



**REPUBLIC OF TURKEY
ADANA ALPARSLAN TÜRKEŞ SCIENCE AND
TECHNOLOGY UNIVERSITY**

**GRADUATE SCHOOL OF NATURAL AND APPLIED
SCIENCES
DEPARTMENT OF NANOTECHNOLOGY AND ENGINEERING
SCIENCES**

**CLASSIFICATION OF BRAIN MR IMAGE DATA USING DATA
MINING TECHNIQUES**

**ÜMİT KILIÇ
MASTER OF SCIENCE**



**REPUBLIC OF TURKEY
ADANA ALPARSLAN TÜRKEŞ SCIENCE AND
TECHNOLOGY UNIVERSITY**

**GRADUATE SCHOOL OF NATURAL AND APPLIED
SCIENCES
DEPARTMENT OF NANOTECHNOLOGY AND ENGINEERING
SCIENCES**

**CLASSIFICATION OF BRAIN MR IMAGE DATA USING DATA
MINING TECHNIQUES**

**ÜMİT KILIÇ
MASTER OF SCIENCE**

**SUPERVISOR
ASST. PROF. DR. MÜMİNE KAYA KELEŞ**

ADANA 2019

Classification of Brain MR Image Data Using Data Mining Techniques

Submitted by **Ümit KILIÇ** in partial fulfillment of the requirements for the degree of **Master of Science in Department of Nanotechnology and Engineering Sciences, Adana Alparslan Türkeş Science and Technology University** by,

Assoc. Prof. Dr. Gözde
BAYDEMİR PEŞİNT
Graduate School Director

Asst. Prof. Dr. Hatice İmge
OKTAY BAŞEĞMEZ
Department Chair

Asst. Prof. Dr. Mümine KAYA
KELEŞ
Supervisor

Thesis Committee

Prof. Dr. Sélma Ayşe ÖZEL
Çukurova University

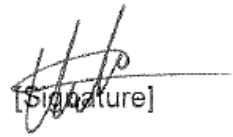
Asst. Prof. Dr. Mümine KAYA
KELEŞ
Adana Alparslan Türkeş
Science and Technology
University

Asst. Prof. Dr. Esra SARAÇ
EŞSİZ
Adana Alparslan Türkeş
Science and Technology
University

Thesis Defense Date

20/06/2019

I hereby declare that all information in this thesis has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all information that is not original to this work.



[Signature]

Ümit KILIÇ

ABSTRACT

Classification of Brain MR Image Data Using Data Mining Techniques

Ümit KILIÇ

Department of Nanotechnology and Engineering Sciences

Supervisor: Asst. Prof. Dr. Mümine KAYA KELEŞ

June 2019, 67 Pages

Alzheimer's Disease is a disease that shows its effects with the progression of the age and it causes the brain to be unable to fulfill its expected functions. The disease's effects show variety according to its phase such as to forget the name of surrounding people or cannot continue daily life without help. As of we know, there is no method has general acceptance for diagnosis and treatment.

In this thesis study, the performance of data mining methods and that of deep learning on diagnosis of Alzheimer's Disease (AD) is assessed by using patients' magnetic resonance image (MRI) data. MRI data from 144 healthy persons named control normal (CN) and 175 patients with AD are used. These two groups consist of 167 male and 152 female aged between 55 and 91. Volumetry statistics of the parts of the brain from related MRI data are obtained and performance of the traditional data mining algorithms on feature selection and classification are evaluated. Also, the performance of Artificial Bee Colony algorithm developed in this thesis and used as a feature selector is compared with the classical feature selector methods. Besides this, collected images segmented as Gray Matter (GM), White Matter (WM) and Cerebrospinal Fluid (CSF) after the normalisation process. Statistical Parametric Mapping (SPM) software's tools are utilized for the normalization and segmentation. These GM images are used in deep learning method. Results are compared.

Keywords: *alzheimer's disease, deep learning, magnetic resonance image, feature selection, data minig, artificial bee colony*

ÖZET

Beyin MR Görüntülerinin Veri Madenciliği Teknikleri Kullanılarak Sınıflandırılması

Ümit KILIÇ

Nanoteknoloji ve Mühendislik Bilimleri

Danışman: Dr. Öğr. Üyesi Mümine KAYA KELEŞ

Haziran 2019, 67 Sayfa

Alzaymır yaşın ilerlemesiyle etkilerini gösteren ve beynin kendisinden beklenen fonksiyonları yerine getirememesine neden olan bir hastalıktır. Bu hastalığın etkileri bulunduğu çevreye göre çevredeki kişilerin adını unutmak veya gündelik hayatına yardım almadan devam edememek gibi değişiklik gösterir. Genel bir kabule sahip olan tedavi yöntemi olmadığı gibi erken teşhis için de kesin bir yöntem bulunmamaktadır.

Bu tez çalışmasında, hastaların beyin manyetik rezonans görüntü (MRI) verileri kullanılarak veri madenciliği ve derin öğrenme yöntemlerinin alzaymır hastalığının teşhisi üzerine performansları değerlendirilmiştir. 144 kontrol grubu olarak adlandırdığımız sağlıklı kişinin ve 175 alzaymır hastalığına sahip kişinin MR görüntüleri kullanılmıştır. Bu iki grup yaşları 55 ile 91 arasında olan 167 erkek ve 152 kadından oluşmaktadır. İlgili MRI verilerinden beyin kısımlarının hacimsel istatistikleri elde edilmiş ve klasik veri madenciliği algoritmalarıyla hacimsel değişiklikler üzerinden nitelik seçimi ve sınıflandırma performansı değerlendirilmiştir. Ayrıca öznitelik seçim yöntemi olarak bu çalışma kapsamında geliştirilen Yapay Arı Kolonisi Algoritmasının performansı klasik öznitelik seçme yöntemleriyle karşılaştırılmıştır. Toplanan görüntüler normalizasyon işleminden sonra, Gri Madde, Beyaz Madde ve Beyin Omurilik Sıvısı olarak bölüntülenmiştir. Bu normalizasyon ve bölüntüleme işlemleri için Statistical Parametric Mapping araçlarından faydalanılmıştır. Bu Gri Madde görüntüler derin öğrenme içerisinde kullanılmıştır. Sonuçlar karşılaştırılmıştır.

Anahtar Kelimeler: *alzaymır, derin öğrenme, manyetik rezonans görüntüleme, öznitelik seçimi, veri madenciliği, yapay arı kolonisi*

ACKNOWLEDGEMENTS

First of all, I thank my supervisor, Asst. Prof. Dr. Mümine KAYA KELEŞ, for her support, precious advices and friendly behavior through this thesis. I am also grateful to Prof. Dr. Selma Ayşe ÖZEL and Asst. Prof. Dr. Esra SARAÇ EŞSİZ for accepting to read and review this thesis and for their invaluable suggestions.

I owe my thanks to my father, Aziz KILIÇ, and my mother, Hüsniye KILIÇ, for bringing me up to this day with a great devotion and love. Also, I wish to thank my brother and sisters for their supports.

