	59.72	58.72	58.56
LGP	69.60	70.99	64.89	59.88	59.34	60.03	61.34
LMP	60.57	59.80	58.64	59.65	62.35	59.49	59.26
LPQ	63.43	62.27	57.33	60.57	58.41	56.25	57.33
LTrP	57.41	58.72	56.02	58.49	59.10	58.72	59.95
MBC	64.20	67.36	50.39	60.34	55.86	55.56	54.94
MBP	62.73	60.73	58.10	60.03	58.87	58.02	58.56
PHOG	63.58	61.65	62.27	60.19	58.10	55.71	57.33
WLD	71.68	70.37	64.04	59.95	62.65	60.11	62.65

Table 14: Coronal plane: Best CN/AD classification accuracy using pre-trained CNNs

Feat.\Classif.	SVM	LDA	QDL	DT	KNN3	KNN6	KNN9
AlexNet	75.93	73.15	59.26	73.15	64.81	65.74	65.74
DarkNet19	70.37	69.44	60.19	63.89	62.04	62.04	64.81
DarkNet53	73.15	69.44	60.19	64.81	64.81	65.74	65.74
EfficientNet	65.74	62.04	58.33	67.59	62.04	62.96	64.81
GoogLeNet	66.67	64.81	65.74	69.44	65.74	63.89	68.52
Inception-ResNet-v2	68.52	71.30	65.74	62.04	63.89	63.89	62.04
Inception-v3	74.07	71.30	56.48	68.52	67.59	63.89	63.89
MobileNet-v2	68.52	69.44	61.11	67.59	62.96	65.74	63.89
NASNet Large	71.30	69.44	60.19	67.59	64.81	65.74	67.59
NASNet Mobile	71.30	64.81	59.26	63.89	62.96	66.67	64.81
ResNet101	71.30	70.37	62.04	64.81	63.89	62.96	61.11
ResNet18	75.00	63.89	61.11	64.81	65.74	62.04	63.89
ResNet50	75.93	71.30	60.19	65.74	65.74	63.89	62.96
ShuffleNet	69.44	72.22	62.04	67.59	66.67	69.44	66.67
SqueezeNet	71.30	69.44	59.26	66.67	65.74	63.89	62.96
VGG16	70.37	65.74	57.41	64.81	62.04	61.11	62.96
VGG19	72.22	64.81	61.11	67.59	62.96	61.11	64.81
Xception	71.30	68.52	62.96	67.59	63.89	65.74	66.67

Table 15: Coronal plane: Top 10% CN/AD classification accuracy using pre-trained CNNs

Feat.\Classif.	SVM	LDA	QDL	DT	KNN3	KNN6	KNN9
AlexNet	68.44	66.67	57.95	65.05	62.42	62.19	64.04
DarkNet19	65.35	64.74	56.79	61.73	60.26	59.41	59.80
DarkNet53	69.68	63.58	56.25	61.03	62.11	61.57	61.88
EfficientNet	60.34	60.57	56.87	62.58	60.03	58.87	59.80
GoogLeNet	64.89	60.96	60.65	63.43	62.19	62.19	63.89
Inception-ResNet-v2	65.20	63.12	58.41	59.88	58.95	59.49	59.88
Inception-v3	70.06	64.12	53.70	62.73	62.58	61.27	61.81
MobileNet-v2	67.13	62.58	58.33	59.88	60.19	59.72	60.57
NASNet Large	67.98	65.05	54.17	61.27	60.80	63.58	63.35
NASNet Mobile	66.82	60.57	57.02	59.80	60.73	60.96	60.88
ResNet101	67.05	62.96	58.41	61.65	59.49	59.95	59.18
ResNet18	67.13	61.42	57.95	60.73	60.03	59.03	60.88
ResNet50	69.29	64.43	56.17	60.57	60.96	59.95	59.57
ShuffleNet	66.28	62.04	57.41	62.27	61.65	62.50	62.96
SqueezeNet	68.67	63.81	56.79	61.34	62.04	60.57	60.19
VGG16	63.97	61.88	55.32	61.50	59.18	58.87	59.88
VGG19	66.59	61.57	58.26	61.73	58.02	58.64	60.26
Xception	68.21	63.12	58.95	62.65	60.88	61.50	62.04

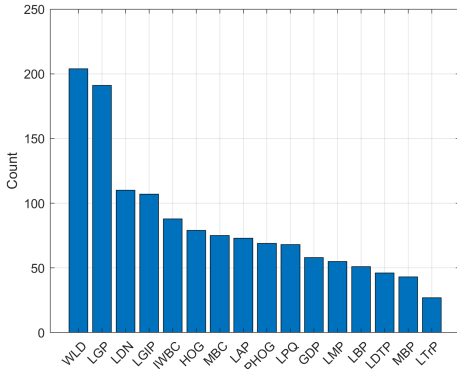


Figure 41: Coronal plane: Histogram descriptor feature ranking

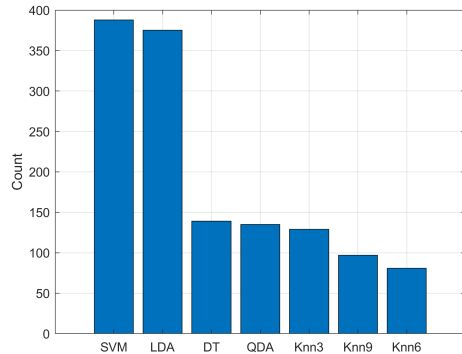


Figure 42: Coronal plane: Classifiers ranking for histogram descriptors

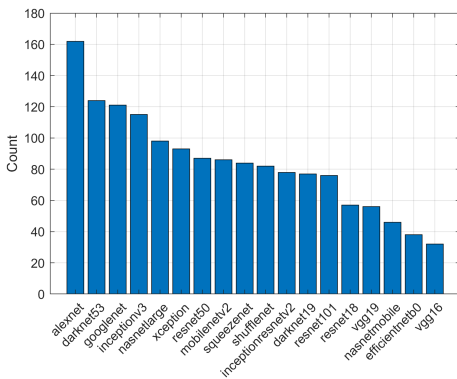


Figure 43: Coronal plane: Pre-trained CNN feature ranking

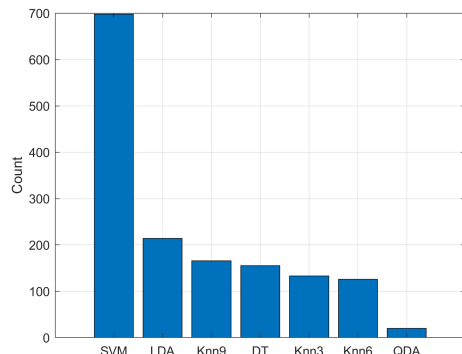


Figure 44: Coronal plane: Classifiers ranking for pre-trained CNN features

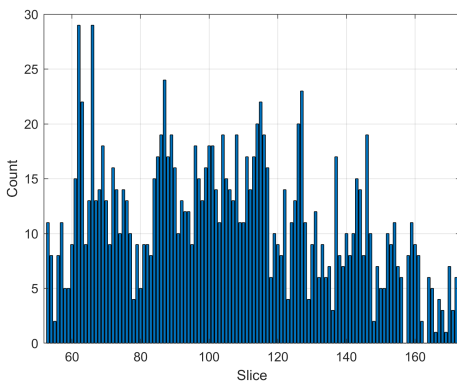


Figure 45: Coronal plane: Slice histogram for histogram descriptors. Slices are ordered from the front to the back of the patient's head

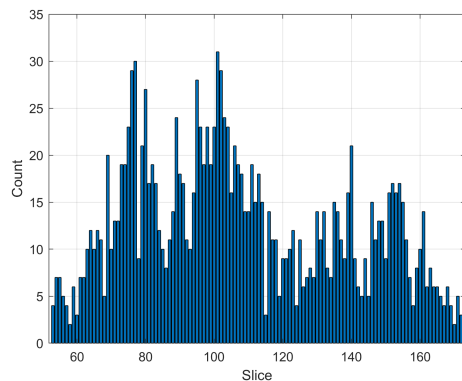


Figure 46: Coronal plane: Slice histogram for pre-trained CNN features. Slices are ordered from the front to the back of the patient's head

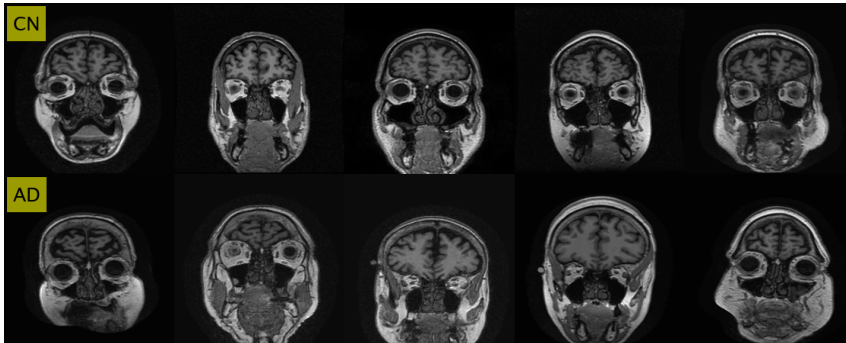


Figure 47: Coronal plane: MRI slice Nr. 62 for CN and AD. Best selection according to histogram descriptors.

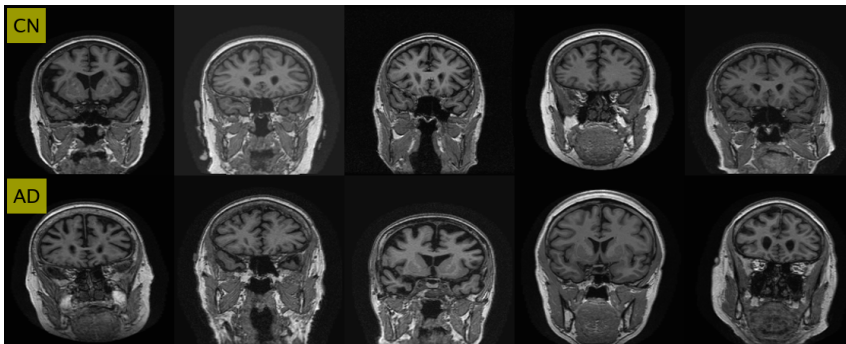


Figure 48: Coronal plane: MRI slice Nr. 87 for CN and AD. Second best selection according to histogram descriptors.

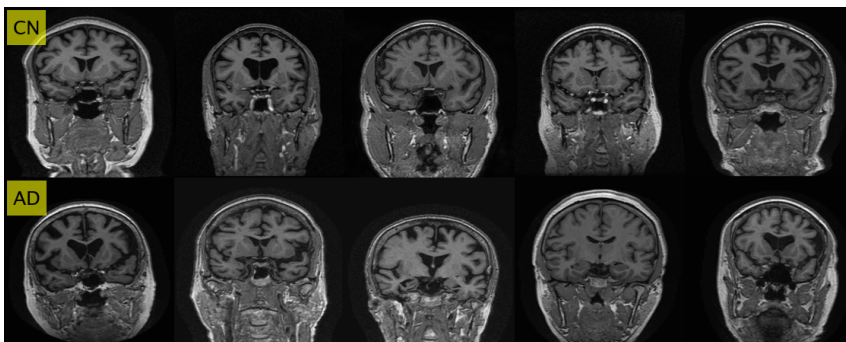


Figure 49: Coronal plane: MRI slice Nr. 101 for CN and AD. Best selection according to pre-trained CNN features.

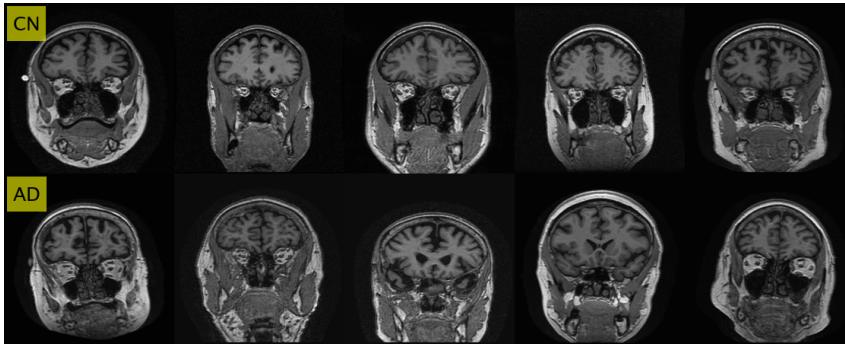


Figure 50: Coronal plane: MRI slice Nr. 77 for CN and AD. Second best selection according to pre-trained CNN features.

7.4.3. Sagittal Plane

Similarly to previous results, for the sagittal plane, SVM and LDA are the best choices for classifiers for histogram descriptors and pre-trained CNNs, as shown in Figures 52 and 54. According to the top 10% model ranking, the best performing features are LGP, WLD, LAP and HOG (Figure 51). From the other plane results, WLD has been the top feature for the axial and coronal planes; LGP was the second best for the coronal plane; and LAP was the second best in the axial plane. As mentioned before, HOG [199]–[201] and its variations is very popular feature descriptor in medical imaging.

On further inspection, HOG+SVM provides the best classification accuracy of 77.78%, followed by LGP with 75.00% and LAP with 73.15% accuracy (Table 16). For the axial and coronal planes, the histogram descriptor feature ranking and top results were consistent, but this is not the case for the sagittal plane. The differences in Figure 51 and the results in Tables 16 and 17 show that LGP has the highest amount of classifiers among the top 10% of models, but the HOG+SVM combination provides the best results in terms of classification accuracy.

When considering the pre-trained CNNs in Figure 53, it shows that AlexNet is the best choice for feature extraction in sagittal plane, followed by DarkNet53, NASNet Large, and SqueezeNet. A similar pattern is visible in Table 18 where the best result of 75.93% accuracy were achieved by SqueezeNet+SVM and the second best was achieved by darkent53+(KNN3, KNN9) with 71.30% accuracy. The averaged top 10% of the models (Table 19) show that SqueezeNet+SVM remains the best combination with 69.29%.

According to histogram descriptors, the left side of the brain has the highest model count for the top 10% of the models (Figure 55), but according to the pre-trained CNN features, the right side of the brain has the highest model count. As indicated by the sagittal plane slice histograms in Figures 55 and 56 the middle part of the brain is not an optimal choice for classification. Similarly, both feature types offer two best locations for sagittal slice selection, one on the left side (Figure 57) and one on the right side (Figure 58) of the brain, and the peak locations

Table 16: Sagittal plane: Best CN/AD classification accuracy using histogram descriptors

Feat.\Classif.	SVM	LDA	QDL	DT	KNN3	KNN6	KNN9
GDP	62.04	62.96	62.04	59.26	63.89	62.96	61.11
HOG	77.78	68.52	50.00	63.89	61.11	57.41	58.33
IWBC	70.37	66.67	50.00	65.74	60.19	61.11	61.11
LAP	61.11	61.11	61.11	73.15	62.96	65.74	64.81
LBP	66.67	64.81	62.04	67.59	62.96	62.96	64.81
LDN	63.89	66.67	60.19	63.89	63.89	69.44	67.59
LDTP	66.67	64.81	65.74	65.74	64.81	66.67	61.11
LGIP	66.67	66.67	59.26	72.22	63.89	64.81	62.96
LGP	70.37	75.00	65.74	64.81	62.96	63.89	65.74
LMP	62.04	62.96	62.96	63.89	62.96	63.89	64.81
LPQ	63.89	64.81	60.19	64.81	60.19	62.04	62.96
LTrP	63.89	63.89	59.26	62.04	67.59	62.96	63.89
MBC	67.59	62.04	50.00	60.19	57.41	57.41	63.89
MBP	65.74	64.81	62.04	62.96	65.74	63.89	64.81
PHOG	68.52	68.52	62.96	64.81	63.89	61.11	61.11
WLD	69.44	69.44	60.19	64.81	63.89	66.67	62.96

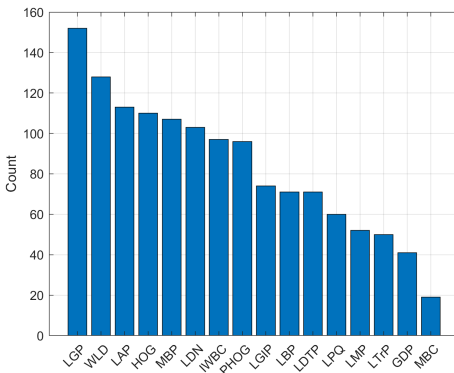


Figure 51: Sagittal plane: Histogram descriptor feature ranking

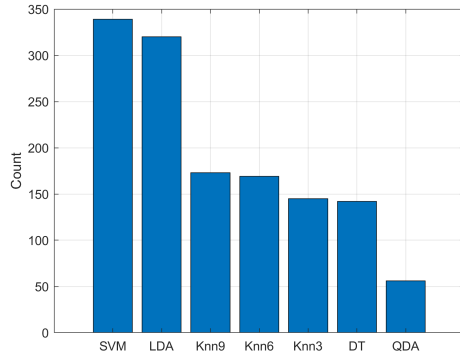


Figure 52: Sagittal plane: Classifiers ranking for histogram descriptors

only differ by a few slices. It is generally accepted that in the case of Alzheimer’s, the left side of the brain is slightly more affected than the right side [202]–[204]. Damage in the left hemisphere is associated with issues with language and semantic memory. The topic is still actively investigated, and there are several studies discussing the differences between the right and left sides of the brain for patients with Alzheimer’s [205]–[207].