I wish to express my appreciation to Department of Computer Engineerin, which I will always be proud to be a part of this department, and Department of Nanotechnology and Engineering Sciences, which I am registered as a master degree student. I would like to thanks Adana Alparslan Türkeş Science and Technology University Scientific Research Office for providing financial support under project no: 18332001

TABLE OF CONTENTS

ABSTRACT.....	v
ÖZET	vi
ACKNOWLEDGEMENTS	vii
TABLE OF CONTENTS.....	viii
LIST OF FIGURES	x
LIST OF TABLES.....	xii
NOMENCLATURE.....	xiii
1 INTRODUCTION.....	1
2 LITERATURE REVIEW	4
2.1 Traditional Data Mining Methods	4
2.2 Deep Learning Methods	8
3 MATERIALS AND METHODS.....	13
3.1 Materials.....	13
3.1.1 ADNI dataset and MRI data.....	13
3.1.2 Statistical parametric mapping software	16
3.1.3 Waikato Environment for Knowledge Analysis (Weka) software	17
3.1.4 volBrain online automated mri brain volumetry system	17
3.2 Methods	21
3.2.1 Feature Selection	21
3.2.2 Artificial bee colony (ABC) algorithm as feature selector.....	22
3.2.3 Deep Learning.....	28
3.2.4 Transfer Learning	38
4 RESULTS AND DISCUSSION	40
4.1 Evaluation Metrics	40
4.2 Result Tables	41
5 CONCLUSION.....	48
REFERENCES	50
APPENDIX - A.....	58

APPENDIX – B 62
VITA..... 67



LIST OF FIGURES

Figure 1.1 Number of people with dementia in low and middle income countries compared to high income countries (Alzheimer's Disease International, 2015).....	1
Figure 3.1 MRI Machine (National Institute of Biomedical Imaging and Bioengineering, 2019)	13
Figure 3.2 Plane representation of the body (IPF Radiology Rounds, 2019)	14
Figure 3.3 Respresentation of the planes for a brain (Walia et al., 2013)	14
Figure 3.4 View of Sagittal (a), Axial (b) and Coronal (c) plane of the brain MRI ..	14
Figure 3.5 A view of the segmentation windows in SPM 12	17
Figure 3.6 Country and usage frequency of volBrain users (volBrain, 2019)	18
Figure 3.7 Jobs processed by the system by the time (volBrain, 2019).....	18
Figure 3.8 The volBrain System Processing Scheme (Manjon & Coupe, 2016) ...	19
Figure 3.9 An example of first page of a result report in pdf format generated from volBrain system.....	20
Figure 3.10 An example of second page of a result report in pdf format generated from volBrain system.....	21
Figure 3.11 Flowchart of ABC algorithm (Akay & Karaboğa, 2012)	25
Figure 3.12 An example of bit vector for feature selection.....	26
Figure 3.13 Representation of the perceptron	28
Figure 3.14 Linear (left) and non-linear (right) models.....	29
Figure 3.15 Linear function (Ronaghan, 2018).....	29
Figure 3.16 Sigmoid function (Ronaghan, 2018)	30
Figure 3.17 Hyperbolic Tangent Function (Ronaghan, 2018)	30
Figure 3.18 ReLU function (Ronaghan, 2018).....	30
Figure 3.19 Leaky ReLU function (Ronaghan, 2018)	31
Figure 3.20 Softmax representation (Ronaghan, 2018).....	31
Figure 3.21 Visualization of Neural Network (Wikipedia, 2019)	32
Figure 3.22 Deep Neural Network (Rezzak et al., 2017)	32
Figure 3.23 Example of convolution process.....	34
Figure 3.24 Example of Max Pooling process	34
Figure 3.25 Example of Average Pooling process.....	35
Figure 3.26 Logical representation of VGG (Neurohive, 2018).....	35
Figure 3.27 VGG Architecture (Simonyan & Zisserman, 2014)	36
Figure 3.28 The architecture of ResNet (He et al., 2016)	37
Figure 3.29 Example of the GM (a), WM (b) and CSF (c) generated using SPM 12 software	37

Figure 3.30 An example of plotted result screen 38
Figure 3.31 Representation of the Transfer Learning (Anonymous) 39



LIST OF TABLES

Table 3.1 Age statistics of data	15
Table 3.2 Sex distribution of used data	15
Table 3.3 Average age of the patients.....	15
Table 3.4 Detailed information of ADNI dataset	16
Table 3.5 Main steps of ABC (Karaboga & Basturk, 2007).....	23
Table 3.6 Steps of ABC for feature selection.....	26
Table 3.7 Brief detail, Advantages and Disadvantages of Deep Learning models (Rezzak et al., 2017)	33
Table 4.1 Confusion Matrix	40
Table 4.2 Result of the raw dataset with the mentioned algorithms.....	42
Table 4.3 Results after feature selection with Info Gain, Gain Ratio and CFS	43
Table 4.4 Features selected by IG, GR and CFS methods.....	44
Table 4.5 Results after feature selection with ABC algorithm	44
Table 4.6 Features selected by ABC feature selector algorithm	45
Table 4.7 Result of OneR and Random Tree with added Cerebrum Asymmetry feature.....	45
Table 4.8 Classification time information for Naive Bayes in WEKA	46
Table 4.9 Time information for ABC, VGG16 and ResNet50	46
Table 4.10 VGG16 Results with GM of MRI Data.....	47
Table 4.11 ResNet50 Results with GM of MRI Data.....	47

NOMENCLATURE

AD	Alzheimer's Disease
MCI	Mild Cognitive Impairment
MMSE	Mini-Mental State Examination
MRI	Magnetic Resonance Imaging
CT	Computerized Tomography
SPECT	Single Proton Emission Computed Tomography
fMRI	Functional Magnetic Resonance Imaging
PET	Positron Emission Tomography
FDG-PET	Fluorodeoxyglucose Positron Emission Tomography
DM	Data Mining
TANH	Hyperbolic Tangent
ReLU	Rectified Linear Unit
RMSE	Root Mean Squared Error
SVM	Support Vector Machine
HC	Healthy Control
SPM	Statistical Parametric Mapping
GM	Grey Matter
WM	White Matter
CSF	Cerebro-Spinal Fluid
LOOV	Leave-One-Out validation
VFI	Voting Feature Intervals

ADNI	Alzheimer's Disease Neuroimaging Initiative
OASIS	Open Access Series of Imaging Studies
AC	Anterior Commissure
PC	Posterior Commissure
ROI	Region of Interest
VOI	Volumes of Interest
BSE	Brain Surface Extractor
BET	Brain Extraction Tool
VBM	Voxel Based Morphometry
TBM	Tensor Based Morphometry
APOE	Apolipoprotein E
OPLS	Orthogonal Partial Least Squares
ANN	Artificial Neural Network
InnoMed	Innovative Medicine in Europe
SEAD-J	Diagnosis of Early Alzheimer's Disease – Japan
RFE	Recursive Feature Elimination
AAL	Automated Anatomical Labeling
BA	Broadmann's Areas
LPBA	Loni Probabilistic Brain Atlas
RBF	Radial Bases Function
GA	Genetic Algorithm
MDS	Multi-Dimensional Scaling
SNP	Single Nucleotide Polymorphisms
SAE	Stacked Auto Encoder
MK-SVM	Multi-Kernel Support Vector Machine
DICOM	Digital Imaging and Communication in Medicine

NifTI	Neuroimaging Informatics Technology Initiative
PNG	Portable Network Graphics
ReLU	Rectified Linear Unit
CNN	Convolutional Neural Network
RNN	Recurrent Neural Network
dA	Deep Autoencoder
DBM	Deep Boltzmann Machine
DBN	Deep Belief Network
CPU	Central Processing Unit
GPU	Graphic Processing Unit
FC	Fully Connected
JPEG	Joint Photographic Experts Group
MP-RAGE	Magnetization-Prepared Rapid Gradient Echo
PCA	Principal Component Analysis
LASSO	Least Absolute Shrinkage and Selection Operator
ABC	Artificial Bee Colony
PSO	Particle Swarm Optimization
ACO	Ant Colony Optimization
AFSA	Artificial Fish Swarm Algorithms
FA	Firefly Algorithms
BA	Bat Algorithms
SMO	Sequential Minimal Optimisation
CAD	Computer-Aided Diagnosis
FDA	the Food and Drug Administration
NIA	the National Institute on Aging
NIBIB	the National Institute of Biomedical Imaging and Bioengineering

IG	Info Gain
GR	Gain Ratio
CFS	Correlation Based Feature Selection
PDF	The Portable Document Format
CSV	Comma-Separated Values
φ	Phi



1 INTRODUCTION

Alzheimer's Disease (AD) is a neurological illness that causes the brain to lose its functionalities such as thinking, memorizing, speaking, reading, writing. Because of this disorder, patients cannot perform their daily life responsibilities and needs. AD is the most well-known form of dementia and accounts for 50-60% all cases (Alzheimer's Disease International, 2018). According to the 2018 World Alzheimer Report from Alzheimer's Disease International, there are 50 million people have dementia in 2018 and one new case of dementia will show up every 3 seconds and expected number of people with dementia is 152 million in 2050. Estimated worldwide cost of dementia is US\$1 trillion in 2018 (Alzheimer's Disease International, 2018).

Figure 1.1 shows people with dementia by grouping income of their countries. It can be seen that increment speed of the number of the people live in countries with low and middle income is much more than that of those who live in high income countries.

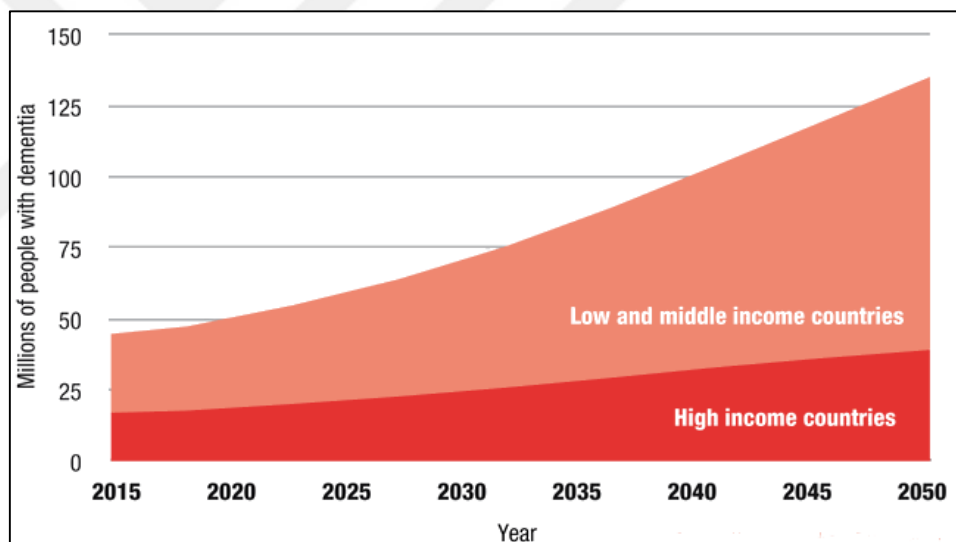


Figure 1.1 Number of people with dementia in low and middle income countries compared to high income countries (Alzheimer's Disease International, 2015)

As it is known, there is not a method for detection or treatment which has general acceptance for the disorder. But early detection is crucial for the proper treatment of AD. Traditional ways for assessment of AD such as Mini-Mental State Examination (MMSE), patient's detailed disorder history, neurobiological and physical exams need an expert and are also time consuming. In addition to those traditional methods, developing technologies help physicians by improving the medical test.

Medical tests, such as blood, urine, and genetic tests, as well as brain scans, are used in the diagnosis of dementia. Blood or urine test are performed to be able to apprehend other symptoms like vitamin and nutrient levels, infections, as well as kidney, liver, and thyroid function. Genetic tests are done to obtain dementia history of the family. Apart from those methods, brain scans can be used to detect dementia by monitoring shrinkage and enlargement of the brain parts. Those brain scans include magnetic resonance imaging (MRI), computerized tomography (CT), single proton emission computed tomography (SPECT), functional magnetic resonance imaging (fMRI) and positron emission tomography (PET) (Dementia Australia, 2019).

Patterns extraction from those data using diverse methods such as machine learning and statistical techniques has become an extensive area for diagnosis of AD. Since it is a non-invasive method, MRI data have common usage for the purpose of diagnosis. Other reasons are that it is cheaper and easier to apply than other techniques. While tracking clue of AD on the brain MRI data or its statistical data, parts of the brain which are related to planning, remembering, thinking, speaking and judgment, such as hippocampus, cerebral cortex, amygdala, ventricles in it, etc., are monitored. Because, AD causes shrinkage or enlargement over those parts, and the level of those volumetry changes depends on the progression of the disease. To find those perceptible discrepancies on the data by analyzing by human requires comprehensive knowledge and considerable experiences. Not only experiences and knowledge but also additional clinical results that should be conjoined with those experiences and knowledge are needed to accomplish a proper classification (Sarraf & Tofighi, 2016).

All those necessities need a bunch of time and the time is a phenomenon which should be shortened as much as possible for diagnosis. A prosperous computer-aided diagnosis (CAD) method which has the ability to categorize may assist physicians and may expedite the processes of the diagnosis of brain disorders. Klöppel et al., (2008) have conducted a study to see whether CAD systems can help and show better performance than physicians. In their study, between radiologists and computerized method have been compared in means of the accuracy of dementia diagnosis, directly. MRI data have been collected from different dataset and physicians from different health-center and with different experience have been employed. MRI data classified after preprocessing and SVM used as the classifier. At the end of the study, it has been found that those CAD systems have been performed superior performance than experts or the system have had an equal performance with the experts.

This thesis focuses on the performance comparison of traditional data mining (DM) methods and that of deep learning, which can be known as relatively new, on AD classification. MRI data from ADNI dataset has been used for the thesis. For this purpose, one of the most known and commonly used nature-inspired optimization algorithm Artificial Bee Colony (ABC) has been implemented as a feature selector because of that the algorithm also shows good performance on feature selection processes. Bayes Net, K-Nearest Neighborhood, Random Tree, Naive Bayes, J48, Decision Table, Sequential Minimal Optimisation (SMO), Bagging, Random Forest, OneR algorithms have been used as the fitness function in the ABC feature selector. Online automated MRI brain volumetry system volBrain (<http://volbrain.upv.es/>, Manjon & Coupe, 2016) has been used to acquire volumetric data from the MRI data. Performance of those algorithms on the acquired data has been concluded as the performance of data mining techniques. For deep learning stage, collected MRI data have been used as input of the deep learning after preprocessing and segmentation by converting PNG format

The purpose of this thesis is to discover the most AD disorder-related brain regions using a feature selection method and classify the MRI data. This kind of CAD studies are needed to help physicians about diagnosis, treatment, and to reduce wasting time. There are studies in the literature that share the similar purposes with this thesis which are such as classification or feature selection using data mining or deep.

In the next chapter, literature studies have been presented. In Chapter 3, the materials and methods used in this thesis has shown while experiments were conducted. In Chapter 4, obtained results have been discussed and conclusions about the study have been added to Chapter 5.

2 LITERATURE REVIEW

In this section of the thesis, studies deal with the diagnosis of AD in the literature are summarized. Some of the studies use traditional data mining methods while the other studies utilize from deep learning methods which are one of the most state-of-art machine learning technique. Literature review with the brief descriptions also can be seen in Appendix - A.

2.1 Traditional Data Mining Methods

Data mining is the science of obtaining utilizable information from huge datasets that has intersections with statistics, machine learning, artificial intelligence, data management and databases, and so on (Hand, Manilla, & Smyth, 2001). Data mining is also used in the medical area to detect, diagnosis and assessments of diseases. One of these areas is dementia (Maroco et al., 2011).

Dementia is a clinical syndrome that affects the brain and causes not be able to sustain daily life. AD, vascular dementia, Lewy body, and frontotemporal dementia are some types of the disease. AD is the most common form of dementia (World Health Organization, 2019). Throughout literature, there are studies to detect AD utilizing data mining methods .