The highest model count in sagittal slices goes through the temporal lobes, where damage in this region is associated with difficulty recognizing well-known faces and objects. An affected individual can appear to have forgotten a familiar face, but they may still recognize that person once they hear them talk because the pathways for hearing are still functional.

Table 17: Sagittal plane: Top 10% averaged CN/AD classification accuracy using histogram descriptors

Feat.\Classif.	SVM	LDA	QDL	DT	KNN3	KNN6	KNN9
GDP	58.64	59.41	57.87	57.87	56.33	58.18	58.87
HOG	68.21	64.35	50.00	60.42	58.80	54.71	55.25
IWBC	63.12	62.04	50.00	61.81	57.64	59.57	59.80
LAP	57.87	58.10	58.33	62.04	61.34	62.65	62.58
LBP	61.50	60.88	57.33	61.03	60.57	61.57	60.19
LDN	61.73	62.73	58.02	59.65	61.65	62.04	61.34
LDTP	62.42	62.35	60.57	59.57	59.41	59.41	59.41
LGIP	62.42	61.65	57.33	61.57	60.49	60.80	59.49
LGP	65.90	69.21	60.03	61.42	61.03	61.03	62.04
LMP	59.80	59.57	57.33	60.42	58.72	60.19	59.88
LPQ	61.50	61.57	57.41	60.49	57.64	59.18	59.49
LTrP	58.95	59.65	57.56	58.87	61.19	59.95	61.19
MBC	60.65	59.26	50.00	56.94	55.32	55.09	57.02
MBP	62.35	61.88	58.64	59.88	62.27	61.96	61.96
PHOG	64.58	65.59	59.34	60.57	57.56	58.33	57.10
WLD	65.59	64.58	58.41	60.57	61.19	61.81	61.11

Table 18: Sagittal plane: Best CN/AD classification accuracy using pre-trained CNNs

Feat.\Classif.	SVM	LDA	QDL	DT	KNN3	KNN6	KNN9
AlexNet	72.22	66.67	60.19	65.74	62.04	63.89	65.74
DarkNet19	70.37	68.52	59.26	65.74	62.96	61.11	62.04
DarkNet53	71.30	68.52	55.56	70.37	71.30	66.67	71.30
EfficientNet	63.89	64.81	58.33	62.04	65.74	62.04	62.04
GoogLeNet	69.44	69.44	57.41	69.44	66.67	61.11	66.67
Inception-ResNet-v2	67.59	65.74	56.48	65.74	64.81	65.74	65.74
Inception-v3	68.52	68.52	58.33	64.81	63.89	65.74	64.81
MobileNet-v2	75.00	62.04	62.96	62.96	66.67	68.52	68.52
NASNet Large	73.15	68.52	57.41	66.67	68.52	66.67	68.52
NASNet Mobile	73.15	65.74	62.04	62.04	62.04	68.52	65.74
ResNet101	75.00	67.59	63.89	62.04	64.81	63.89	63.89
ResNet18	66.67	63.89	60.19	62.96	63.89	63.89	59.26
ResNet50	70.37	63.89	57.41	66.67	63.89	62.04	65.74
ShuffleNet	70.37	63.89	55.56	66.67	64.81	63.89	62.04
SqueezeNet	75.93	63.89	58.33	68.52	64.81	63.89	66.67
VGG16	68.52	69.44	59.26	65.74	66.67	61.11	66.67
VGG19	70.37	64.81	60.19	63.89	63.89	62.96	60.19
Xception	69.44	66.67	61.11	64.81	69.44	62.04	62.96

Table 19: Sagittal plane: Top 10% CN/AD classification accuracy using pre-trained CNNs

Feat.\Classif.	SVM	LDA	QDL	DT	KNN3	KNN6	KNN9
AlexNet	68.67	63.89	56.87	62.65	59.65	60.96	62.27
DarkNet19	66.20	62.04	56.40	60.88	59.18	58.49	60.03
DarkNet53	68.06	64.74	53.47	61.65	62.04	61.34	62.58
EfficientNet	60.34	61.50	55.48	59.49	61.50	59.72	60.19
GoogLeNet	64.51	62.19	54.71	62.42	60.65	58.80	61.34
Inception-ResNet-v2	64.58	61.81	55.17	60.42	61.88	61.19	62.27
Inception-v3	66.28	63.12	52.78	61.81	61.42	61.34	62.89
MobileNet-v2	67.36	60.03	56.48	61.19	61.50	62.58	62.89
NASNet Large	67.28	65.28	54.63	61.19	61.96	61.96	63.27
NASNet Mobile	65.59	62.65	58.18	60.49	59.95	59.95	61.34
ResNet101	67.82	63.66	54.32	59.88	61.65	60.03	61.73
ResNet18	62.89	61.27	57.48	61.03	59.34	58.80	57.79
ResNet50	66.74	61.50	53.63	62.58	58.95	58.95	60.19
ShuffleNet	66.28	60.73	54.09	62.11	60.57	59.18	60.03
SqueezeNet	69.29	61.81	55.25	61.65	60.19	61.34	62.65
VGG16	64.12	62.35	57.10	60.26	59.65	58.26	58.18
VGG19	65.74	62.19	56.79	60.88	59.10	58.80	57.87
Xception	65.43	62.42	56.56	61.34	61.19	59.65	61.03

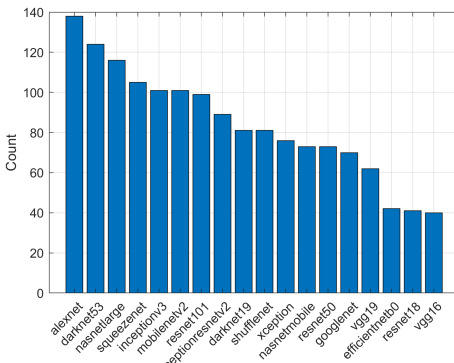


Figure 53: Sagittal plane: Pre-trained CNN feature ranking

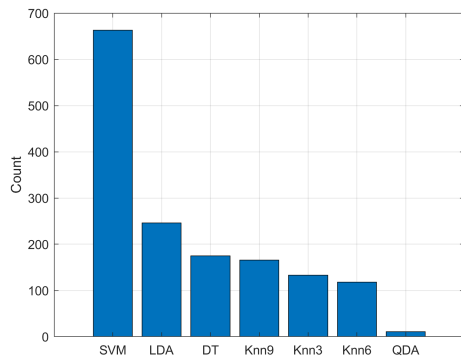


Figure 54: Sagittal plane: Classifiers ranking for pre-trained CNN features

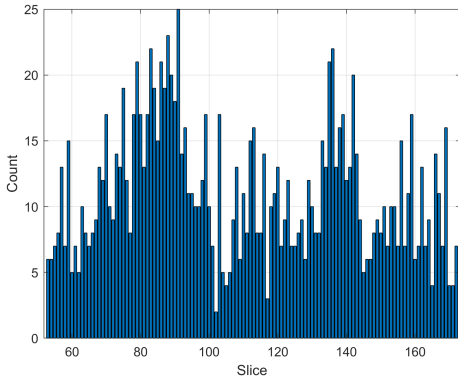


Figure 55: Sagittal plane: Slice histogram for histogram descriptors. Peak locations are at slice Nr. 91 and 136. Slices are ordered from the left to the right of the patient's head

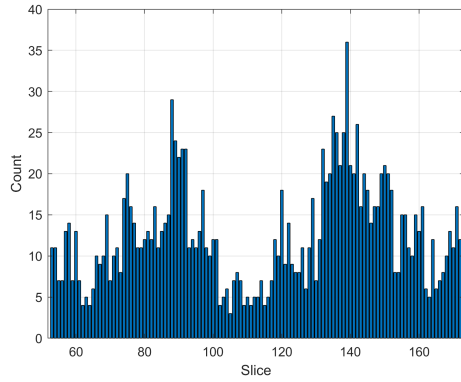


Figure 56: Sagittal plane: Slice histogram for pre-trained CNN features. Peak locations are at slice Nr. 88 and 139. Slices are ordered from the left to the right of the patient's head

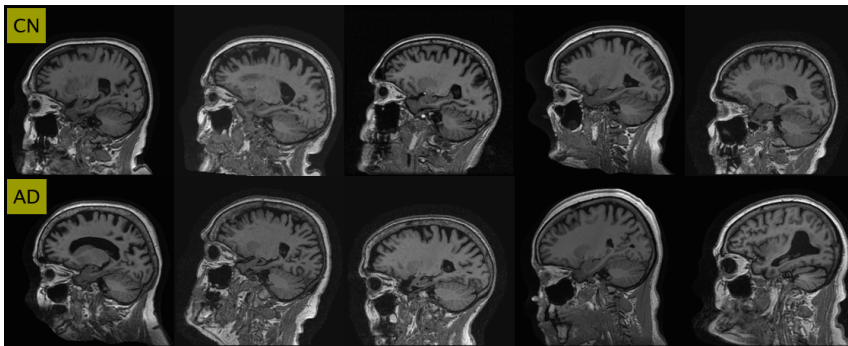


Figure 57: Sagittal plane: Slice Nr. 88 for CN and AD. Left side of the brain

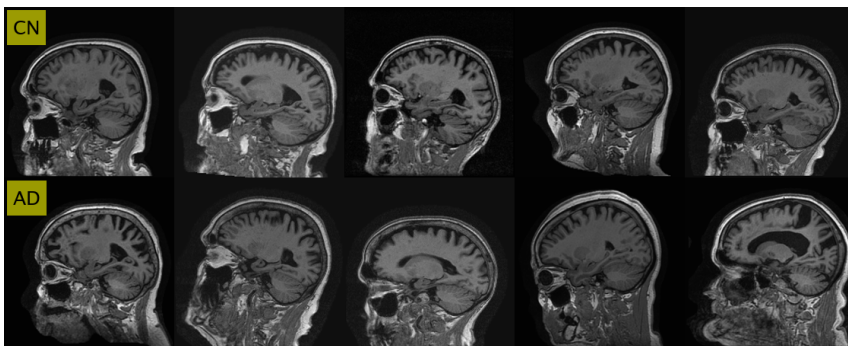


Figure 58: Sagittal plane: Slice Nr. 139 for CN and AD. Right side of the brain

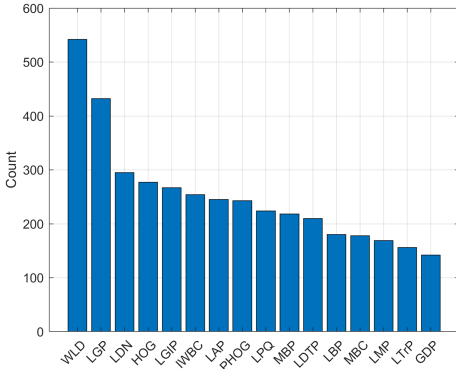


Figure 59: Global: Top 10% histogram descriptor ranking

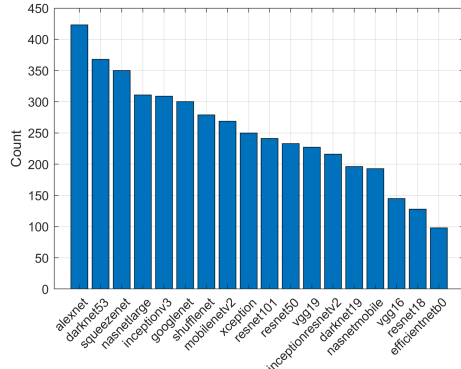


Figure 60: Global: Top 10% pre-trained CNN feature ranking

7.4.4. Multi-slice View

The combination of multiple slices in the same plane or orthogonal planes, which is referred to as the multi-slice view, has the potential to improve the classification accuracy, and this approach has been demonstrated in several publications [27], [208]–[210]. Using the knowledge gained from evaluating each MRI plane, a multi-slice view approach was tested for both feature types. The selection of both feature types is based on global (among all three planes) ranking, which can be seen in Figures 59 and 60. The top five features according to global and individual MRI plane ranking were selected for further evaluation. To make the feature values more uniform across all the slices, feature standardization was applied, which is a scaling method where the features are centered around the feature mean value with a unit standard deviation.

Regarding the slice selection, finding the best slice combination using brute force calculations is not feasible, as 120^3 options would need to be tested with the current setup. For that reason, the locations were selected according to the previously established plane histogram plots; another option would be to select the slices empirically according to other AD studies.

The histogram-based feature classification results are shown in Tables 20 and 21 showing concatenated features of the best slice locations, where the only difference in input feature configuration is in the best and second best coronal slice selection. The tables are divided into two sections, which show classification results according to global feature ranking on the top half and plane based feature ranking on the bottom half. A good example case is WLD, which is the best feature among the tested histogram descriptors. By changing the selected coronal slice, the classification accuracy with SVM differs by almost 10%. The second best feature that can be recommended is HOG, as it performs well individually and with mixed features. In the case of LDN, the features perform well with individual slices but completely fails when the features are concatenated. In this study, it was found that for histogram-based features, the best result of 75% accuracy with SVM is achieved for the following axis and feature combination: axial-88 WLD,

Table 20: Concatenated feature CN/AD classification accuracy using histogram descriptors for slices: Axial-88, Coronal-62, and Sagittal-91 for top 5 global (from Figure 59) and plane (from Figures 34, 41, and 51) feature ranking

Axial-88	Coronal-62	Sagittal-91	SVM	LDA	QDA	DT	KNN3	KNN6	KNN9
WLD	WLD	WLD	62.96	57.41	60.19	49.07	52.78	50.00	60.19
LGP	LGP	LGP	62.04	62.96	62.04	51.85	52.78	52.78	60.19
LDN	LDN	LDN	52.78	49.07	44.44	45.37	50.93	50.00	47.22
HOG	HOG	HOG	69.44	65.74	50.00	60.19	51.85	48.15	47.22
LGIP	LGIP	LGIP	55.56	56.48	54.63	48.15	49.07	55.56	52.78
WLD	WLD	LGP	71.30	67.59	54.63	55.56	62.04	57.41	57.41
LAP	LGP	WLD	62.04	61.11	53.70	56.48	54.63	62.04	60.19
LDTP	LDN	LAP	49.07	49.07	50.00	45.37	59.26	57.41	71.30
HOG	LGIP	HOG	68.52	50.00	50.00	56.48	47.22	45.37	50.00
IWBC	IWBC	MBP	68.52	55.56	50.00	50.00	57.41	58.33	63.89

Table 21: Concatenated feature CN/AD classification accuracy using histogram descriptors for slices: Axial-88, Coronal-87, and Sagittal-91 for top 5 global (from Figure 59) and plane (from Figures 34, 41, and 51) feature ranking

Axial-88	Coronal-87	Sagittal-91	SVM	LDA	QDA	DT	KNN3	KNN6	KNN9
WLD	WLD	WLD	73.15	68.52	57.41	41.67	52.78	51.85	55.56
LGP	LGP	LGP	59.26	56.48	56.48	52.78	44.44	44.44	53.70
LDN	LDN	LDN	52.78	55.56	46.30	59.26	40.74	49.07	46.30
HOG	HOG	HOG	65.74	62.96	50.00	52.78	45.37	48.15	48.15
LGIP	LGIP	LGIP	55.56	56.48	51.85	50.93	46.30	47.22	49.07
WLD	WLD	LGP	75.00	74.07	57.41	58.33	52.78	51.85	53.70
LAP	LGP	WLD	62.96	54.63	49.07	54.63	52.78	60.19	59.26
LDTP	LDN	LAP	47.22	60.19	50.93	49.07	55.56	55.56	60.19
HOG	LGIP	HOG	70.37	51.85	50.00	56.48	47.22	45.37	49.07
IWBC	IWBC	MBP	66.67	54.63	50.00	56.48	61.11	49.07	44.44

coronal-87 WLD, and sagittal-91 LGP.