Klöppel et al. (2008) have conducted a study to see how successfully Support Vector Machine (SVM) works on individual diagnosis and different datasets. Those three datasets have been from divergent centers and diverse scanners. Three groups have created with a different number of subjects as Group 1, Group 2 and Group 3. Group 1 has consisted of 20 AD patient and 20 healthy control (HC) persons. Group 2 has included 14 AD and 14 HC, while 3 AD and 57 HC have been in Group 3. Images have been segmented as grey matter (GM), white matter (WM) and cerebro-spinal fluid (CSF) using Statistical Parametric Mapping 5 (SPM) (Wellcome Trust Centre for Neuroimaging, Institute of Neurology, UCL, London UK—<http://www.fil.ion.ucl.ac.uk/spm>) software. Then, the diffeomorphic registration algorithm has been implemented to minimize structural variation among subjects' data. Finally, SVM with the leave one out validation (LOOV) has been used to AD-HC classification and 95%, 92.9%, 81.1% accuracy values have been obtained for Group 1, Group 2, Group 3, respectively.

Plant et al. (2010) have presented their study to detect AD-related regions using three classifiers including SVM, Bayesian Classifier (Bayes) and Voting Feature Intervals

(VFI). Private dataset sourced MRI data that belong to 74 persons, which are 32 AD patients, 24 mild cognitive impairment (MCI) patients, and 18 HC persons have preprocessed by SPM 2 and those preprocessed data have been used to apply feature selection, clustering, and classification operations to be able to detect the most discriminative parts of the brain. AD-HC classification has been done by using LOOV and abovementioned three classifiers. 92% accuracy value has been acquired with the Bayes classifier. According to their results, it has been found that major changes in the brain regions to distinguish AD-HC forms have been the prefrontal cortex, especially the middle and inferior frontal gyri, the adjacent subcortical basal ganglia, hippocampal region and posterior brain regions which are in the parietal lobe.

Poulin et al. (2011) have applied a study that has had a similar purpose to Plant et al. (2010). They have wanted to compare some regions of the brain such as amygdala and hippocampus and detect disease-related area. MRI data of 174 patients from Alzheimer's Disease Neuroimaging Initiative (ADNI) database and that of 90 patients from Open Access Series of Imaging Studies (OASIS) have been utilized by dividing into two sample groups as Sample 1 and Sample 2. To be able to compare the level of amygdala atrophy to that of the hippocampus, FreeSurfer (<http://surfer.nmr.mgh.harvard.edu/>) software has been used while volumetric analysis was performed. After extraction of the regions' statistical information, Variance and chi-square analyzes have been employed to correlate sample groups. As a result, it has been found that amygdala atrophy and hippocampal atrophy have paralleled each other at the different stage of dementia severity, strongly.

Not only MRI but also other biomarkers such as fluorodeoxyglucose positron emission tomography (FDG-PET) and CSF are can be used for the studies. Zhang et al. (2011) have proposed to combine three modalities of biomarkers which were MRI, FDG-PET, and CSF to detect the difference between AD and normal groups. 51 AD, 99 MCI, and 52 HC participants' MRI data have been obtained from ADNI dataset for the study. After anterior commissure (AC), posterior commissure (PC), skull stripping using brain surface extractor (BSE) and brain extraction tool (BET), MR images have been segmented into three tissue; GM, WM and CSF and 93 regions for MRI, 93 regions for FDG-PET and 3 original value from CSF have been used directly. Volumetric features have been extracted from MRI and FDG-PET images' regions of interest (ROI) by atlas warping algorithm. Then SVM has been used with 10-fold cross-validation. 86.2% has been obtained as an accuracy value using only MRI. When MRI combined with FDG-PET, the value has raised to 90.6%. Lastly, Three biomarkers have been used and the accuracy value of the classification has been received as 93.2%. Apart from these results, features have been ranked considering

their averaged SVM weights. Amygdala right, hippocampal formation left and hippocampal formation right have been the best three of the top 11 regions selected for MCI classification.

Likewise, Hinrichs et al. (2011) have utilized multi markers to analyze MCI progression by using ADNI dataset. MRI and FDG-PET imaging modalities, CSF, Apolipoprotein E (APOE), Cognitive Scores non-imaging modalities have been used on 233 subjects including 48 AD, 66 HC, and 119 MCI. For features, voxel-based morphometry (VBM) and tensor-based morphometry (TBM) have been applied. Then Multi-Kernel Learning (MKL) and SVM have been compared for classification. MKL has outperformed SVM about 3%-4%. 87.6% accuracy value has been acquired with MRI and PET modalities, while 92.4 % accuracy value has been obtained with all MRI, PET, CSF, APOE, Cognitive Scores measurements.

Westman et al. (2012) have been utilized ADNI dataset to combine MRI and CSF for the prediction of MCI conversion. The study has included 96 AD, 162 MCI, and 111 HC subjects who had have successful MRI and CSF measures. Orthogonal partial least squares (OPLS) has been used to combine the measures for individual classification. For MCI-HC classification, 77.6% for CSF and MRI, 71.8 % for MRI only, 70.3 % for CSF only values have been obtained. For AD-HC classification, the results have been 81.6 % for CSF only, 87 % for MRI only and 91.8 % for MRI and CSF combination.

Aguilar et al. (2013) have used MRI data from AddNeuroMed project which is part of InnoMed (Innovative Medicine in Europe) which had been designed to make more efficient drug discovery which are compatible with the ADNI dataset. 345 participants with 116 AD, 119 MCI, and 110 HC have been included for the study. FreeSurfer software has been used to process and analyze the images including regional volume segmentation and cortical thickness parcellation. Four classifiers which are OPLS, Decision Tree, SVM, and Artificial Neural Networks (ANN) have been used to classify results. The values of 81.9% for Decision Tree, 84.9% for ANN, 83.6% for SVM and 84.5% for OPLS have been achieved for AD-HC classification. When looking at the classification results produced by MRI data, it can be seen that there is no big difference between classification accuracies. Hippocampus, amygdala and entorhinal are three of the most important measures according to SVM and OPLS.

Ota et al. (2014) have compared three brain atlases to see their effect on the performance of the AD-MCI classification. The study has included a voxel-based feature extraction by atlas-based parcellation. 37 patients, who had developed AD within 3 years and 38 patients who had not, have been used from Diagnosis of Early Alzheimer's Disease - Japan (SEAD-J) dataset. For feature selection, VBM and SVM with recursive feature

elimination (RFE) have been utilized. The compared brain atlases have been the Automated Anatomical Labeling (AAL) Atlas, Brodmann's Areas (BA) and the LONI Probabilistic Brain Atlas (LPBA40). SVM with radial bases function (RBF) has been used for classification and the best performance has been achieved by LPBA40 with 77.9% accuracy.

Zhou et al. (2014) have carried out the study to classify AD status using 309 participants' MRI and MMSE data from Wien Center for Alzheimer's Disease and Memory Disorders private dataset. The participants have consisted of 59 AD, 67 amnesic MCI, 56 nonamnesic MCI, and 127 HC patients. FreeSurfer software has been used to calculate volumetric variables and Student's t-test has been used for feature rank. SVM has been implemented to classify results. For AD-HC classification, values of 78.2% with MRI only and 92.4% with combined MRI and MMSE have been acquired. Right hippocampus, left hippocampus and left amygdala have been three of the five most important regions according to the feature ranking.

Another study that uses measurements of the brain is conducted by Sorensen et al. (2017). MRI biomarkers cortical thickness measurements, volumetric measurements, hippocampal shape, and hippocampal texture have been combined in their study to the purpose of the diagnosis of HC, MCI, and AD. ADNI, the imaging arm of the Australian Imaging Biomarkers and Lifestyle flagship study of ageing (AIBL) and the data from CADDementia Challenge have been employed. For extraction of the abovementioned measurements, FreeSurfer software has been used and the combination has been realized by sending all biomarkers in a linear discriminant analysis (LDS) as features. Feature selection has been applied by using sequential forward feature selection (SFS). During the selection process, the top three most frequently selected features have been hippocampal volume (66.7%), ventricular volume (53.3%) and hippocampal texture (50%). Furthermore, three of the earliest selected features and the most frequently selected subset have been the same which is hippocampal texture, hippocampal volume, and ventricular volume. In this study, a multi-class (3 class) classification has been performed and results have been obtained via LDS as 62.7% for ADNI and AIBL, 63.0% for the CADDementia Challenge.

Behesti & Demirel (2016) have conducted another study that uses the t-test by aiming to classify AD data. ADNI dataset with 38 AD and 68 HC patients have been employed in the study. VBM has been used and with the making use of the VBM results, VOIs have been produced. The created features have been ranked with t-test scores. Then, to select the optimal number of discriminative features, Fisher Criterion between AD and HC have been calculated. SVM has been used with 10- fold cross-validation and 96.3 % accuracy value has been obtained using MRI only. In their another study (Behesti et al.,

2017), the method has been changed slightly. ADNI has been used as a source of data with 160 AD, 162 HC, 65 stable MCI, and 71 progressive MCI patients. Again VBM has been used to detect important volume reduction from gray matter. Found regions have been segmented as VOIs and values of those VOIs have been used as a feature vector. All found feature vectors have been ranked with the t-test score. Then, the binary genetic algorithm (GA), whose part of the objective function is the Fisher criterion, has been employed to identify the ideal feature subset. SVM with 10-fold cross validation has been used for classification. 93% for AD-HC classification and 75% for stable MCI - progressive MCI classification have been obtained.

Long et al. (2017) have collected MRI data of 135 HC, 132 stable MCI who had not converted to AD within 36 months, 95 progressive MCI who had converted to AD after 36 months, and 65 AD subjects from ADNI. Data with FreeSurfer processing have been chosen to bypass software-related differences and problems about the standard of quality. They have proposed a deformation-based machine learning method and the metric multi-dimensional scaling (MDS) algorithm have been employed for embedding. SVM with a linear kernel was implemented on the embedded space to classify data. As a result, 96.5% has been calculated as the accuracy value for AD-HC classification.

The study conducted by Peng et al. (2019) is another example of the multi-modality for AD classification. MRI and PET data have been obtained as imaging modalities. Single nucleotide polymorphisms (SNPs) which are the most common type of the genetic variation and have attracted attention for AD association have been collected as genetic data. Those ADNI dataset sourced data have been divided into two groups as Dataset I and Dataset II to evaluate the method. Dataset I has consisted of 189 subjects with 49 AD, 93 MCI, and 47 HC while Dataset II has comprised 360 subjects including 85 patients with AD, 185 with MCI and 90 HC. To not lose sight of feature-wise importance, distinct kernels have been used for each feature. Structured sparsity regularizer has been employed for being able to select not only discriminative but also complementary features at the feature selection stage. For Dataset II, results have been found as 96.1% by using MRI, PET and SNPs combination, 88.4% using only MRI, 86.3% using only PET and 92.3% using MRI and PET combination. Highest accuracy for Dataset I has been acquired as 94.5% with the combination of MRI, PET, and SNPs.

2.2 Deep Learning Methods

Deep learning is a sub-technique of the machine learning systems employed for various operations such as object detection in an image, speech recognition and transcribes it into text, product identification with users' interests so on (LeCun, Bengio, & Hinton, Deep

Learning, 2015). Deep learning uses representation learning methods that enable a system to use raw data and automatically find out representations needed for clustering, classification or detection. Deep learning has had a good performance at recognizing complex structures in high-dimensional data. At the problems artificial intelligence communities had been dealt with have had advances thanks to deep learning (LeCun et al., 2015).

This state-of-the-art method has gained attention by researchers from many fields including medical. It has been thought that deep learning systems can be used as assistance for the detection and diagnosis of illness (Rezzak et al., 2017). Researchers who have the same opinion have conducted studies about the detection of dementia.

Suk & Shen (2013) have applied the study using MRI, PET, and CSF from 51 AD, 99 MCI, and 52 HC along with the MMSE and AD Assessment Scale-Cognitive subscale from ADNI dataset. MRI images have been segmented into GM, WM, and CSF and have been parceled into 93 ROIs. The PET images have been normalized spatially by coregistering them to respective MRI data. Nonetheless, 3 CSF biomarkers have been selected as features along with GM tissue volume from MRI and the mean intensity from PET for the related ROIs. Then, latent feature representation has been done with stacked auto-encoder (SAE), which is a special type of neural networks consists of three layers as input, hidden and output, from low-level features in MRI, PET and CSF. Feature selection has been done by multi-task learning and the best accuracy has been produced by Multi-Kernel SVM (MK-SVM) as 95.9% for AD-HC classification.

Sarraf & Tofighi (2016) have performed the first study that has used fMRI in deep learning to classify patients as AD or HC. They also have utilized MRI data and both data have been collected from ADNI dataset. Total 144 subjects with 52 AD and 92 HC patients have been included in the first dataset that uses fMRI while 302 subjects that consist of 211 AD and 91 HC have been contained to the second dataset that uses MRI. fMRI data that were in Digital Imaging and Communication in Medicine (DICOM) format have been converted to Neuroimaging Informatics Technology Initiative (NIFTI) format while MRI data have already been collected in nifti format. Then, preprocessing operations have been applied to the data. Furthermore, the imaging data have been converted to a Portable Network Graphics (PNG) format to have been used as input for the Convolutional Neural Network (CNN) classifier after the data had been split as 75%-25% for training and testing. The classification stage, that has used LeNet (LeCun et al., 1998) and GoogleNet (Szegedy et al., 2015) architectures, has produced 99% and 98.8% values as average accuracy.

Farooq et al. (2017) have implemented a CNN based pipeline with a 4-way classifier to classify AD, MCI, Late MCI and HC using ADNI dataset. The data has contained 33 AD, 22 late MCI, 49 MCI, and 45 HC patients. After data had gathered, GM segmentation has been carried out using SPM 8 and slices have been converted to Joint Photographic Experts Group (JPEG) format. To augment the data, MRI data have been augmented by flipping the images horizontally. 25% of data have been used for testing and remaining data have been employed for the train. %10 of the train data has been set as a validation set. GoogleNet and ResNet (He et al., 2015) have been used for classification. 98.8% with GoogleNet, 98.01% with ResNet-18 and 98.14% for ResNet-152 accuracy values have been acquired.

Gunawardena et al. (2017) have arranged two experiments that use SVM with RBF kernel and CNN separately to see performance of them on classification using ADNI dataset. SVM has been used for two-classes classification as AD-HC while CNN has been used for AD-MCI and HC. Noise reduction, image normalization, image sharpening, edge detection, image segmentation steps have been applied as image processing steps. The architecture of CNN has been convolution layer + rectified Linear Unit layer (ReLU) + convolution layer + ReLU + pooling layer + Fully Connected layer (FC) + output (3 classes). Also, two sub-experiments have been done in the CNN method. For the first sub-experiment, full image without detecting any edge, full image after detecting the edges, extended ROI with and without edge detection, limited ROI without detecting edges, a limited ROI with detecting edges methods have been used, respectively. After the best method had been selected from the first sub-experiment, the method has been used in the second experiment. The dataset has been divided into two sets and used in the second sub-experiment to help to prove that CNN is not dataset biased. SVM has produced 84.4%, 95.3% and 71.4% as classification accuracy, sensitivity, and specificity, respectively. For the first sub-experiment in the second experiment, the third method which was the extended ROI with detecting edges has produced the best results as 96% and 98% for sensitivity and specificity, respectively. Also second sub-experiment has produced that the CNN had not database biased by having produced 96% sensitivity, 98% specificity for database 1 and 95% sensitivity, 98% specificity for database 2.