The pre-trained CNN feature ranking classification results are shown in Table 22 for the right side and Table 23 for the left side of the brain, according to sagittal slice selection. AlexNet followed by DarkNet53 took the first and second places in global and individual MRI plane ranking; therefore, in global and plane ranking, the first and second positions are the same (both results are added in the tables for consistency). From the tables, it can be seen that AlexNet performs well with several classifiers, but, for example, DarkNet53 only has good results with SVM. The best results were achieved by SVM in combination with SqueezeNet with 69.44% accuracy, followed by AlexNet and DarkNet53 with 67.59% accuracy, when sagittal slice of the left side of the brain was used. For the right side of the brain, DarkNet53+SVM had the best results with 69.44% accuracy. Consequently, the mentioned three CNNs would be good choices for feature extraction and transfer learning for AD classification tasks based on MRI slices.

The last test shows the combined classification results for the best feature and slice combinations in Tables 24 and 25. The obtained results indicate that the combined feature accuracy is lower than the best individual slice accuracies and some combinations are as good as random guessing. The listed combinations were also tested with PCA with various principal component amounts, but it did not improve the accuracy in comparison to using all the features. As already seen

Table 22: Concatenated feature CN/AD classification accuracy using pre-trained CNNs for slices: Axial-88, Coronal-87, and Sagittal-139 for top 5 global (from Figure 60) and plane (from Figures 36, 43, and 53) feature ranking

Axial-88	Coron.-101	Sagit.-88	SVM	LDA	QDA	DT	KNN3	KNN6	KNN9
AlexNet	AlexNet	AlexNet	67.59	65.74	47.22	58.33	54.63	47.22	48.15
DarkNet53	DarkNet53	DarkNet53	67.59	57.41	50.00	56.48	56.48	56.48	56.48
Sque.Net	Sque.Net	Sque.Net	69.44	49.07	50.00	57.41	49.07	53.70	52.78
NASNet L.	NASNet L.	NASNet L.	62.04	62.96	51.85	58.33	53.70	59.26	56.48
Incept.V3	Incept.V3	Incept.V3	65.74	59.26	50.00	54.63	46.30	50.00	51.85
AlexNet	AlexNet	AlexNet	67.59	65.74	47.22	58.33	54.63	47.22	48.15
DarkNet53	DarkNet53	DarkNet53	67.59	57.41	50.00	56.48	56.48	56.48	56.48
SqueezeNet	GoogLeNet	NASNet L.	66.67	55.56	50.00	52.78	58.33	55.56	60.19
GoogLeNet	Incept.V3	Sque.Net	59.26	63.89	49.07	55.56	53.70	47.22	58.33
Incept.V3	NASNet L.	Incept.V3	67.59	56.48	51.85	54.63	44.44	54.63	54.63

Table 23: Concatenated feature CN/AD classification accuracy using pre-trained CNNs for slices: Axial-88, Coronal-101, and Sagittal-139 for top 5 global (from Figure 60) and plane (from Figures 36, 43, and 53) feature ranking

Axial-88	Coron.-101	Sagit.-139	SVM	LDA	QDA	DT	KNN3	KNN6	KNN9
AlexNet	AlexNet	AlexNet	64.81	64.81	48.15	59.26	62.04	61.11	61.11
DarkNet53	DarkNet53	DarkNet53	69.44	53.70	50.00	57.41	57.41	48.15	50.00
Sque.Net	Sque.Net	Sque.Net	64.81	58.33	50.00	61.11	55.56	50.93	53.70
NASNet L.	NASNet L.	NASNet L.	62.96	62.04	49.07	57.41	48.15	55.56	52.78
Incept.V3	Incept.V3	Incept.V3	66.67	58.33	50.00	62.04	52.78	55.56	50.93
AlexNet	AlexNet	AlexNet	64.81	64.81	48.15	59.26	62.04	61.11	61.11
DarkNet53	DarkNet53	DarkNet53	69.44	53.70	50.00	57.41	57.41	48.15	50.00
SqueezeNet	GoogLeNet	NASNet L.	65.74	48.15	49.07	54.63	51.85	53.70	55.56
GoogLeNet	Incept.V3	Sque.Net	67.59	61.11	49.07	66.67	58.33	54.63	60.19
Incept.V3	NASNet L.	Incept.V3	62.96	62.04	51.85	54.63	51.85	62.04	66.67

before, in the case of histogram descriptors, SVM and LDA have top classification accuracy, but, for example, KNN3 and KNN6 perform at a level of random guessing for concatenated features. In the case of pre-trained CNNs, LDA had the best results, followed by KNN6.

There is a notable difference between the selected slices in Tables 24 and 25. The selection from pre-trained CNNs are from similar locations; for example, the sagittal slice is selected from the right side of the brain from number 133 to 163 range, axial slice locations are close to number 90, and coronal is slightly more varied. For histogram descriptors, the slices center around different locations; for sagittal slices, four slices are on the left side and three are on the right side of the brain; coronal slices are split between the frontal lobe and midsection of the brain; and for axial slices, most of the features focus on the top part of the head.

According to the presented results, histogram-based descriptors slightly outperform pre-trained CNN feature extraction. The process of improving the histogram-based approach is to apply more consistent MRI voxel value normalization, followed by feature extraction and selection algorithms, and combined with a sophisticated classification algorithm with various parameters to explore.

In the case of NNs, different architectures could be tested to see if the extracted features provide enough information for classification. Given the availability of data, the applied NNs should outperform the feature engineering approach, as for

Table 24: Concatenated feature CN/AD classification accuracy using histogram descriptors selected according to Tables 8, 12, and 16

Axial	Slice	Coronal	Slice	Sagittal	Slice	Classif.	Acc.(%)
WLD	77	WLD	111	HOG	66	SVM	75.93
LAP	161	WLD	111	WLD	87	LDA	69.44
PHOG	163	LGP	116	LDTP	99	QDA	51.85
LDN	83	LAP	62	LAP	84	DT	62.04
LDN	64	LDN	154	LTrP	123	KNN3	49.07
LDN	57	IWBC	63	LDN	143	KNN6	49.07
LTrP	53	LAP	66	LDN	122	KNN9	61.11

Table 25: Concatenated feature CN/AD classification accuracy using pre-trained CNNs selected according to Tables 10, 14, and 18

Axial	Slice	Coronal	Slice	Sagittal	Slice	Classif.	Acc.(%)
AlexNet	92	AlexNet	82	Sque.Net	137	SVM	65.74
DarkNet53	95	AlexNet	103	GoogLeNet	138	LDA	69.44
AlexNet	93	GoogLeNet	89	ResNet101	155	QDA	49.07
AlexNet	90	AlexNet	88	DarkNet53	163	DT	58.33
Incept.V3	78	Incept.V3	144	DarkNet53	138	KNN3	50.93
Sque.Net	77	ShuffleNet	152	MobileNetV2	133	KNN6	68.52
shufflenet	103	GoogLeNet	74	DarkNet53	139	KNN9	64.81

most of the computer vision problems, as NNs define the current state-of-the-art.

The main advantage of local descriptors is their ease of implementation, as all the algorithms can run on CPU, but when it comes to CNNs, a dedicated hardware and more advanced ML frameworks are needed, such as PyTorch, TensorFlow, or Keras. Nevertheless, as hardware and software development progresses, the field of hand-crafted feature engineering will be replaced by NN architecture engineering due to its data driven approach and unbound potential.

The future plans for this work will focus on reducing the influence of brain size and location variance. Currently, for a specific slice, there is a slight shift in what part of the brain is visible when comparing MRI images from different patients. To address this problem, one solution can be brain region segmentation, which would allow for adjusting the slice selection according to a specific location in the brain. Another option is to directly use a 3D voxel grid as input and subsequently process the data with 3D feature extractors such as 3D-HOG [27] or use 3D CNNs [211]–[214].

8. ENSEMBLE APPROACH FOR DETECTION OF DEPRESSION USING EEG FEATURES

8.1. Introduction

According to the World Health Organization (WHO), depression is predicted to be the second most prevalent cause of disability, or an inability to work, after cardiovascular disease [215]. Therefore, depression constitutes a significant public health issue worldwide. In European countries, depression affects the life course of individuals most dramatically [216]. Within a 12 month period in 2011, 6.9% of the European population was found to suffer from clinical depression, as evidenced by the most extensive study of mental ill health in Europe [216].

Typically, medical professionals use interviews to diagnose depression, and a clinical questionnaire is often conducted in tandem with the interview. The latter might be assessed by a doctor, as with the Hamilton Depression Rating Scale (HDRS) [217], or self-reported, as with the Emotional State Questionnaire (EST-Q) [128] and the Mini-Mental State Examination (MMSE) [218]. As part of the screening process, the Beck Depression Inventory (BDI) [219] and the HDRS [220] questionnaires may also be used. In addition to the interpretation of clinical questionnaires, patients' brain activity can be monitored directly using several imaging modalities. Examples of this approach include CT, fMRI, and EEG. The latter method, EEG, has garnered significant research interest because it is a cost-efficient and straightforward tool that has shown promising results [46], [221]–[224]. Using several linear and non-linear features [225] and providing broad insights into the feature calculation, this chapter presents classification results based on the Tallinn University of Technology (TalTech) EEG dataset by utilising feature selection and classifier configuration as a means to improve the accuracy for classification of clinical depression.

8.2. Features

EEG signals are typically analyzed by segmenting the signal into functionally distinct frequency bands, such as delta (1-4 Hz), theta (4-8 Hz), alpha (8-12 Hz), beta (12-30 Hz), and gamma (30-45 Hz). In the present study, the power spectral density of the EEG signal was determined using Welch's method in accordance with Bachmann et al. [46]. By integrating the power spectral density within the boundary frequencies of the EEG spectral bands, the EEG powers in the theta, alpha, beta, and gamma frequency bands were found. Relative Band Power (RBP) of EEG signals (T_{RBP} , A_{RBP} , B_{RBP} , G_{RBP}) are expressed as the power in the specific EEG frequency band as a percentage of the total power of the signal.

Linear methods were tested using a consistent sample rate of 400 Hz. Contrastingly, due to a substantial computational load, non-linear methods were tested using downsampled signals.

8.2.1. Linear Features

Alpha Power Variability. Using a pass-band filter, the alpha band signal (8-12 Hz) was determined. Subsequently, the APV was quantified using three steps for the artifact-free 10-second segments. The alpha band signal power in the time window T for $N = 4000$ samples was calculated as follows:

$$W_i = \frac{1}{N} \sum_{r=1}^N [V(r)]^2 \quad (8.1)$$

$V(r)$ represents the amplitude of the alpha band signal in a sample, r and N indicate the number of samples in the time window T . In light of this, APV was calculated as follows:

$$APV = \frac{\sigma}{W_0} \quad (8.2)$$

W_0 denotes the value of alpha band power on average over a five-minute period, and σ represents the standard deviation of those segments.

Spectral Asymmetry Index. The power at low and high frequencies was evaluated using the SASI and quantified as a relative difference between the band power at higher and lower EEG frequencies. The EEG spectral symmetry is indicated by the power balance [47]. The powers within the frequency bands were determined as follows:

$$P_{\delta mn} = \sum_{f_i=F_c-6}^{f_i=F_c-2} S_{mn} \quad (8.3)$$

and

$$P_{\beta mn} = \sum_{f_i=F_c+2}^{f_i=F_c+26} S_{mn} \quad (8.4)$$

F_c represents the central frequency of the EEG spectrum maximum in the alpha band and was determined for each individual. In channel m , the SASI for each subject (n) was calculated as follows:

SASI in channel m for a subject n is calculated as

$$SASI_{mn} = \frac{P_{\beta mn} - P_{\delta mn}}{P_{\beta mn} + P_{\delta mn}} \quad (8.5)$$

8.2.2. Non-linear Features

Underlying physiological brain activity causes chaotic behavior within EEG signals. Non-linear methods were used to apprehend this phenomenon [222]. Higuchi Fractal Dimension (HFD), Lempel Ziv Complexity (LZC), and Detrended Fluctuation Analysis (DFA) were used to describe participants' brain activity.

Higuchi Fractal Dimension. The complexity and fractal dimension of time series signals such as the EEG can be measured using HFD. With a parameter $k_{max} = 8$, the HFD values of each electrode were determined in accordance with Higuchi's methods [49].

Lempel Ziv Complexity. LZC [226] can be used to determine the complexity of signals. Further, by representing spatio-temporal activity patterns in high dimensionality non-linear systems, LZC can reveal the regularity and randomness of EEG signals. Each signal segment is converted into a binary sequence $s(n)$ in order to calculate LZC:

$$s(n) = \begin{cases} 1, & \text{if } x(n) > m \\ 0, & \text{if } x(n) \leq m \end{cases} \quad (8.6)$$

$x(n)$ describes the signal segment, n describes the segment's sample index from 1 to N (segment length), and m denotes the threshold value. The number of patterns is counted by the binary sequence $s(n)$ as it is scanned from left to right. Each time a new pattern arises, the complexity value $c(n)$ increases. LZC values are quantified as follows:

$$C(N) = \frac{c(N)}{b(N)} \quad (8.7)$$

$b(N)$ describes the upper bound of $c(n)$,

$$\lim_{n \rightarrow \infty} c(n) = b(N) = \frac{N}{\log_a N} \quad (8.8)$$

and $c(n)$ is used to both normalize LZC values and avoid discrepancies in segment length.

Detrended Fluctuation Analysis. In EEG signals, the presence and persistence of long-range correlations over time are assessed using DFA. The resting EEG of healthy participants was found to present persistent long-range correlation in time [55]. In the time domain, DFA was quantified in accordance with the approach outlined by Peng et al. [54], [227].

8.3. Methodology

The 1D vectors, constructed from a set of feature calculations, were used to depict the calculated features: T_{RBP} , A_{RBP} , B_{RBP} , G_{RBP} , APV, SASI, HFD, LZC, and DFA for electrodes FP1, FP2, F7, F3, FZ, F4, F8, T3, C3, C4, T4, T5, P3, PZ, P4, T6, O1, and O2, where each feature group had 18 features that corresponded with each of the electrode locations. Employing F-Test or ReliefF evaluation methods can limit the number of features, and more diverse feature sets can be assembled by concatenating these features. 10-fold cross-validation was used to assess each method, and trained models were dispensed with after each iteration. The

cross-validation procedure was also reproduced 100 times using a randomized sample order as a means by which to limit the effect of the sample order. Further, each fold involved the same number of healthy and depressed participants in order to maintain balance within the training data. For each fold, in order to ascertain predictions about the weighted and boosted ensemble, the training set was assessed using another 9-fold procedure (see Figure 61). Subsequently, the weight of each classifier vote was tailored to the results produced in the training set. Accordingly, AdaBoost used predictions from the training set to quantify the weights of each classifier within the ensemble.