Luo et al. (2017) have employed ADNI dataset with 47 AD and 34 HC subjects' MRI data in their study. The architecture of CNN used in the study has consisted of three sequential groups of processing layers, two FC layers, and a classification layer. Every group in the CNN has contained convolutional, pooling and normalization layers. The data have had magnetization-prepared rapid gradient echo (MP-RAGE) which includes preprocessing steps such as Gradwarp, B1 non-uniformity, N3, and scale. Data augmentation has been done by random zooming in and out and cropping. Randomly

selected 66% data have been employed for training and the rest of the data have been utilized for testing. The method has obtained 0.69 for sensitivity and 0.98 for specificity for AD and HC.

Cheng & Liu (2017) have used MRI and PET image data from ADNI dataset with 93 AD and 100 HC for the experiment. The deep 3D CNN has been created for transforming the brain information into compact high-level features. Then, a 2D CNN has been cascaded to combine the features in order to use for classification. The proposed CNN model has consisted of four convolutional layers and three pooling layers. According to their experiment results, 85.4%, 83.9%, and 90.0% values have been obtained with MRI, while PET data has produced 87.1%, 87.1%, and 84% values for accuracy, sensitivity, and specificity, respectively. In means of the combination of two modalities, all of the three rates have increased to 89.64%, 87.1%, and 92% values.

Unlike most studies, Islam & Zhang (2018) have used 416 subjects' MRI data aged 18-96 from OASIS dataset and four different stages of AD such as non-demented, very mild, mild and moderate has been tried to classify. Data have been augmented by cropping with pre-defined dimension similar to the dimension of proposed CNN classifier. For training and testing, 80% and 20% of the data have been employed, respectively. 10% of the training data have been utilized as a validation set. Small CNNs have been created and new architectures have been implemented from the combination of those pre-created small CNNs. Each model has included several convolutional layers, pooling layers, dense blocks, and transitions layers. The results of the proposed ensemble model have had 0.97, 1.00, 0.67, 0.50, 0.94 values as precision values for non-demented, very mild, mild, moderate and average/total, respectively.

Lin et al. (2018) have proposed a framework that uses CNN to classify the MRI data that collected from ADNI dataset including 118 AD, 229 HC, and 401 MCI subjects. The method has started with the two paths, the left path has had feature extraction with the FreeSurfer software and age correction has been done. The right path had included image preprocessing steps, such as skull-stripping, brain alignment to MNI151 template which is a brain template, and 2.5D patches extraction has been done after age correction. Then feature extraction has been applied to the data with CNN. The overall architecture of CNN used in their work has contained convolutional layer, max pooling, convolutional layer, average pooling, convolutional layer, average pooling, by turns. Principal component analysis (PCA) and least absolute shrinkage and selection operator (LASSO) have been used to reduce the number of features and to get rid of redundant features at the end of both paths. After that, the paths have been merged into the classification step. For the

classification step, the extreme learning machine, which is a feed-forward neural network that includes a single hidden layer, has been employed. Finally, 79.9% for AD-MCI conversion and 88.79% for AD-HC classification has been acquired as the best performance in the classification accuracy.



3 MATERIALS AND METHODS

In this section, the used dataset and the used tools were presented under Materials title. After the title, the used methods including feature selection, Artificial Bee Colony algorithm, and deep learning were explained in the Methods section.

3.1 Materials

3.1.1 ADNI dataset and MRI data

MRI is a non-invasive imaging technology. It is employed for disease detection, diagnosis, and treatment monitoring. The method utilizes from magnets, radio waves and computer to make three-dimensional anatomical images. The patients are placed inside a magnet as shown in Figure 3.1 which is large enough for the human body and body must be remained motionless throughout the imaging process (National Institute of Biomedical Imaging and Bioengineering, 2019).



Figure 3.1 MRI Machine (National Institute of Biomedical Imaging and Bioengineering, 2019)

The method creates a three-dimensional image of the body parts by capturing the patient's body from the coronal plane, sagittal plane and axial plane (IPF Radiology Rounds, 2019). Representation of the body plane is as shown in Figure 3.2.

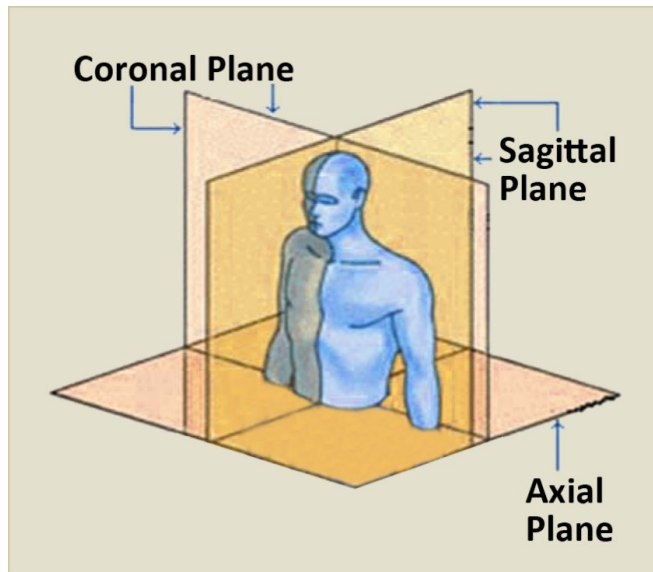


Figure 3.2 Plane representation of the body (IPF Radiology Rounds, 2019)
 Same planes for the brain is as shown in Figure 3.3.

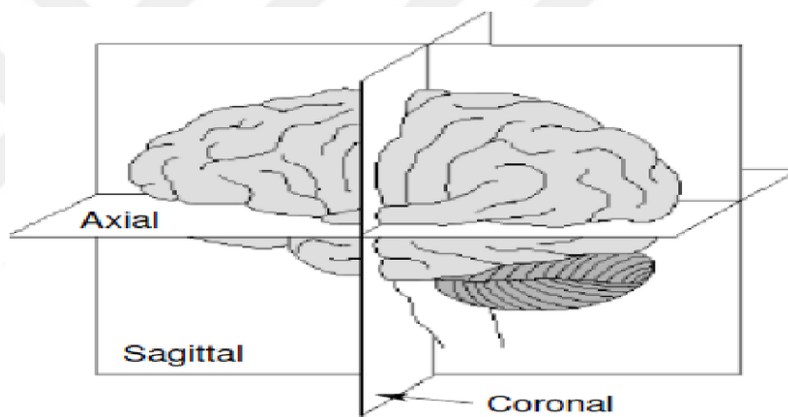


Figure 3.3 Representation of the planes for a brain (Walia et al., 2013)
 Each of the planes of the brain MR image obtained from ADNI is as shown in Figure 3.4.

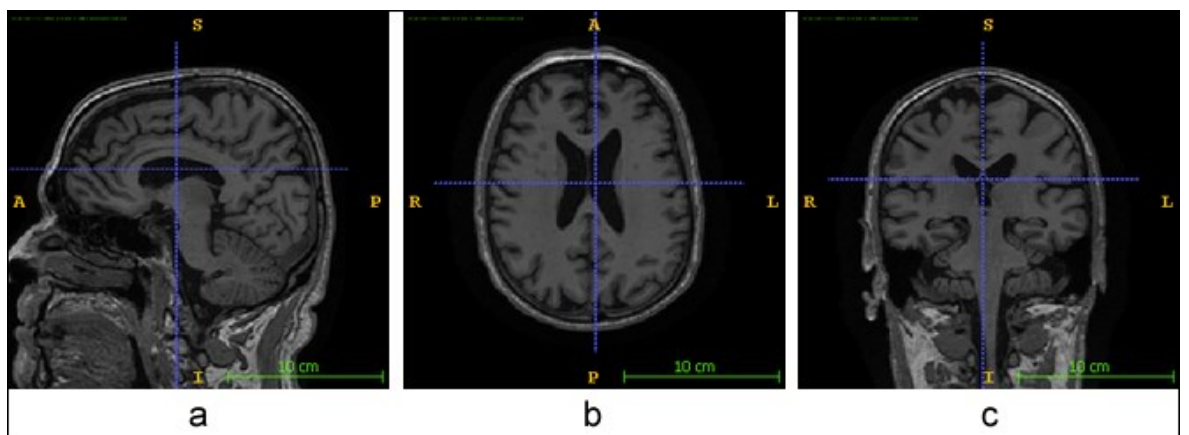


Figure 3.4 View of Sagittal (a), Axial (b) and Coronal (c) plane of the brain MRI

MRI data used in this thesis are obtained from an ADNI (<http://adni.loni.usc.edu/>) collection called ADNI1:Screening 1.5T. The dataset that launched in 2003 by the Food and Drug Administration (FDA), the National Institute on Aging (NIA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), non-profit organizations and private pharmaceutical companies and it has been running since 2004. ADNI, whose purpose is to improve diagnosis of AD and MCI using MRI, PET, other biological, clinical and neuropsychological assessments, is a multisite study. Details of the system that enrolls patients 55-91 ages can be seen in Table 3.4 (Wikipedia, 2019).