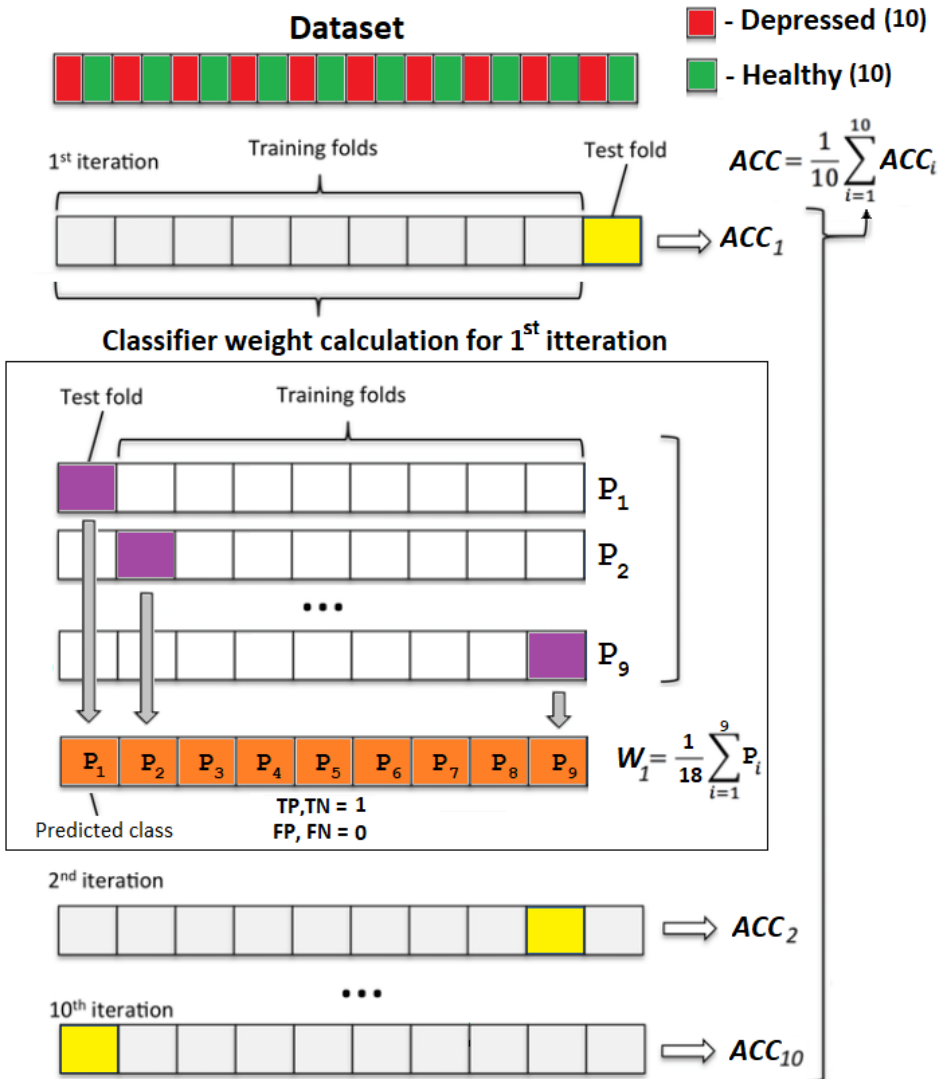


Figure 61: Cross validation procedure for ensemble classifier with weight calculation

8.3.1. Feature Selection

Cognitive disorders may introduce observable changes in EEG recordings. Compared to cognitively normal brains, a brain affected by a cognitive disorder may present with statistically significant differences across each region depending on the feature calculations used. In this study, feature subset selection was used as a preparatory step prior to the model being learned in order to select the most relevant electrode locations. More specifically, the F-Test and ReliefF were applied. The former is frequently used to indicate the statistical significance between two classes, and the latter is a rank-based feature selector.

Univariate Feature Ranking using F-Tests. Using an F-Test, the univariate feature ranking algorithm provides further information about the significance of each feature by examining the importance of each predictor. F-Tests assess whether the response values (as grouped together according to predictor variable values) are drawn from populations with the same mean or different means [228].

ReliefF. The ReliefF algorithm approximates the quality of attributes in accordance with how successfully these attributes can distinguish between instances in close proximity. The algorithm penalizes predictors that attribute different values to neighbors within the same class and rewards predictors that attribute different values to differing classes [229].

8.3.2. Machine Learning Algorithms

In accordance with survey papers published by Lakshmi et al. [230] and Alotaiby et al. [231], the supervised ML algorithms employed in the present study have been extensively used in existing research. For the process of binary classification, this study uses the following algorithms:

- Support Vector Machine (SVM) with Radial Basis Function kernel
- Linear Discriminant Analysis (LDA)
- Naive Bayes (NB)
- K-Nearest Neighbours (KNN)
- Decision Tree (DT)

An ensemble approach was used, in which classifiers were trained in relation to each feature group. Simultaneously, an individual evaluation of each of the listed classifiers and feature groups was undertaken. Ultimately, this made predictions about the class label possible.

8.4. Results and Discussion

By assessing each feature group individually, a baseline accuracy was ascertained. HFD and DFA generated an acceptable accuracy rate of 80%, and used at least one of the classifiers (see Table 26). Weak classifiers were found among all of the feature groups, thereby confirming the relevance of the data, as a minimum

of one classifier always generated an accuracy rate of over 50%. Evidently, the aforementioned features have functioned as successful biomarkers of depression [44]; however, certain classifiers underperformed. This may have been an effect of redundant features on classifier performance.

Table 26: EEG feature classification accuracy

Feature group	Classifier accuracy (%)				
	RBF SVM	LDA	NB	KNN	DT
T _{RBP}	54.40	65.00	73.30	44.45	38.65
A _{RBP}	50.00	64.15	70.95	58.70	66.55
B _{RBP}	70.00	52.90	65.05	62.20	79.85
G _{RBP}	38.45	52.95	59.40	50.90	54.85
APV	35.40	27.05	31.85	37.70	64.65
SASI	54.55	55.00	54.60	59.10	53.15
HFD	55.40	41.55	51.80	71.55	82.55
LZC	80.70	57.10	58.50	75.95	63.50
DFA	68.15	45.75	63.15	70.80	74.55

F-Tests, a form of univariate feature ranking, were utilized to determine the most effective features. Classification results were gathered for each feature group, and features were categorized according to their relevance.

The evaluation progressed using the most relevant feature and incorporated a feature of less relevance during each subsequent iteration. Table 27 presents the highest accuracy produced by certain features, and Table 29 contains the list of selected features. In accordance with the highest accuracy among all five classifiers within each feature group, the most favorable feature set was chosen. Excluding LZC, the average classifier accuracy increased among almost all feature groups in comparison with the results shown in Tables 26 and 27. Similarly to the application of F-Tests, the most relevant features among each group were selected using ReliefF. Within almost all feature groups, ReliefF results were always higher (Table 28) compared to the baseline and F-Test feature selection classification results (Tables 26 and 27). B_{RBP} results remained consistent, as both feature selection algorithms utilized features from O1 and O2 electrodes and produced the most effective results. The selected features that were used for classification tasks mentioned in Table 27 for F-Tests and Table 28 for ReliefF are shown in Table 29. FP1 and FP2, which correspond to the brain’s frontal lobes, and O1 and O2 in the occipital lobes, were found to be the most used electrode locations.

The selected features that were used for classification tasks mentioned in Table 27 for F-tests and Table 28 for ReliefF are shown in Table 29. The selected features that represent electrode locations show that the most commonly used features are FP1, FP2, which correspond to the frontal lobes, and O1, O2, which correspond to the occipital lobes of the brain.

Feature groups were assessed in combination, rather than being classified individually, and the feature vectors were established by concatenating all available

Table 27: Classifier accuracy for EEG features selected by univariate feature ranking using F-Tests

Feature group	Classifier accuracy (%)				
	RBF SVM	LDA	naive Bayes	KNN	DT
T _{RBP} (2)	55.20	65.60	69.85	55.75	45.70
A _{RBP} (6)	67.80	62.45	65.50	57.30	75.05
B _{RBP} (2)	90.00	80.85	79.90	90.00	90.00
G _{RBP} (10)	51.25	60.09	69.90	55.00	55.80
APV (7)	34.90	36.30	33.30	46.25	73.05
SASI (3)	67.25	65.00	60.90	48.25	58.25
HFD (5)	88.00	41.50	48.70	75.00	71.10
LZC (10)	60.65	67.50	56.25	75.00	68.15
DFA (3)	66.90	64.45	57.90	68.65	67.85

(*) Number of features used

Table 28: Classifier accuracy for EEG features selected by ReliefF algorithm

Feature group	Classifier accuracy (%)				
	RBF SVM	LDA	naive Bayes	KNN	DT
T _{RBP} (5)	66.15	79.60	80.00	72.25	55.85
A _{RBP} (2)	81.20	78.70	75.95	90.00	85.30
B _{RBP} (2)	90.00	80.85	79.90	90.00	90.00
G _{RBP} (1)	75.00	69.00	75.00	70.00	63.25
APV (3)	58.85	51.55	48.65	60.60	65.85
SASI (2)	63.35	72.35	70.70	55.00	72.55
HFD (4)	78.75	49.35	65.65	75.95	85.70
LZC (2)	81.25	78.25	72.75	81.95	69.55
DFA (3)	78.80	56.40	72.55	86.00	72.90

(*) Number of features used

Table 29: Selected electrodes based on F-test and ReliefF

Feature group	Selected electrodes	
	Univariate feature ranking using F-Tests	ReliefF
T _{RBP}	F4 F8	O1 PZ P4 C4 P3
A _{RBP}	F4 F8 C3 T4 P4 O1	O1 O2
B _{RBP}	O1 O2	O2 O1
G _{RBP}	F7 F3 FZ F4 T3 T4 O1 O2 FP1 FP2	FP1
APV	T3 C3 F3 F4 F8 I3 FP1	F3 O2 C3
SASI	FP1 FP2 F7	FP1 F3
HFD	FP1 FP2 FZ F8 C3	FP1 O1 FP2 T5
LZC	F3 F4 T4 FP1 FP2 FZ F8 P3 PZ O1	FP1 FZ
DFA	O1 O2 FP2	FP1 FP2 O1

features. Only the features listed in Table 29 are concatenated, and Table 30 indicates the beneficial effects of feature selection. The use of F-Test and ReliefF

in the selection of features produced higher results than those produced at the baseline for all features.

Table 30: Concatenated EEG feature classifier accuracy

Features	Classifier accuracy (%)				
	RBF SVM	LDA	NB	KNN	DT
All features	53.25	52.35	65.20	55.50	56.70
F-test features	52.10	57.55	65.35	62.15	73.75
ReliefF features	55.10	71.75	72.35	80.00	65.05

An ensemble approach that utilizes weak classifiers within voting can facilitate the creation of a more effective solution. In Table 31, each result is produced by combining 10 classifiers of the same type. One classifier was used for each feature group. The features used from each feature group were selected in accordance with the data contained in Table 29. Using all features and one classifier produced less effective results than the majority voting ensemble approach. The ensemble approach improves results to a greater extent when F-Tests and ReliefF feature selection algorithms are used (see Table 30). Overall, ReliefF classification results exceeded the capability of an ensemble approach and the features of which were selected using F-Tests.

In most instances, using AdaBoost to select classifier weight did not produce better results than majority voting and weighted ensemble approaches. The AdaBoost algorithm can determine optimal weights using limited classifiers and disregard the remaining data, but it underperforms in comparison to other ensembles.

All five classifiers for B_{RBP} electrodes O1, and O2 reached 86.15% accuracy on average. In turn, ReliefF features that utilized majority voting generated an average accuracy of 86.25%. Overall, both of these approaches produced the most effective results using the available dataset.

Table 31: Ensemble classifier accuracy

Feature group ensemble	Classifier accuracy (%)				
	RBF SVM	LDA	NB	KNN	DT
All + Majority	70.55	58.05	61.85	68.65	79.05
F-test + Majority	81.95	70.60	78.60	78.25	74.75
ReliefF + Majority	87.75	81.30	86.60	86.30	89.30
All + Weighted	65.50	56.55	63.85	69.00	73.80
F-test + Weighted	83.25	70.15	76.50	78.85	73.20
ReliefF + Weighted	84.75	80.85	82.55	88.15	87.50
All + Adaboost	70.80	56.50	71.40	63.75	69.15
F-test + Adaboost	80.70	70.35	80.70	82.90	79.25
ReliefF + Adaboost	79.85	72.70	78.65	84.85	81.15

9. CONCLUSION

This thesis provides an introduction to data acquisition devices such as magnetic resonance imaging and electroencephalograms, along with further insights into machine learning. The research part was carried out for Alzheimer's disease classification using magnetic resonance imaging and clinical depression classification using electroencephalograms.

Utilizing brain magnetic resonance imaging scans obtained from the Alzheimer's Disease Neuroimaging Initiative dataset, machine learning algorithms have been used to identify the difference between cognitively normal and Alzheimer's disease patients. By using the histograms of sagittal slices and a Gaussian support vector machine, the first study showed that it was possible to classify slices from brain scans with an accuracy of 64.0% and the study also showed how classification is affected when using slices at different locations of the brain. Further on, this study was expanded with 16 histogram-based local descriptors and feature extraction using 18 pre-trained neural networks for Alzheimer's classification. The imaging planes were investigated individually, and the following feature types achieved the best results when classified using support vector machine; in the axial plane, using Weber local descriptor 72.22% and AlexNet 74.07%, coronal plane, with Weber local descriptor 79.63%, AlexNet and RasNet50 75.93%, sagittal plane, using histogram of oriented gradients 77.78% and SqueezeNet 75.93%. The top classification results also indicated the best locations for slice selection in each plane, and the found locations also corresponded to brain regions mentioned in medical publications in Alzheimer's studies. The future work of this approach will focus on performing slice selection according to segmented brain regions as reference points.

The human brain was also examined using electroencephalograms obtained from Tallinn University of Technology, with the use of linear and non-linear features to categorize the long-lasting effects of depression in electroencephalogram relative band power. Up to 90% accuracy in classification was achieved with the help of feature selection methods such as univariate feature ranking. Majority voting and a weighted ensemble of classifiers consistently gave good results, with the highest accuracy recorded at 89.30%. These results point to the possibility of using machine learning and electroencephalogram data to precisely identify and categorize neurological illnesses, but more clinically valid data is needed to further validate the developed approach.

In conclusion, the academic research in incorporating machine learning with magnetic resonance imaging and electroencephalograms has revolutionized the detection of brain abnormalities, paving the way for enhanced patient outcomes and individualized treatments. Even though the future for machine learning in the field of medical diagnostics is very promising, the practical application of such tools is currently limited by the challenges associated with demonstrating their effectiveness in clinical trials and certifying the developed methods.

BIBLIOGRAPHY

- [1] A. Association. (2019). “Alzheimer’s disease and dementia.” Accessed: 2019-05-02, [Online]. Available: <https://www.alz.org/alzheimers-dementia/what-is-dementia>.
- [2] R. J. Castellani, G. Perry, and G. L. Iverson, “Chronic effects of mild neurotrauma: Putting the cart before the horse?” *Journal of Neuropathology & Experimental Neurology*, vol. 74, no. 6, pp. 493–499, 2015.
- [3] A. Association. (2019). “Alzheimer’s disease and dementia.” Accessed: 2019-05-02, [Online]. Available: http://www.alz.org/alzheimers_disease_what_is_alzheimers.asp.
- [4] R. J. Castellani and G. Perry, “The complexities of the pathology - pathogenesis relationship in alzheimer disease,” *Biochemical pharmacology*, vol. 88, no. 4, pp. 671–676, 2014.
- [5] Y. Zhang, S. Wang, P. Phillips, J. Yang, and T.-F. Yuan, “Three-dimensional eigenbrain for the detection of subjects and brain regions related with alzheimer’s disease,” *Journal of Alzheimer’s Disease*, vol. 50, no. 4, pp. 1163–1179, 2016.
- [6] R. Brookmeyer, E. Johnson, K. Ziegler-Graham, and H. M. Arrighi, “Forecasting the global burden of alzheimer’s disease,” *Alzheimer’s & dementia*, vol. 3, no. 3, pp. 186–191, 2007.
- [7] A. Association. (2019). “Alzheimer’s disease and dementia.” Accessed: 2019-05-02, [Online]. Available: https://www.alz.org/alzheimer_s_dementia.
- [8] M. Prince, R. Bryce, and C. Ferri. (2019). “World alzheimer report 2011: The benefits of early diagnosis and intervention.” Accessed: 2019-05-02, [Online]. Available: <https://alz.co.uk/research/WorldAlzheimerReport2011.pdf>.
- [9] J. M. Roe, D. Vidal-Piñeiro, Ø. Sørensen, A. M. Brandmaier, S. Düzel, H. A. Gonzalez, R. A. Kievit, E. Knights, S. Kühn, U. Lindenberger, *et al.*, “Asymmetric thinning of the cerebral cortex across the adult lifespan is accelerated in alzheimer’s disease,” *Nature communications*, vol. 12, no. 1, p. 721, 2021.
- [10] A. Association. (2019). “Alzheimer’s disease and dementia.” Accessed: 2019-05-02, [Online]. Available: <https://www.alz.org/alzheimers-dementia/research-progress/earlier-diagnosis>.
- [11] J. Neugroschl and S. Wang, “Alzheimer’s disease: Diagnosis and treatment across the spectrum of disease severity,” *Mount Sinai Journal of Medicine: A Journal of Translational and Personalized Medicine*, vol. 78, no. 4, pp. 596–612, 2011.
- [12] T. Sunderland, J. L. Hill, A. M. Mellow, B. A. Lawlor, J. Gundersheimer, P. A. Newhouse, and J. H. Grafman, “Clock drawing in alzheimer’s disease: A novel measure of dementia severity,” *Journal of the American Geriatrics society*, vol. 37, no. 8, pp. 725–729, 1989.
- [13] G. B. Frisoni, N. C. Fox, C. R. Jack Jr, P. Scheltens, and P. M. Thompson, “The clinical use of structural mri in alzheimer disease,” *Nature Reviews Neurology*, vol. 6, no. 2, p. 67, 2010.
- [14] K. A. Johnson, N. C. Fox, R. A. Sperling, and W. E. Klunk, “Brain imaging in alzheimer disease,” *Cold Spring Harbor perspectives in medicine*, vol. 2, no. 4, a006213, 2012.
- [15] J.-H. Park, S. Kim, C.-H. Kim, A. Cichocki, and K. Kim, “Multiscale entropy analysis of eeg from patients under different pathological conditions,” *Fractals*, vol. 15, no. 04, pp. 399–404, 2007.
- [16] L. Orgo, M. Bachmann, K. Kalev, M. Järvelaid, J. Raik, and H. Hinrikus, “Resting eeg functional connectivity and graph theoretical measures for discrimination of depression.” in *2017 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, IEEE, 2017, pp. 389–392.
- [17] D. Zhang, D. Shen, A. D. N. Initiative, *et al.*, “Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in alzheimer’s disease,” *NeuroImage*, vol. 59, no. 2, pp. 895–907, 2012.

- [18] R. Cuingnet, E. Gerardin, J. Tessieras, G. Auzias, S. Lehéricy, M.-O. Habert, M. Chupin, H. Benali, O. Colliot, A. D. N. Initiative, *et al.*, “Automatic classification of patients with alzheimer’s disease from structural mri: A comparison of ten methods using the adni database,” *neuroimage*, vol. 56, no. 2, pp. 766–781, 2011.
- [19] E. Westman, J.-S. Muehlboeck, and A. Simmons, “Combining mri and csf measures for classification of alzheimer’s disease and prediction of mild cognitive impairment conversion,” *Neuroimage*, vol. 62, no. 1, pp. 229–238, 2012.
- [20] Q. Zhou, M. Goryawala, M. Cabrerizo, J. Wang, W. Barker, D. A. Loewenstein, R. Duara, and M. Adjouadi, “An optimal decisional space for the classification of alzheimer’s disease and mild cognitive impairment,” *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 8, pp. 2245–2253, 2014.
- [21] K. R. Gray, P. Aljabar, R. A. Heckemann, A. Hammers, D. Rueckert, A. D. N. Initiative, *et al.*, “Random forest-based similarity measures for multi-modal classification of alzheimer’s disease,” *NeuroImage*, vol. 65, pp. 167–175, 2013.
- [22] G. A. Papakostas, A. Savio, M. Graña, and V. G. Kaburlasos, “A lattice computing approach to alzheimer’s disease computer assisted diagnosis based on mri data,” *Neurocomputing*, vol. 150, pp. 37–42, 2015.
- [23] D. S. Marcus, T. H. Wang, J. Parker, J. G. Csernansky, J. C. Morris, and R. L. Buckner, “Open access series of imaging studies (oasis): Cross-sectional mri data in young, middle aged, nondemented, and demented older adults,” *Journal of cognitive neuroscience*, vol. 19, no. 9, pp. 1498–1507, 2007.
- [24] D. Zhang, Y. Wang, L. Zhou, H. Yuan, D. Shen, A. D. N. Initiative, *et al.*, “Multimodal classification of alzheimer’s disease and mild cognitive impairment,” *Neuroimage*, vol. 55, no. 3, pp. 856–867, 2011.
- [25] K. S. Nayaki and A. Varghese, “Alzheimer’s detection at early stage using local measures on mri: A comparative study on local measures,” in *2014 International Conference on Data Science & Engineering (ICDSE)*, IEEE, 2014, pp. 224–227.
- [26] O. Ben Ahmed, J. Benois-Pineau, M. Allard, C. Ben Amar, G. Catheline, and A. D. N. Initiative, “Classification of alzheimer’s disease subjects from mri using hippocampal visual features,” *Multimedia Tools and Applications*, vol. 74, pp. 1249–1266, 2015.
- [27] D. Sarwinda and A. Bustamam, “3d-hog features–based classification using mri images to early diagnosis of alzheimer’s disease,” in *2018 IEEE/ACIS 17th International Conference on Computer and Information Science (ICIS)*, IEEE, 2018, pp. 457–462.
- [28] T. Altaf, S. M. Anwar, N. Gul, M. N. Majeed, and M. Majid, “Multi-class alzheimer’s disease classification using image and clinical features,” *Biomedical Signal Processing and Control*, vol. 43, pp. 64–74, 2018.
- [29] R. Mahmood and B. Ghimire, “Automatic detection and classification of alzheimer’s disease from mri scans using principal component analysis and artificial neural networks,” in *2013 20th International Conference on Systems, Signals and Image Processing (IWSSIP)*, IEEE, 2013, pp. 133–137.
- [30] Y. Ding, C. Zhang, T. Lan, Z. Qin, X. Zhang, and W. Wang, “Classification of alzheimer’s disease based on the combination of morphometric feature and texture feature,” in *2015 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, IEEE, 2015, pp. 409–412.
- [31] P. Cao, J. Gao, and Z. Zhang, “Multi-view based multi-model learning for mci diagnosis,” *Brain Sciences*, vol. 10, no. 3, p. 181, 2020.
- [32] D. S. Cohen, K. A. Carpenter, J. T. Jarrell, X. Huang, A. D. N. Initiative, *et al.*, “Deep learning-based classification of multi-categorical alzheimer’s disease data,” *Current neurobiology*, vol. 10, no. 3, p. 141, 2019.
- [33] S. Basaia, F. Agosta, L. Wagner, E. Canu, G. Magnani, R. Santangelo, M. Filippi, A. D. N. Initiative, *et al.*, “Automated classification of alzheimer’s disease and mild cognitive im-

- pairment using a single mri and deep neural networks,” *NeuroImage: Clinical*, vol. 21, p. 101–645, 2019.
- [34] C. R. Jack, R. C. Petersen, P. C. O’Brien, and E. G. Tangalos, “Mr-based hippocampal volumetry in the diagnosis of alzheimer’s disease,” *Neurology*, vol. 42, no. 1, pp. 183–183, 1992.
- [35] B. Magnin, L. Mesrob, S. Kinkingnéhun, M. Péligrini-Issac, O. Colliot, M. Sarazin, B. Dubois, S. Lehericy, and H. Benali, “Support vector machine-based classification of alzheimer’s disease from whole-brain anatomical mri,” *Neuroradiology*, vol. 51, no. 2, pp. 73–83, 2009.
- [36] E. Hosseini-Asl, R. Keynton, and A. El-Baz, “Alzheimer’s disease diagnostics by adaptation of 3d convolutional network,” in *Image Processing (ICIP), 2016 IEEE International Conference on*, IEEE, 2016, pp. 126–130.
- [37] X. Yang, Q. Wu, D. Hong, and J. Zou, “Spatial regularization for neural network and application in alzheimer’s disease classification,” in *Future Technologies Conference (FTC)*, IEEE, 2016, pp. 831–837.
- [38] B. Mahanand, S. Suresh, N. Sundararajan, and M. A. Kumar, “Alzheimer’s disease detection using a self-adaptive resource allocation network classifier,” in *Neural Networks (IJCNN), The 2011 International Joint Conference on*, IEEE, 2011, pp. 1930–1934.
- [39] G. S. Babu, S. Suresh, and B. Mahanand, “Alzheimer’s disease detection using a projection based learning meta-cognitive rbf network,” in *Neural Networks (IJCNN), The 2012 International Joint Conference on*, IEEE, 2012, pp. 1–8.
- [40] C. D. Billones, O. J. L. D. Demetria, D. E. D. Hostallero, and P. C. Naval, “Demnet: A convolutional neural network for the detection of alzheimer’s disease and mild cognitive impairment,” in *Region 10 Conference (TENCON), 2016 IEEE*, IEEE, 2016, pp. 3724–3727.
- [41] S. Sarraf and G. Tofighi, “Deep learning-based pipeline to recognize alzheimer’s disease using fmri data,” in *Future Technologies Conference (FTC)*, IEEE, 2016, pp. 816–820.
- [42] C. Hu, R. Ju, Y. Shen, P. Zhou, and Q. Li, “Clinical decision support for alzheimer’s disease based on deep learning and brain network,” in *Communications (ICC), 2016 IEEE International Conference on*, IEEE, 2016, pp. 1–6.
- [43] L. Zhou, L. Wang, L. Liu, P. Ogunbona, and D. Shen, “Discriminative brain effective connectivity analysis for alzheimer’s disease: A kernel learning approach upon sparse gaussian bayesian network,” in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, IEEE, 2013, pp. 2243–2250.
- [44] F. S. de Aguiar Neto and J. L. G. Rosa, “Depression biomarkers using non-invasive eeg: A review,” *Neuroscience & Biobehavioral Reviews*, vol. 105, pp. 83–93, 2019.
- [45] V. Knott, C. Mahoney, S. Kennedy, and K. Evans, “Eeg power, frequency, asymmetry and coherence in male depression,” *Psychiatry Research: Neuroimaging*, vol. 106, no. 2, pp. 123–140, 2001.
- [46] M. Bachmann, L. Päeske, K. Kaley, K. Aarma, A. Lehtmetts, P. Ööpik, J. Lass, and H. Hinrikus, “Methods for classifying depression in single channel eeg using linear and nonlinear signal analysis,” *Computer methods and programs in biomedicine*, vol. 155, pp. 11–17, 2018.
- [47] H. Hinrikus, A. Suhhova, M. Bachmann, K. Adamssoo, Ü. Võhma, J. Lass, and V. Tuulik, “Electroencephalographic spectral asymmetry index for detection of depression,” *Medical & biological engineering & computing*, vol. 47, no. 12, p. 1291, 2009.
- [48] M. Bachmann, J. Lass, and H. Hinrikus, “Single channel eeg analysis for detection of depression,” *Biomedical Signal Processing and Control*, vol. 31, pp. 391–397, 2017.
- [49] T. Higuchi, “Approach to an irregular time series on the basis of the fractal theory,” *Physica D: Nonlinear Phenomena*, vol. 31, no. 2, pp. 277–283, 1988.

- [50] M. Bachmann, J. Lass, A. Suhhova, and H. Hinrikus, "Spectral asymmetry and higuchi's fractal dimension measures of depression electroencephalogram," *Computational and mathematical methods in medicine*, vol. 2013, 2013.
- [51] A. Lempel and J. Ziv, "On the complexity of finite sequences," *IEEE Transactions on information theory*, vol. 22, no. 1, pp. 75–81, 1976.
- [52] X.-S. Zhang, R. J. Roy, and E. W. Jensen, "Eeg complexity as a measure of depth of anesthesia for patients," *IEEE transactions on biomedical engineering*, vol. 48, no. 12, pp. 1424–1433, 2001.
- [53] K. Kalev, M. Bachmann, L. Orgo, J. Lass, and H. Hinrikus, "Lempel-ziv and multiscale lempel-ziv complexity in depression," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE, 2015, pp. 4158–4161.
- [54] C.-K. Peng, S. V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley, and A. L. Goldberger, "Mosaic organization of dna nucleotides," *Physical review e*, vol. 49, no. 2, p. 1685, 1994.
- [55] M. Bachmann, A. Suhhova, J. Lass, K. Adamsoo, Ü. Vöhma, and H. Hinrikus, "Detrended fluctuation analysis of eeg in depression," in *XIII Mediterranean Conference on Medical and Biological Engineering and Computing 2013*, Springer, 2014, pp. 694–697.
- [56] A. Berger, "How does it work?: Magnetic resonance imaging," *BMJ: British Medical Journal*, vol. 324, no. 7328, p. 35, 2002.
- [57] K. Coyne. (2012). "Mri: A guided tour." Accessed: 2020-04-06, [Online]. Available: <https://nationalmaglab.org/education/magnet-academy/learn-the-basics/stories/mri-a-guided-tour>.
- [58] I. Despotović, B. Goossens, and W. Philips, "Mri segmentation of the human brain: Challenges, methods, and applications," *Computational and mathematical methods in medicine*, vol. 2015, 2015.
- [59] S. Marchesotti, M. Bassolino, A. Serino, H. Bleuler, and O. Blanke, "Quantifying the role of motor imagery in brain-machine interfaces," *Scientific reports*, vol. 6, no. 1, pp. 1–12, 2016.
- [60] K. E. Sprecher, B. A. Riedner, R. F. Smith, G. Tononi, R. J. Davidson, and R. M. Benca, "High resolution topography of age-related changes in non-rapid eye movement sleep electroencephalography," *PLoS One*, vol. 11, no. 2, e0149770, 2016.
- [61] W. Wang, *Machine Audition: Principles, Algorithms and Systems: Principles, Algorithms and Systems*. IGI Global, 2010.
- [62] R. C. Deo, "Machine learning in medicine," *Circulation*, vol. 132, no. 20, pp. 1920–1930, 2015.
- [63] T. Panch, P. Szolovits, and R. Atun, "Artificial intelligence, machine learning and health systems," *Journal of global health*, vol. 8, no. 2, 2018.
- [64] J. A. Nichols, H. W. H. Chan, and M. A. Baker, "Machine learning: Applications of artificial intelligence to imaging and diagnosis," *Biophysical reviews*, vol. 11, no. 1, pp. 111–118, 2019.
- [65] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics and intelligent laboratory systems*, vol. 2, no. 1-3, pp. 37–52, 1987.
- [66] I. Jolliffe, *Principal Component Analysis*. Springer, 2002.
- [67] W. Wu, D. Massart, and S. De Jong, "The kernel pca algorithms for wide data. part i: Theory and algorithms," *Chemometrics and Intelligent Laboratory Systems*, vol. 36, no. 2, pp. 165–172, 1997.
- [68] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Annals of eugenics*, vol. 7, no. 2, pp. 179–188, 1936.
- [69] A. Subasi and M. I. Gursoy, "Eeg signal classification using pca, ica, lda and support vector machines," *Expert systems with applications*, vol. 37, no. 12, pp. 8659–8666, 2010.
- [70] A. Bhardwaj, A. Gupta, P. Jain, A. Rani, and J. Yadav, "Classification of human emotions from eeg signals using svm and lda classifiers," in *2015 2nd International Conference on Signal Processing and Integrated Networks (SPIN)*, IEEE, 2015, pp. 180–185.