The data used in the thesis consist of 167 female and 152 male. The participants aged between 55-91.

Table 3.1 Age statistics of data

Statistic	Value
Minimum	55
Maximum	91
Mean	75.583
StdDev	6.603

Distribution of patients' sex grouped by their health status is as shown in Table 3.2.

Table 3.2 Sex distribution of used data

Sex	AD	HC
Female	92	75
Male	83	69

The average age of patients grouped by their health status and sex is as shown in Table 3.3.

Table 3.3 Average age of the patients

Sex	Average age for AD	Average age for HC
Female	74.47	75.81
Male	75.72	76.44

Table 3.4 Detailed information of ADNI dataset

Study characteristics	ADNI-1	ADNI-GO	ADNI-2	ADNI-3
Primary Goal	Developed biomarkers as outcome measures for clinical trials	Examine biomarkers in earlier stages of disease	Develop biomarkers as predictors of cognitive decline, and as outcome measures	Study the use of tau PET and functional imaging techniques and clinical trials
Funding	\$40 million federal (NIA), \$27 million industry and foundation	\$24 million American Recovery Act funds	\$40 million federal (NIA), \$27 million industry and foundation	\$40 million federal (NIA), \$20 million industry and foundation
Duration/start date	5 years / October 2004	2 years / September 2009	5 years / September 2011	5 years / September 2016
Cohort	200 elderly controls 400 MCI 200 AD	Existing ADNI-1 + 200 early MCI	Existing ADNI-GO+ 150 elderly controls 100 early MCI 150 late MCI 150 late MCI 150 AD	Existing ADNI-2 + 133 elderly controls

3.1.2 Statistical parametric mapping software

Preprocessing and segmentation for obtained MRI data have been done using SPM 12 (<http://www.fil.ion.ucl.ac.uk/spm>). SPM that created for examining fMRI or PET data is a statistical technique. It is written by the Wellcome Department of Imaging Neuroscience at University College London using MATLAB (Wikipedia, 2019). A view of the segmentation window in SPM 12 can be seen in Figure 3.5.

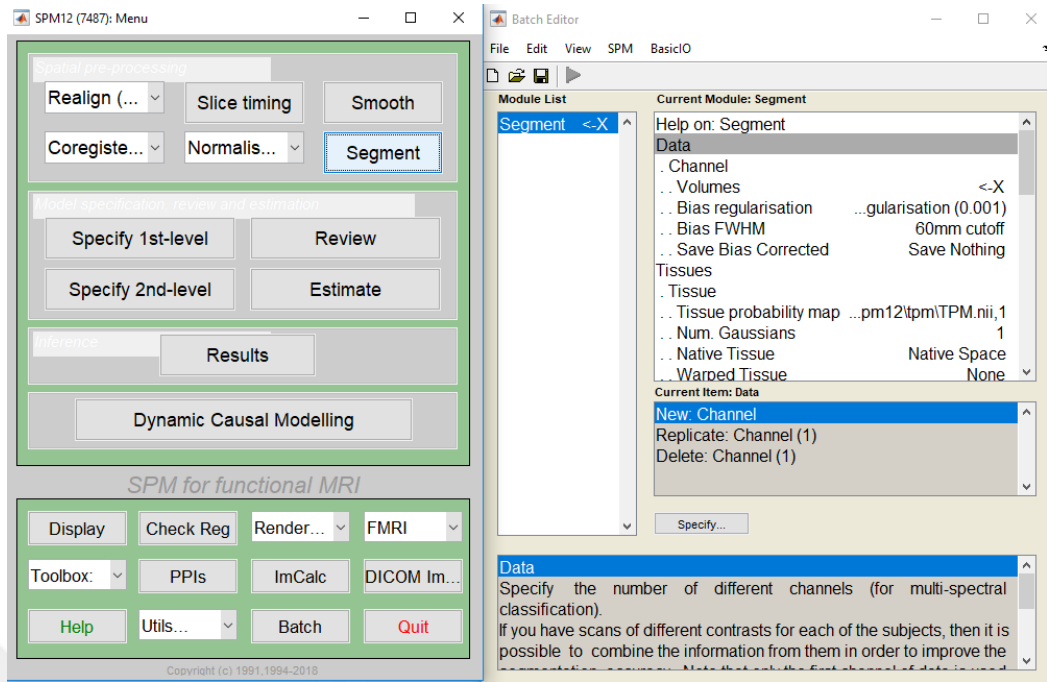


Figure 3.5 A view of the segmentation windows in SPM 12

3.1.3 Waikato Environment for Knowledge Analysis (Weka) software

Weka is a machine learning suite software written using Java programming language by the University of Waikato. It contains visualization tools and algorithms for data analysis and predictive modeling with the graphical user interface (Wikipedia, 2019). In this thesis, Weka is used to get feature selection and classification results.

3.1.4 volBrain online automated mri brain volumetry system

The volBrain system developed by Jose V. Manjon and Pierrick Coupe aims to help researchers to get the brain's volumetric information with their MRI data, automatically. Figure 3.6 shows researchers' countries and how often using the system according to the country from all around the world (volBrain, 2019).

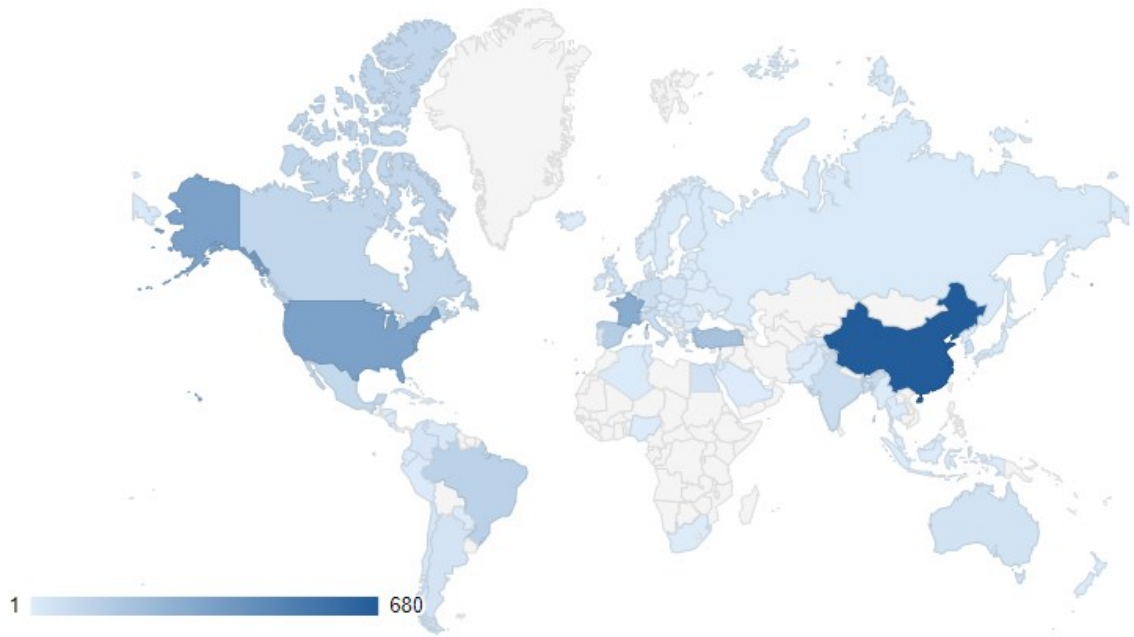


Figure 3.6 Country and usage frequency of volBrain users (volBrain, 2019)

Their number of users has been increasing day-to-day as well as the daily job number processed in the system as it can be seen in Figure 3.7.