- [71] A. Ali. (2019). "Dimensionality reduction (pca and lda) with practical implementation." Accessed: 2023-08-23, [Online]. Available: <https://medium.com/machine-learning-researcher/dimensionality-reduction-pca-and-lda-6be91734f567>.
- [72] C. M. Bishop and N. M. Nasrabadi, *Pattern recognition and machine learning*, 4. Springer, 2006, vol. 4.
- [73] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural networks*, vol. 61, pp. 85–117, 2015.
- [74] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.
- [75] T. Hastie, R. Tibshirani, and J. Friedman, *The elements of statistical learning: data mining, inference, and prediction*. Springer Science & Business Media, 2009.
- [76] A. Sharmila and P. Geethanjali, "Dwt based detection of epileptic seizure from eeg signals using naive bayes and k-nn classifiers," *Ieee Access*, vol. 4, pp. 7716–7727, 2016.
- [77] T. M. Cover and P. E. Hart, "Nearest neighbor pattern classification," *IEEE transactions on information theory*, vol. 13, no. 1, pp. 21–27, 1967.
- [78] R. O. Duda, P. E. Hart, *et al.*, *Pattern classification*. John Wiley & Sons, 2006.
- [79] M. D. Steenwijk, P. J. Pouwels, M. Daams, J. W. van Dalen, M. W. Caan, E. Richard, F. Barkhof, and H. Vrenken, "Accurate white matter lesion segmentation by k nearest neighbor classification with tissue type priors (knn-ttps)," *NeuroImage: Clinical*, vol. 3, pp. 462–469, 2013.
- [80] L. Breiman and J. Friedman, "Ra olshen and cj stone," *Classification and regression trees*, 1984.
- [81] S. Sun, C. Zhang, and D. Zhang, "An experimental evaluation of ensemble methods for eeg signal classification," *Pattern Recognition Letters*, vol. 28, no. 15, pp. 2157–2163, 2007.
- [82] C. Kingsford and S. L. Salzberg, "What are decision trees?" *Nature biotechnology*, vol. 26, no. 9, p. 1011, 2008.
- [83] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [84] T. Platt, "Probabilistic outputs for support vector machines and comparison to regularized likelihood methods," *Advances in Large Margin Classifiers*, pp. 61–74, 2000.
- [85] N. Cristianini, J. Shawe-Taylor, *et al.*, *An introduction to support vector machines and other kernel-based learning methods*. Cambridge university press, 2000.
- [86] G. Guo, S. Z. Li, and K. L. Chan, "Support vector machines for face recognition," *Image and Vision computing*, vol. 19, no. 9, pp. 631–638, 2001.
- [87] S.-i. Amari and S. Wu, "Improving support vector machine classifiers by modifying kernel functions," *Neural Networks*, vol. 12, no. 6, pp. 783–789, 1999.
- [88] J. M.-M. Melišek and M. O.-J. Pavlovicová, "Support vector machines, pca and lda in face recognition," *J. Electr. Eng.*, vol. 59, pp. 203–209, 2008.
- [89] K. M. Ali and M. J. Pazzani, "Error reduction through learning multiple descriptions," *Machine learning*, vol. 24, no. 3, pp. 173–202, 1996.
- [90] T. G. Dietterich, "Ensemble methods in machine learning," in *International workshop on multiple classifier systems*, Springer, 2000, pp. 1–15.
- [91] K. Abualsaud, M. Mahmuddin, M. Saleh, and A. Mohamed, "Ensemble classifier for epileptic seizure detection for imperfect eeg data," *The Scientific World Journal*, vol. 2015, 2015.
- [92] A. Datta and R. Chatterjee, "Comparative study of different ensemble compositions in eeg signal classification problem," in *Emerging Technologies in Data Mining and Information Security*, Springer, 2019, pp. 145–154.
- [93] R. Polikar, "Ensemble based systems in decision making," *IEEE Circuits and systems magazine*, vol. 6, no. 3, pp. 21–45, 2006.
- [94] L. Breiman, "Bagging predictors," *Machine learning*, vol. 24, no. 2, pp. 123–140, 1996.

- [95] L. Lam and C. Y. Suen, "A theoretical analysis of the application of majority voting to pattern recognition," in *Proceedings of the 12th IAPR International Conference on Pattern Recognition, Vol. 3-Conference C: Signal Processing (Cat. No. 94CH3440-5)*, IEEE, vol. 2, 1994, pp. 418–420.
- [96] K.-W. Hsu, "A theoretical analysis of why hybrid ensembles work," *Computational intelligence and neuroscience*, vol. 2017, 2017.
- [97] M. Mohandes, M. Deriche, and S. O. Aliyu, "Classifiers combination techniques: A comprehensive review," *IEEE Access*, vol. 6, pp. 19 626–19 639, 2018.
- [98] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of computer and system sciences*, vol. 55, no. 1, pp. 119–139, 1997.
- [99] A. Arnx. (2019). "First neural network for beginners explained (with code) - understand and create a perceptron." Accessed: 2023-04-06, [Online]. Available: <https://towardsdatascience.com/first-neural-network-for-beginners-explained-with-code-4cfd37e06eaf>.
- [100] H. J. Kelley, "Gradient theory of optimal flight paths," *Ars Journal*, vol. 30, no. 10, pp. 947–954, 1960.
- [101] S. Narayan, "The generalized sigmoid activation function: Competitive supervised learning," *Information sciences*, vol. 99, no. 1-2, pp. 69–82, 1997.
- [102] A. Namin, K. Leboeuf, R. Muscedere, H. Wu, and M. Ahmadi, "Efficient hardware implementation of the hyperbolic tangent sigmoid function," Jun. 2009, pp. 2117–2120. DOI: 10.1109/ISCAS.2009.5118213.
- [103] A. F. Agarap, "Deep learning using rectified linear units (relu)," *arXiv preprint arXiv:1803.08375*, 2018.
- [104] N. Johnson, P. Vulimiri, A. To, X. Zhang, C. Brice, B. Kappes, and A. Stebner, "Machine learning for materials developments in metals additive manufacturing," *arXiv preprint arXiv:2005.05235*, 2020.
- [105] W. Rammer and R. Seidl, "Harnessing deep learning in ecology: An example predicting bark beetle outbreaks," *Frontiers in plant science*, p. 1327, 2019.
- [106] K. Janocha and W. M. Czarnecki, "On loss functions for deep neural networks in classification," *arXiv preprint arXiv:1702.05659*, 2017.
- [107] X. Zhang, "Regularization," in *Encyclopedia of Machine Learning*, C. Sammut and G. I. Webb, Eds. Boston, MA: Springer US, 2010, pp. 845–849, ISBN: 978-0-387-30164-8. DOI: 10.1007/978-0-387-30164-8_712. [Online]. Available: https://doi.org/10.1007/978-0-387-30164-8_712.
- [108] M. T. Priya Pedamkar. (2023). "Gradient descent in machine learning:- a basic introduction." Accessed: 2023-06-06, [Online]. Available: <https://www.educba.com/gradient-descent-in-machine-learning/>.
- [109] A. Wagh. (2022). "Gradient descent and its types." Accessed: 2023-06-06, [Online]. Available: <https://www.analyticsvidhya.com/blog/2022/07/gradient-descent-and-its-types/>.
- [110] H. Tunanyan. (2021). "Optimization methods - gradient descent." Accessed: 2023-06-06, [Online]. Available: <https://datascience.stackexchange.com/questions/84167/what-is-momentum-in-neural-network>.
- [111] R. Kurban. (2021). "Boost your network performance - hyperparameter tuning and training optimization." Accessed: 2023-06-06, [Online]. Available: <https://towardsdatascience.com/boost-your-network-performance-cc0a2a95c5ef>.
- [112] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *The bulletin of mathematical biophysics*, vol. 5, pp. 115–133, 1943.
- [113] Y. Freund and R. E. Schapire, "Large margin classification using the perceptron algorithm," in *Proceedings of the eleventh annual conference on Computational learning theory*, 1998, pp. 209–217.

- [114] S. E. Karimboyevich and A. O. Nematullayevich, “Single layer artificial neural network: Perceptron,” *Eur. Multidiscip. J. Mod. Sci*, vol. 5, pp. 230–238, 2022.
- [115] Javatpoint. (2021). “Single layer perceptron in tensorflow.” Accessed: 2023-06-06, [Online]. Available: <https://www.javatpoint.com/single-layer-perceptron-in-tensorflow>.
- [116] S. Haykin, *Neural networks: a comprehensive foundation*. Prentice Hall PTR, 1994.
- [117] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, “Densely connected convolutional networks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.
- [118] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “Imagenet classification with deep convolutional neural networks,” in *Proceedings of the 25th International Conference on Neural Information Processing Systems - Volume 1*, ser. NIPS’12, Lake Tahoe, Nevada: Curran Associates Inc., 2012, pp. 1097–1105.
- [119] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” *arXiv preprint arXiv:1409.1556*, 2014.
- [120] A. Sengupta, Y. Ye, R. Wang, C. Liu, and K. Roy, “Going deeper in spiking neural networks: Vgg and residual architectures,” *Frontiers in neuroscience*, vol. 13, p. 95, 2019.
- [121] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, “Going deeper with convolutions,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.
- [122] S. Targ, D. Almeida, and K. Lyman, “Resnet in resnet: Generalizing residual architectures,” *arXiv preprint arXiv:1603.08029*, 2016.
- [123] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, *et al.*, “Imagenet large scale visual recognition challenge,” *International journal of computer vision*, vol. 115, pp. 211–252, 2015.
- [124] ADNI. (2019). “About adni.” Accessed: 02-05-2019, [Online]. Available: <http://adni.loni.usc.edu/about/>.
- [125] —, (2019). “Background & rationale.” Accessed: 02-05-2019, [Online]. Available: <http://adni.loni.usc.edu/study-design/>.
- [126] —, (2019). “Mri pre-processing.” Accessed: 02-05-2019, [Online]. Available: <http://adni.loni.usc.edu/methods/mri-tool/mri-analysis/#mri-pre-processing-container>.
- [127] —, (2019). “Mri pre-processing.” Accessed: 02-05-2019, [Online]. Available: <http://adni.loni.usc.edu/study-design/#study-objective-container>.
- [128] A. Aluoja, J. Shlik, V. Vasar, K. Luuk, and M. Leinsalu, “Development and psychometric properties of the emotional state questionnaire, a self-report questionnaire for depression and anxiety,” *Nordic Journal of Psychiatry*, vol. 53, no. 6, pp. 443–449, 1999.
- [129] P. S. Kumar and V. Dharun, “A study of mri segmentation methods in automatic brain tumor detection,” *Int. J. Eng. Technol*, vol. 8, no. 2, pp. 609–614, 2016.
- [130] R. C. Gonzalez and R. E. Woods, *Digital Image Processing Second Edition*. Upper Saddle River, New Jersey 07458: Prentice-Hall, Inc., 2002.
- [131] G. Anbarjafari, “An objective no-reference measure of illumination assessment,” *Measurement Science Review*, vol. 15, no. 6, pp. 319–322, 2015.
- [132] I. Beheshti, N. Maikusa, H. Matsuda, H. Demirel, and G. Anbarjafari, “Histogram-based feature extraction from individual gray matter similarity-matrix for alzheimer’s disease classification,” *Journal of Alzheimer’s Disease*, vol. 55, no. 4, pp. 1571–1582, 2017.
- [133] I. Beheshti, N. Maikusa, M. Daneshmand, H. Matsuda, H. Demirel, and G. Anbarjafari, “Classification of alzheimer’s disease and prediction of mild cognitive impairment conversion using histogram-based analysis of patient-specific anatomical brain connectivity networks,” *Journal of Alzheimer’s Disease*, vol. 60, no. 1, pp. 295–304, 2017.

- [134] J. Lu, V. E. Liong, X. Zhou, and J. Zhou, "Learning compact binary face descriptor for face recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 37, no. 10, pp. 2041–2056, 2015.
- [135] S. Shojaeilangari, W.-Y. Yau, and E.-K. Teoh, "Pose-invariant descriptor for facial emotion recognition," *Machine Vision and Applications*, vol. 27, no. 7, pp. 1063–1070, 2016.
- [136] C. Turan and K.-M. Lam, "Histogram-based local descriptors for facial expression recognition (fer): A comprehensive study," *Journal of visual communication and image representation*, vol. 55, pp. 331–341, 2018.
- [137] T. S. Kumar, K. N. Rajesh, S. Maheswari, V. Kanhangad, and U. R. Acharya, "Automated schizophrenia detection using local descriptors with eeg signals," *Engineering Applications of Artificial Intelligence*, vol. 117, p. 105 602, 2023.
- [138] D. Sarwinda and A. Bustamam, "Detection of alzheimer's disease using advanced local binary pattern from hippocampus and whole brain of mr images," in *2016 International Joint Conference on Neural Networks (IJCNN)*, IEEE, 2016, pp. 5051–5056.
- [139] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 24, no. 7, pp. 971–987, 2002.
- [140] M. P. T Ojala and M. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification width local binary patterns," *Transactions on Pattern Analysis and Machine Intelligence*, pp. 971–987, 2002.
- [141] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, Ieee, vol. 1, 2005, pp. 886–893.
- [142] A. Bosch, A. Zisserman, and X. Munoz, "Representing shape with a spatial pyramid kernel," in *Proceedings of the 6th ACM international conference on Image and video retrieval*, 2007, pp. 401–408.
- [143] E. Kaplan, E. Altunisik, Y. E. Firat, P. D. Barua, S. Dogan, M. Baygin, F. B. Demir, T. Tuncer, E. Palmer, R.-S. Tan, *et al.*, "Novel nested patch-based feature extraction model for automated parkinson's disease symptom classification using mri images," *Computer Methods and Programs in Biomedicine*, vol. 224, p. 107 030, 2022.
- [144] V. Ojansivu and J. Heikkilä, "Blur insensitive texture classification using local phase quantization," in *Image and Signal Processing: 3rd International Conference, ICISP 2008. Cherbourg-Octeville, France, July 1-3, 2008. Proceedings 3*, Springer, 2008, pp. 236–243.
- [145] A. Dhall, A. Asthana, R. Goecke, and T. Gedeon, "Emotion recognition using phog and lpq features," in *2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG)*, IEEE, 2011, pp. 878–883.
- [146] G. Macin, B. Tasci, I. Tasci, O. Faust, P. D. Barua, S. Dogan, T. Tuncer, R.-S. Tan, and U. R. Acharya, "An accurate multiple sclerosis detection model based on exemplar multiple parameters local phase quantization: Exmplpq," *Applied Sciences*, vol. 12, no. 10, p. 4920, 2022.
- [147] T. Mohammad and M. L. Ali, "Robust facial expression recognition based on local monotonic pattern (lmp)," in *14th International Conference on Computer and Information Technology (ICCIT 2011)*, IEEE, 2011, pp. 572–576.
- [148] T. Jabid and O. Chae, "Local transitional pattern: A robust facial image descriptor for automatic facial expression recognition," in *Proceedings of the International Conference on Computer Convergence Technology, Seoul, Korea, 2011*, pp. 333–344.
- [149] T. Jabid and O. Chae, "Facial expression recognition based on local transitional pattern," *International Information Institute (Tokyo). Information*, vol. 15, no. 5, p. 2007, 2012.
- [150] M. Yang, L. Zhang, S. C.-K. Shiu, and D. Zhang, "Monogenic binary coding: An efficient local feature extraction approach to face recognition," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 6, pp. 1738–1751, 2012.

- [151] J. Lu, V. E. Liong, and J. Zhou, “Cost-sensitive local binary feature learning for facial age estimation,” *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 5356–5368, 2015.
- [152] L. Zhou and H. Wang, “Local gradient increasing pattern for facial expression recognition,” in *2012 19th IEEE International Conference on Image Processing*, IEEE, 2012, pp. 2601–2604.
- [153] A. R. Rivera, J. R. Castillo, and O. O. Chae, “Local directional number pattern for face analysis: Face and expression recognition,” *IEEE transactions on image processing*, vol. 22, no. 5, pp. 1740–1752, 2012.
- [154] F. Ahmed, “Gradient directional pattern: A robust feature descriptor for facial expression recognition,” *Electronics letters*, vol. 48, no. 19, pp. 1203–1204, 2012.
- [155] W. Chu, “Facial expression recognition based on local binary pattern and gradient directional pattern,” in *2013 IEEE International Conference on Green Computing and Communications and IEEE Internet of Things and IEEE Cyber, Physical and Social Computing*, IEEE, 2013, pp. 1458–1462.
- [156] R. V. Rao and T. Prasad, “A new optimized hybrid local lifting wavelet co-occurrence texture pattern for content based medical image retrieval.,” *International Journal of Online & Biomedical Engineering*, vol. 17, no. 11, 2021.
- [157] S. Li, D. Gong, and Y. Yuan, “Face recognition using weber local descriptors,” *Neurocomputing*, vol. 122, pp. 272–283, 2013.
- [158] S. Liu, Y. Zhang, and K. Liu, “Facial expression recognition under partial occlusion based on weber local descriptor histogram and decision fusion,” in *Proceedings of the 33rd Chinese Control Conference*, IEEE, 2014, pp. 4664–4668.
- [159] M. S. Islam and S. Auwatanamo, “Facial expression recognition using local arc pattern,” *Trends Appl. Sci. Res*, vol. 9, no. 2, p. 113, 2014.
- [160] M. S. Islam, “Local gradient pattern-a novel feature representation for facial expression recognition,” *Journal of AI and Data Mining*, vol. 2, no. 1, pp. 33–38, 2014.
- [161] F. Bashar, A. Khan, F. Ahmed, and M. H. Kabir, “Robust facial expression recognition based on median ternary pattern (mtp),” in *2013 International Conference on Electrical Information and Communication Technology (EICT)*, IEEE, 2014, pp. 1–5.
- [162] A. R. Rivera, J. R. Castillo, and O. Chae, “Local directional texture pattern image descriptor,” *Pattern Recognition Letters*, vol. 51, pp. 94–100, 2015.
- [163] B.-Q. Yang, T. Zhang, C.-C. Gu, K.-J. Wu, and X.-P. Guan, “A novel face recognition method based on iwld and iwbc,” *Multimedia Tools and Applications*, vol. 75, pp. 6979–7002, 2016.
- [164] H. Greenspan, B. Van Ginneken, and R. M. Summers, “Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique,” *IEEE transactions on medical imaging*, vol. 35, no. 5, pp. 1153–1159, 2016.
- [165] A. Ashraf, S. Naz, S. H. Shirazi, I. Razzak, and M. Parsad, “Deep transfer learning for alzheimer neurological disorder detection,” *Multimedia Tools and Applications*, pp. 1–26, 2021.
- [166] H.-I. Suk, S.-W. Lee, D. Shen, A. D. N. Initiative, *et al.*, “Hierarchical feature representation and multimodal fusion with deep learning for ad/mci diagnosis,” *NeuroImage*, vol. 101, pp. 569–582, 2014.
- [167] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, “Imagenet: A large-scale hierarchical image database,” in *2009 IEEE conference on computer vision and pattern recognition*, Ieee, 2009, pp. 248–255.
- [168] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the inception architecture for computer vision,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818–2826.
- [169] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

- [170] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "Squeezenet: Alexnet-level accuracy with 50x fewer parameters and < 0.5 mb model size," *arXiv preprint arXiv:1602.07360*, 2016.
- [171] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 31, 2017.
- [172] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 1251–1258.
- [173] J. Redmon and A. Farhadi, "Yolo9000: Better, faster, stronger," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 7263–7271.
- [174] X. Zhang, X. Zhou, M. Lin, and J. Sun, "Shufflenet: An extremely efficient convolutional neural network for mobile devices," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 6848–6856.
- [175] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, "Learning transferable architectures for scalable image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 8697–8710.
- [176] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 4510–4520.
- [177] M. Tan and Q. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," in *International conference on machine learning*, PMLR, 2019, pp. 6105–6114.
- [178] N. I. of Health (NIH). (2019). "Neuroimaging informatics technology initiative." Accessed: 25-04-2019, [Online]. Available: <https://nifti.nimh.nih.gov/>.
- [179] S. S. Gudadhe, A. D. Thakare, and D. Oliva, "Classification of intracranial hemorrhage ct images based on texture analysis using ensemble-based machine learning algorithms: A comparative study," *Biomedical Signal Processing and Control*, vol. 84, p. 104832, 2023.
- [180] B. F. Stanley and S. W. Franklin, "Effective feature extraction for cerebral microbleed detection using edge emphasized weber maximum directional co-occurrence matrix," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–14, 2022.
- [181] A. Saif, T. Imtiaz, C. Shahnaz, W.-P. Zhu, and M. O. Ahmad, "Exploiting cascaded ensemble of features for the detection of tuberculosis using chest radiographs," *IEEE Access*, vol. 9, pp. 112 388–112 399, 2021.
- [182] M. Bramarambika, "Brain tumor classification for mr images using hybrid glcm-ldtp-le-net feature extraction and bi- lstm model.," *International Journal of Intelligent Engineering & Systems*, vol. 15, no. 2, 2022.
- [183] M. W. Mwadulo, "Alocal directional ternary pattern texture descriptor for mammographic breast cancer classification," Ph.D. dissertation, MMUST, 2020.
- [184] D. S. Uplaonkar, Virupakshappa, and N. Patil, "Modified otsu thresholding based level set and local directional ternary pattern technique for liver tumor segmentation," *International Journal of System Assurance Engineering and Management*, pp. 1–11, 2022.
- [185] W. A. Hassan, Y. H. Ali, and N. J. Ibrahim, "A survey of latest techniques in medical image classification," in *2021 International Conference on Communication & Information Technology (ICICT)*, IEEE, 2021, pp. 68–73.
- [186] R. A. A. Al-falluji, "Mri based techniques for detection of alzheimer: A survey," *International Journal of Computer Applications*, vol. 159, no. 5, pp. 20–24, 2017.
- [187] R. R. Khasawneh, E. Abu-El-Rub, A. Alzu'bi, G. T. Abdelhady, and H. S. Al-Soudi, "Corpus callosum anatomical changes in alzheimer patients and the effect of acetylcholinesterase inhibitors on corpus callosum morphometry," *PLoS One*, vol. 17, no. 7, e0269082, 2022.
- [188] S. Lu, Z. Lu, and Y.-D. Zhang, "Pathological brain detection based on alexnet and transfer learning," *Journal of computational science*, vol. 30, pp. 41–47, 2019.

- [189] L. S. Kumar, S. Hariharasitaraman, K. Narayanasamy, K. Thinakaran, J. Mahalakshmi, and V. Pandimurugan, "Alexnet approach for early stage alzheimer's disease detection from mri brain images," *Materials Today: Proceedings*, vol. 51, pp. 58–65, 2022.
- [190] S. Deepak and P. Ameer, "Retrieval of brain mri with tumor using contrastive loss based similarity on googlenet encodings," *Computers in biology and medicine*, vol. 125, p. 103 993, 2020.
- [191] S. Ahuja, B. K. Panigrahi, and T. K. Gandhi, "Enhanced performance of dark-nets for brain tumor classification and segmentation using colormap-based superpixel techniques," *Machine Learning with Applications*, vol. 7, p. 100 212, 2022.
- [192] A. Farooq, S. Anwar, M. Awais, and S. Rehman, "A deep cnn based multi-class classification of alzheimer's disease using mri," in *2017 IEEE International Conference on Imaging systems and techniques (IST)*, IEEE, 2017, pp. 1–6.
- [193] Y. Eroglu, M. Yildirim, and A. Cinar, "Mrmr-based hybrid convolutional neural network model for classification of alzheimer's disease on brain magnetic resonance images," *International Journal of Imaging Systems and Technology*, vol. 32, no. 2, pp. 517–527, 2022.
- [194] J. Li, P. Wang, Y. Li, Y. Zhou, X. Liu, and K. Luan, "Transfer learning of pre-trained inception-v3 model for colorectal cancer lymph node metastasis classification," in *2018 IEEE International Conference on Mechatronics and Automation (ICMA)*, IEEE, 2018, pp. 1650–1654.
- [195] M. Shoaib and N. Sayed, "Yolo object detector and inception-v3 convolutional neural network for improved brain tumor segmentation..," *Traitement Du Signal*, vol. 39, no. 1, 2022.
- [196] R. Singh, N. Sharma, and R. Gupta, "Detection of alzheimer's risk level using inception v3 transfer learning model," in *2023 International Conference on Distributed Computing and Electrical Circuits and Electronics (ICDCECE)*, IEEE, 2023, pp. 1–6.
- [197] P. Li, W. Quan, Z. Wang, Y. Liu, H. Cai, Y. Chen, Y. Wang, M. Zhang, Z. Tian, H. Zhang, et al., "Early-stage differentiation between alzheimer's disease and frontotemporal lobe degeneration: Clinical, neuropsychology, and neuroimaging features," *Frontiers in Aging Neuroscience*, vol. 14, p. 981 451, 2022.
- [198] M. Lindau, O. Almkvist, J. Kushi, K. Boone, S. Johansson, L. Wahlund, J. Cummings, and B. Miller, "First symptoms—frontotemporal dementia versus alzheimer's disease," *Dementia and geriatric cognitive disorders*, vol. 11, no. 5, pp. 286–293, 2000.
- [199] X. Zhu, H.-I. Suk, Y. Zhu, K.-H. Thung, G. Wu, and D. Shen, "Multi-view classification for identification of alzheimer's disease," in *Machine Learning in Medical Imaging: 6th International Workshop, MLMI 2015, Held in Conjunction with MICCAI 2015, Munich, Germany, October 5, 2015, Proceedings 6*, Springer, 2015, pp. 255–262.
- [200] M. Saim and A. Feroui, "A new hybrid method based on bias-correction fuzzy c means and histogram of oriented gradient for alzheimer disease detection," in *2022 First International Conference on Computer Communications and Intelligent Systems (I3CIS)*, IEEE, 2022, pp. 31–36.
- [201] E. Kaplan, M. Baygin, P. D. Barua, S. Dogan, T. Tuncer, E. Altunisik, E. E. Palmer, and U. R. Acharya, "Exhif: Alzheimer's disease detection using exemplar histogram-based features with ct and mr images," *Medical Engineering & Physics*, vol. 115, p. 103 971, 2023.
- [202] D. Chan, N. C. Fox, R. I. Scahill, W. R. Crum, J. L. Whitwell, G. Leschziner, A. M. Rossor, J. M. Stevens, L. Cipolotti, and M. N. Rossor, "Patterns of temporal lobe atrophy in semantic dementia and alzheimer's disease," *Annals of neurology*, vol. 49, no. 4, pp. 433–442, 2001.
- [203] C. J. Galton, K. Patterson, K. Graham, M. A. Lambon-Ralph, G. Williams, N. Antoun, B. Sahakian, and J. Hodges, "Differing patterns of temporal atrophy in alzheimer's disease and semantic dementia," *Neurology*, vol. 57, no. 2, pp. 216–225, 2001.
- [204] L. Ferrarini, W. M. Palm, H. Olofsen, M. A. van Buchem, J. H. Reiber, and F. Admiraal-Behloul, "Shape differences of the brain ventricles in alzheimer's disease," *Neuroimage*, vol. 32, no. 3, pp. 1060–1069, 2006.

- [205] R. Duara, C. Grady, J. Haxby, M. Sundaram, N. Cutler, L. Heston, A. Moore, N. Schlageter, S. Larson, and S. I. Rapoport, "Positron emission tomography in alzheimer's disease," *Neurology*, vol. 36, no. 7, pp. 879–879, 1986.
- [206] P. Bartolomeo, G. Dalla Barba, M.-F. Boissé, A.-C. Bachoud-Lévi, J.-D. Degos, and F. Boller, "Right-side neglect in alzheimer's disease," *Neurology*, vol. 51, no. 4, pp. 1207–1209, 1998.
- [207] F. Kumfor, R. Landin-Romero, E. Devenney, R. Hutchings, R. Grasso, J. R. Hodges, and O. Piguet, "On the right side? A longitudinal study of left- versus right-lateralized semantic dementia," *Brain*, vol. 139, no. 3, pp. 986–998, Jan. 2016, ISSN: 0006-8950. DOI: 10 . 1093 / brain / awv387. [Online]. Available: <https://doi.org/10.1093/brain/awv387>.
- [208] W. Kang, L. Lin, B. Zhang, X. Shen, S. Wu, A. D. N. Initiative, *et al.*, "Multi-model and multi-slice ensemble learning architecture based on 2d convolutional neural networks for alzheimer's disease diagnosis," *Computers in Biology and Medicine*, vol. 136, p. 104678, 2021.
- [209] J. Jang and D. Hwang, "M3t: Three-dimensional medical image classifier using multi-plane and multi-slice transformer," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 20718–20729.
- [210] L. Chen, H. Qiao, and F. Zhu, "Alzheimer's disease diagnosis with brain structural mri using multiview-slice attention and 3d convolution neural network," *Frontiers in Aging Neuroscience*, vol. 14, p. 871706, 2022.
- [211] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri, "Learning spatiotemporal features with 3d convolutional networks," in *Computer Vision (ICCV), 2015 IEEE International Conference on*, IEEE, 2015, pp. 4489–4497.
- [212] H. Mzoughi, I. Njeh, A. Wali, M. B. Slima, A. BenHamida, C. Mhiri, and K. B. Mahfoudhe, "Deep multi-scale 3d convolutional neural network (cnn) for mri gliomas brain tumor classification," *Journal of Digital Imaging*, vol. 33, pp. 903–915, 2020.
- [213] B. Khagi and G.-R. Kwon, "3d cnn design for the classification of alzheimer's disease using brain mri and pet," *IEEE Access*, vol. 8, pp. 217830–217847, 2020.
- [214] P. Agrawal, N. Katal, and N. Hooda, "Segmentation and classification of brain tumor using 3d-unet deep neural networks," *International Journal of Cognitive Computing in Engineering*, vol. 3, pp. 199–210, 2022.
- [215] C. J. Murray, A. D. Lopez, W. H. Organization, *et al.*, *The global burden of disease: a comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020: summary*. World Health Organization, 1996.
- [216] H.-U. Wittchen, F. Jacobi, J. Rehm, A. Gustavsson, M. Svensson, B. Jönsson, J. Olesen, C. Allgulander, J. Alonso, C. Faravelli, *et al.*, "The size and burden of mental disorders and other disorders of the brain in europe 2010," *European neuropsychopharmacology*, vol. 21, no. 9, pp. 655–679, 2011.
- [217] M. Hamilton, "A rating scale for depression," *Journal of neurology, neurosurgery, and psychiatry*, vol. 23, no. 1, p. 56, 1960.
- [218] T. Tombaugh, I. McDowell, B. Kristjansson, and A. Hubble, "Mini-mental state examination (mmse) and the modified mmse (3ms): A psychometric comparison and normative data," *Psychological Assessment*, vol. 8, no. 1, p. 48, 1996.
- [219] A. Beck, R. Steer, and G. Brown, "Manual for the beck depression inventory-ii. 1996," *San Antonio, TX: Psychological Corporation*, vol. 2, 1996.
- [220] R. Mowbray, "The hamilton rating scale for depression: A factor analysis," *Psychological Medicine*, vol. 2, no. 3, pp. 272–280, 1972.
- [221] H. Cai, J. Han, Y. Chen, X. Sha, Z. Wang, B. Hu, J. Yang, L. Feng, Z. Ding, Y. Chen, *et al.*, "A pervasive approach to eeg-based depression detection," *Complexity*, vol. 2018, 2018.
- [222] S. Mahato and S. Paul, "Detection of major depressive disorder using linear and non-linear features from eeg signals," *Microsystem Technologies*, vol. 25, no. 3, pp. 1065–1076, 2019.

- [223] ———, “Electroencephalogram (eeg) signal analysis for diagnosis of major depressive disorder (mdd): A review,” in *Nanoelectronics, Circuits and Communication Systems*, Springer, 2019, pp. 323–335.
- [224] A. Khosla, P. Khandnor, and T. Chand, “A comparative analysis of signal processing and classification methods for different applications based on eeg signals,” *Biocybernetics and Biomedical Engineering*, 2020.
- [225] H. Hinrikus, A. Suhhova, M. Bachmann, K. Aadamsoo, Ü. Vöhma, H. Pehlak, and J. Lass, “Spectral features of eeg in depression,” *Biomedical Engineering/Biomedizinische Technik*, vol. 55, no. 3, pp. 155–161, 2010.
- [226] A. J. Ibanez-Molina, S. Iglesias-Parro, M. F. Soriano, and J. I. Aznarte, “Multiscale lempel–ziv complexity for eeg measures,” *Clinical Neurophysiology*, vol. 126, no. 3, pp. 541–548, 2015.
- [227] C.-K. Peng, S. Havlin, H. E. Stanley, and A. L. Goldberger, “Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series,” *Chaos: an interdisciplinary journal of nonlinear science*, vol. 5, no. 1, pp. 82–87, 1995.
- [228] Mathworks. (2021). “Univariate feature ranking for regression using f-tests.” Accessed: 2021-01-01, [Online]. Available: <https://se.mathworks.com/help/stats/fsrftest.html>.
- [229] M. Robnik-Šikonja and I. Kononenko, “Theoretical and empirical analysis of relieff and relieff,” *Machine learning*, vol. 53, no. 1-2, pp. 23–69, 2003.
- [230] M. R. Lakshmi, T. Prasad, and D. V. C. Prakash, “Survey on eeg signal processing methods,” *International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 4, no. 1, 2014.
- [231] T. N. Alotaiby, S. A. Alshebeili, T. Alshawi, I. Ahmad, and F. E. Abd El-Samie, “Eeg seizure detection and prediction algorithms: A survey,” *EURASIP Journal on Advances in Signal Processing*, vol. 2014, no. 1, p. 183, 2014.

ACKNOWLEDGEMENTS

This study is supported by the "Estonian Centre of Excellence in IT (EXCITE) (TK148)" (1.01.2016-1.03.2023) project funded by Archimedes Foundation.

Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012). ADNI is funded by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and through generous contributions from the following: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research is providing funds to support ADNI clinical sites in Canada. Private sector contributions are facilitated by the Foundation for the National Institutes of Health (www.fnih.org). The grantee organization is the Northern California Institute for Research and Education, and the study is coordinated by the Alzheimer's Therapeutic Research Institute at the University of Southern California. ADNI data are disseminated by the Laboratory for Neuro Imaging at the University of Southern California.

Tallinn University of Technology (TalTech) supported the electroencephalogram (EEG) study by providing EEG data and questionnaires from consenting patients for research purposes. The gathering of data was carried out by TalTech in accordance with the Declaration of Helsinki, and the process was formally approved by the Tallinn Medical Research Ethics Committee.

SISUKOKKUVÕTE

Aju ebanormaalsuse tuvastamine individuaalsete struktuuriliste ühenduvuse võrkude ja EEG signaali statistilise analüüsi abil

Aju kõrvalekalded avalduvad anatoomiliste või funktsionaalsete muutustena, mis võivad ilmned a sünnist saati või tekkida hiljem trauma, haiguse või muude asjaolude tõttu. Need esinevad muutustena konkreetsete ajupiirkondade suuruses ja struktuuris ning neuronaaelses aktiivsuses või ühenduvuses. See doktoritöö käsitleb Alzheimeri tõve ja kliinilise depressiooniga seotud aju kõrvalekaldeid, mis on mõõdetavad tänapäevaste ajukuvameetoditega (MRI ja EEG). Doktoritöö tutvustab esialgu lugejale Alzheimeri tõve ja kliinilise depressiooni eripärasid, seonduvaid töid, annab ülevaate MRI ja EEG tehnoloogiatest, masinõppest, andmekogumiku kirjeldusest ning jätkub kolme peatükiga, mis kirjeldavad läbi viidud uuringuid.

Uurimuses kasutati MRI ajukuvapilte ADNI andmebaasist ning aju mustrite muutused jäädvustati histogrammidena. Neid histograme kasutati terviklike kognitiivsete võimetega ja Alzheimeri tõvega inimeste eristamiseks mitmete klassifikaatorite kaudu. Esimene uuring näitas, et ühendades konkreetsete lõikude histograme ja kasutades häälteenamuse algoritmi, on võimalik MRI piltide alusel lõike klassifitseerida. Hiljem laiendati uuringut laiendati histogrammi deskriptorite ja eeltreenitud närvivõrgustike närvivõrkude kasutamisega, et tunnused eraldada, ning uuriti Alzheimeri tõve klassifitseerimist uuriti kolme MRI tasandiga. Parimad klassifitseerimistulemused näitasid ka igas tasapinnas asuvaid optimaalsemaid lõike asukohti.

Järgmine uuring näitas, kuidas EEG salvestustest eraldatud lineaarseid ja mitelinearseid tunnuseid (saadud Tallinna Tehnikaülikoolist) kasutati kogumi konfiguratsioonis kliinilise depressiooniga patsientide klassifitseerimiseks. Tunnuste valiku meetodite — ühemõõtmeliste tunnuste järjestamine (univariate feature ranking), häälteenus (majority voting) ja klassifikaatorite kaalutud kogum (weighted ensemble of classifiers) — abil saavutati järjepidevalt head tulemused. Doktoritöö tulemused näitavad, et masinõpet ja EEG andmeid on võimalik kasutada neuroloogiliste haiguste täpseks tuvastamiseks ja klassifitseerimiseks. Väljatöötatud lähenemisviisi täiendamiseks kinnitamiseks on vaja rohkem kliiniliselt usaldusväärseid andmeid.

PUBLICATIONS

CURRICULUM VITAE

Personal data

Name: Egils Avots
Birth: 1991, Latvia
Citizenship: Latvian
Languages: Latvian, English
Address: Narva mnt 18, 51009 Tartu, Estonia
Contact: egils@ut.ee

Education

2017– Doctor of philosophy – Engineering and Technology – Faculty of Science and Technology, University of Tartu
2015–2017 Master of science in engineering – Robotics and computer engineering – Faculty of Science and Technology, University of Tartu
2013–2015 Master’s degree of engineering sciences in electronic – Ventspils University College
2010–2013 Bachelor’s degree of engineering sciences in electronic – Ventspils University College

Employment

2017– University of Tartu, Tartu, Estonia
2014–2015 Ventspils University College, Ventspils, Latvia
2013–2014 SIA "Salons AIS", internship, Riga, Latvia

Scientific work

Main fields of interest:

- Image processing
- Pattern recognition
- Machine learning

ELULOOKIRJELDUS

Isikuandmed

Nimi: Egils Avots
Sünniaeg ja -koht: 1991, Läti
Kodakondsus: lätlane
Keelteoskus: läti, inglise
Aadress: Narva mnt 18, 51009 Tartu, Eesti
Kontaktandmed: egils@ut.ee

Haridus

2017– Filosoofiadoktor – Teadus ja Tehnoloogia – Tartu Ülikool
2015–2017 Tehnikateaduste magister - Arvutitehnika ja robotika – Tartu Ülikool
2013–2015 Elektroonikateaduste magistrikraad – Ventspilsi Ülikooli Kolledž
2010–2013 Elektroonikateaduste bakalaureusekraad – Ventspilsi Ülikooli Kolledž

Teenistuskäik

2017– Tartu Ülikool, Tartu, Eesti
2014–2015 Ventspilsi Ülikooli Kolledž, Ventspils, Läti
2013–2014 SIA "Salons AIS", praktika, Riia, Läti

Teadustegevus

Peamised uurimisvaldkonnad:

- Pildi töötlemine
- Muustrituvastus
- Masinõpe

DISSERTATIONES TECHNOLOGIAE UNIVERSITATIS TARTUENSIS

1. **Imre Mäger.** Characterization of cell-penetrating peptides: Assessment of cellular internalization kinetics, mechanisms and bioactivity. Tartu 2011, 132 p.
2. **Taavi Lehto.** Delivery of nucleic acids by cell-penetrating peptides: application in modulation of gene expression. Tartu 2011, 155 p.
3. **Hannes Luidalepp.** Studies on the antibiotic susceptibility of *Escherichia coli*. Tartu 2012, 111 p.
4. **Vahur Zadin.** Modelling the 3D-microbattery. Tartu 2012, 149 p.
5. **Janno Torop.** Carbide-derived carbon-based electromechanical actuators. Tartu 2012, 113 p.
6. **Julia Suhorutšenko.** Cell-penetrating peptides: cytotoxicity, immunogenicity and application for tumor targeting. Tartu 2012, 139 p.
7. **Viktoryia Shyp.** G nucleotide regulation of translational GTPases and the stringent response factor RelA. Tartu 2012, 105 p.
8. **Mardo Kõivomägi.** Studies on the substrate specificity and multisite phosphorylation mechanisms of cyclin-dependent kinase Cdk1 in *Saccharomyces cerevisiae*. Tartu, 2013, 157 p.
9. **Liis Karo-Astover.** Studies on the Semliki Forest virus replicase protein nsP1. Tartu, 2013, 113 p.
10. **Piret Arukuusk.** NickFects—novel cell-penetrating peptides. Design and uptake mechanism. Tartu, 2013, 124 p.
11. **Piret Villo.** Synthesis of acetogenin analogues. Asymmetric transfer hydrogenation coupled with dynamic kinetic resolution of α -amido- β -keto esters. Tartu, 2013, 151 p.
12. **Villu Kasari.** Bacterial toxin-antitoxin systems: transcriptional cross-activation and characterization of a novel *mqsRA* system. Tartu, 2013, 108 p.
13. **Margus Varjak.** Functional analysis of viral and host components of alpha-virus replicase complexes. Tartu, 2013, 151 p.
14. **Liane Viru.** Development and analysis of novel alphavirus-based multi-functional gene therapy and expression systems. Tartu, 2013, 113 p.
15. **Kent Langel.** Cell-penetrating peptide mechanism studies: from peptides to cargo delivery. Tartu, 2014, 115 p.
16. **Rauno Temmer.** Electrochemistry and novel applications of chemically synthesized conductive polymer electrodes. Tartu, 2014, 206 p.
17. **Indrek Must.** Ionic and capacitive electroactive laminates with carbonaceous electrodes as sensors and energy harvesters. Tartu, 2014, 133 p.
18. **Veiko Voolaid.** Aquatic environment: primary reservoir, link, or sink of antibiotic resistance? Tartu, 2014, 79 p.
19. **Kristiina Laanemets.** The role of SLAC1 anion channel and its upstream regulators in stomatal opening and closure of *Arabidopsis thaliana*. Tartu, 2015, 115 p.

20. **Kalle Pärn.** Studies on inducible alphavirus-based antitumour strategy mediated by site-specific delivery with activatable cell-penetrating peptides. Tartu, 2015, 139 p.
21. **Anastasia Selyutina.** When biologist meets chemist: a search for HIV-1 inhibitors. Tartu, 2015, 172 p.
22. **Sirle Saul.** Towards understanding the neurovirulence of Semliki Forest virus. Tartu, 2015, 136 p.
23. **Marit Orav.** Study of the initial amplification of the human papillomavirus genome. Tartu, 2015, 132 p.
24. **Tormi Reinson.** Studies on the Genome Replication of Human Papillomaviruses. Tartu, 2016, 110 p.
25. **Mart Ustav Jr.** Molecular Studies of HPV-18 Genome Segregation and Stable Replication. Tartu, 2016, 152 p.
26. **Margit Mutso.** Different Approaches to Counteracting Hepatitis C Virus and Chikungunya Virus Infections. Tartu, 2016, 184 p.
27. **Jelizaveta Geimanen.** Study of the Papillomavirus Genome Replication and Segregation. Tartu, 2016, 168 p.
28. **Mart Toots.** Novel Means to Target Human Papillomavirus Infection. Tartu, 2016, 173 p.
29. **Kadi-Liis Veiman.** Development of cell-penetrating peptides for gene delivery: from transfection in cell cultures to induction of gene expression *in vivo*. Tartu, 2016, 136 p.
30. **Ly Pärnaste.** How, why, what and where: Mechanisms behind CPP/cargo nanocomplexes. Tartu, 2016, 147 p.
31. **Age Utt.** Role of alphavirus replicase in viral RNA synthesis, virus-induced cytotoxicity and recognition of viral infections in host cells. Tartu, 2016, 183 p.
32. **Veiko Vunder.** Modeling and characterization of back-relaxation of ionic electroactive polymer actuators. Tartu, 2016, 154 p.
33. **Piia Kivipõld.** Studies on the Role of Papillomavirus E2 Proteins in Virus DNA Replication. Tartu, 2016, 118 p.
34. **Liina Jakobson.** The roles of abscisic acid, CO₂, and the cuticle in the regulation of plant transpiration. Tartu, 2017, 162 p.
35. **Helen Isok-Paas.** Viral-host interactions in the life cycle of human papillomaviruses. Tartu, 2017, 158 p.
36. **Hanna Hõrak.** Identification of key regulators of stomatal CO₂ signalling via O₃-sensitivity. Tartu, 2017, 260 p.
37. **Jekaterina Jevtuševskaja.** Application of isothermal amplification methods for detection of *Chlamydia trachomatis* directly from biological samples. Tartu, 2017, 96 p.
38. **Ülar Allas.** Ribosome-targeting antibiotics and mechanisms of antibiotic resistance. Tartu, 2017, 152 p.
39. **Anton Paier.** Ribosome Degradation in Living Bacteria. Tartu, 2017, 108 p.
40. **Vallo Varik.** Stringent Response in Bacterial Growth and Survival. Tartu, 2017, 101 p.

41. **Pavel Kudrin.** In search for the inhibitors of *Escherichia coli* stringent response factor RelA. Tartu, 2017, 138 p.
42. **Liisi Henno.** Study of the human papillomavirus genome replication and oligomer generation. Tartu, 2017, 144 p.
43. **Katrin Krõlov.** Nucleic acid amplification from crude clinical samples exemplified by *Chlamydia trachomatis* detection in urine. Tartu, 2018, 118 p.
44. **Eve Sankovski.** Studies on papillomavirus transcription and regulatory protein E2. Tartu, 2018, 113 p.
45. **Morteza Daneshmand.** Realistic 3D Virtual Fitting Room. Tartu, 2018, 233 p.
46. **Fatemeh Noroozi.** Multimodal Emotion Recognition Based Human-Robot Interaction Enhancement. Tartu, 2018, 113 p.
47. **Krista Freimann.** Design of peptide-based vector for nucleic acid delivery in vivo. Tartu, 2018, 103 p.
48. **Rainis Venta.** Studies on signal processing by multisite phosphorylation pathways of the *S. cerevisiae* cyclin-dependent kinase inhibitor Sic1. Tartu, 2018, 155 p.
49. **Inga Põldsalu.** Soft actuators with ink-jet printed electrodes. Tartu, 2018, 85 p.
50. **Kadri Künnapuu.** Modification of the cell-penetrating peptide PepFect14 for targeted tumor gene delivery and reduced toxicity. Tartu, 2018, 114 p.
51. **Toomas Mets.** RNA fragmentation by MazF and MqsR toxins of *Escherichia coli*. Tartu, 2019, 119 p.
52. **Kadri Tõldsepp.** The role of mitogen-activated protein kinases MPK4 and MPK12 in CO₂-induced stomatal movements. Tartu, 2019, 259 p.
53. **Pirko Jalakas.** Unravelling signalling pathways contributing to stomatal conductance and responsiveness. Tartu, 2019, 120 p.
54. **S. Sunjai Nakshatharan.** Electromechanical modelling and control of ionic electroactive polymer actuators. Tartu, 2019, 165 p.
55. **Eva-Maria Tombak.** Molecular studies of the initial amplification of the oncogenic human papillomavirus and closely related nonhuman primate papillomavirus genomes. Tartu, 2019, 150 p.
56. **Meeri Visnapuu.** Design and physico-chemical characterization of metal-containing nanoparticles for antimicrobial coatings. Tartu, 2019, 138 p.
57. **Jelena Beljantseva.** Small fine-tuners of the bacterial stringent response – a glimpse into the working principles of Small Alarmone Synthetases. Tartu, 2020, 104 p.
58. **Egon Urgard.** Potential therapeutic approaches for modulation of inflammatory response pathways. Tartu, 2020, 120 p.
59. **Sofia Raquel Alves Oliveira.** HPLC analysis of bacterial alarmone nucleotide (p)ppGpp and its toxic analogue ppApp. Tartu, 2020, 122 p.
60. **Mihkel Örd.** Ordering the phosphorylation of cyclin-dependent kinase Cdk1 substrates in the cell cycle. Tartu, 2021, 228 p.
61. **Fred Elhi.** Biocompatible ionic electromechanically active polymer actuator based on biopolymers and non-toxic ionic liquids. Tartu, 2021, 140 p.

62. **Liisi Talas.** Reconstructing paleo-diversity, dynamics and response of eukaryotes to environmental change over the Late-Glacial and Holocene period in lake Lielais Svētiņū using sedaDNA. Tartu, 2021, 118 p.
63. **Livia Matt.** Novel isosorbide-based polymers. Tartu, 2021, 118 p.
64. **Koit Aasumets.** The dynamics of human mitochondrial nucleoids within the mitochondrial network. Tartu, 2021, 104 p.
65. **Faiza Summer.** Development and optimization of flow electrode capacitor technology. Tartu, 2022, 109 p.
66. **Olavi Reinsalu.** Cancer-testis antigen MAGE-A4 is incorporated into extracellular vesicles and is exposed to the surface. Tartu, 2022, 130 p.
67. **Tetiana Brodiazhenko.** RelA-SpoT Homolog enzymes as effectors of Toxin-Antitoxin systems. Tartu, 2022, 132 p.
68. **Georg-Marten Lanno.** Development of novel antibacterial drug delivery systems as wound scaffolds using electrospinning technology. Tartu, 2022, 175 p.
69. **Liubov Cherkashchenko.** New insights into alphaviral nsP2 functions. Tartu, 2023, 171 p.
70. **Kristina Kiisholts.** Peptide-based drug carriers and preclinical nanomedicine applications for endometriosis treatment. Tartu, 2023, 138 p.
71. **Kai Rausalu.** Alphaviral nsP2 protease: From requirements for functionality to inhibition. Tartu, 2023, 175 p.
72. **Laura Sandra Lello.** Unraveling the intricate nature of the alphavirus RNA replicase. Tartu, 2023, 219 p.
73. **Houman Masnavi.** Visibility Aware Navigation. Tartu, 2023, 180 p.
74. **Kadir Aktas.** Cosmic Ray Tomography based Object Reconstruction and Recognition. Tartu, 2023, 104 p.