1287 institutions including 66 of that from Turkey use volBrain from all over the world (volBrain, 2019).

According to their website (<http://volbrain.upv.es/about.php>), 30 publications have already used the system for their study.

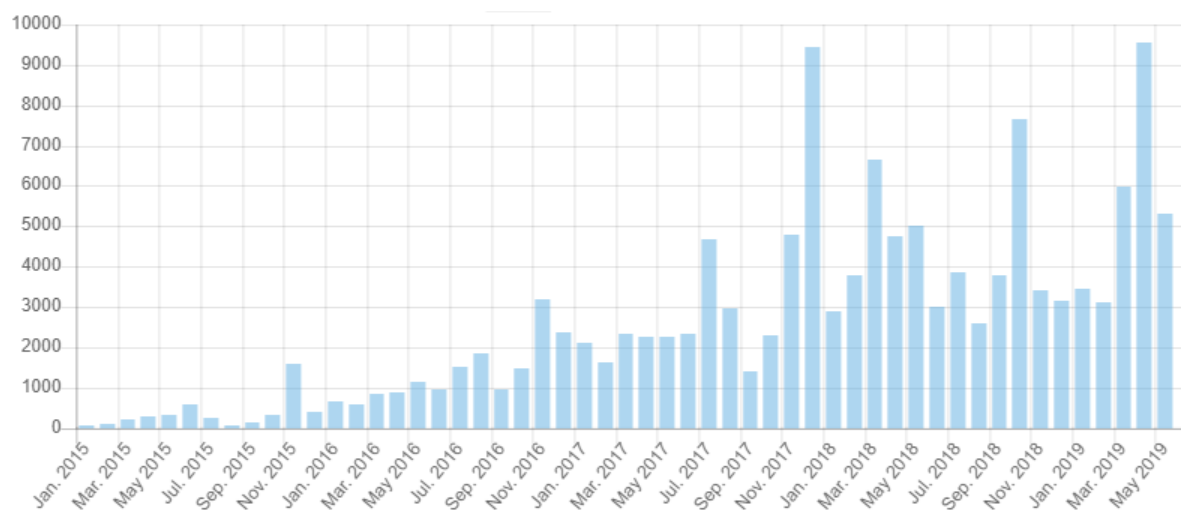


Figure 3.7 Jobs processed by the system by the time (volBrain, 2019)

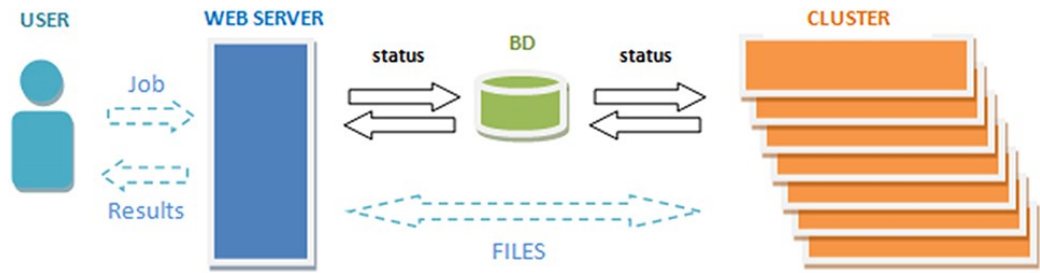


Figure 3.8 The volBrain System Processing Scheme (Manjon & Coupe, 2016)

As can be seen from Figure 3.8, the volBrain system needs the raw data that are not exposed to any preprocessing from the user. The system takes the data in nifti format and process it to generate result report. An example of a result report in the portable document format (PDF) is represented in Figure 3.9 and 3.10. The result reports also can be downloaded from the system in comma-separated values (CSV) format.

volBrain Volumetry Report version 1.0 release 04-03-2015

Patient ID	Sex	Age	Report Date
job133540	Female	71	10-Apr-2019

Tissue type	Volume (cm ³ %)		Image information
White Matter (WM)	447.66 (32.45%)	[26.09, 40.51]	Orientation neurological
Grey Matter (GM)	575.42 (41.72%)	[42.12, 53.88]	Scale factor 0.76
Cerebro Spinal Fluid (CSF)	356.28 (25.83%)	[13.99, 23.40]	SNR 29.54
Brain (WM + GM)	1023.09 (74.17%)	[76.60, 86.01]	
Intracranial Cavity (IC)	1379.36 (100.00%)		

Structure				
Cerebrum	Total (cm ³ %)	Right (cm ³ %)	Left (cm ³ %)	Asym. (%)
	889.92 (64.52%)	457.57 (33.17%)	432.35 (31.34%)	5.6667
	[66.03, 74.85]	[33.09, 37.56]	[32.88, 37.34]	[-1.28, 2.49]
	GM WM	GM WM	GM WM	
	484.55 405.38	250.08 207.49	234.47 197.88	
	(35.13%) (29.39%)	(18.13%) (15.04%)	(17.00%) (14.35%)	
	[35.32, 45.36]	[23.83, 36.37]	[17.63, 22.67]	[11.84, 18.08]

Cerebellum	Total (cm ³ %)	Right (cm ³ %)	Left (cm ³ %)	Asym. (%)
	111.64 (8.09%)	55.42 (4.02%)	56.22 (4.08%)	-1.4371
	[7.81, 10.55]	[3.89, 5.27]	[3.90, 5.29]	[-4.59, 3.80]
	GM WM	GM WM	GM WM	
	86.06 25.57	42.57 12.85	43.50 12.72	
	(6.24%) (1.85%)	(3.09%) (0.93%)	(3.15%) (0.92%)	
	[5.88, 8.64]	[2.88, 4.29]	[2.98, 4.37]	[0.47, 1.37]

Brainstem	Total (cm ³ %)
	21.59 (1.57%) [1.41, 1.97]

Structure	Total (cm ³ %)	Right (cm ³ %)	Left (cm ³ %)	Asymmetry (%)
Lateral ventricles	53.62 (3.89%)	23.00 (1.67%)	30.62 (2.22%)	-28.4397
	[0.61, 2.82]	[0.25, 1.39]	[0.31, 1.49]	[-66.0018, 48.02]
Caudate	6.73 (0.49%)	3.51 (0.25%)	3.22 (0.23%)	8.5261
	[0.38, 0.59]	[0.19, 0.30]	[0.19, 0.30]	[-7.5516, 10.90]
Putamen	6.79 (0.49%)	3.51 (0.25%)	3.28 (0.24%)	6.6779
	[0.45, 0.66]	[0.22, 0.33]	[0.22, 0.33]	[-6.7962, 5.03]
Thalamus	8.71 (0.63%)	4.32 (0.31%)	4.38 (0.32%)	-1.3849
	[0.62, 0.83]	[0.31, 0.42]	[0.31, 0.42]	[-6.9684, 5.07]
Globus Pallidus	2.20 (0.16%)	1.16 (0.08%)	1.03 (0.07%)	11.8791
	[0.13, 0.21]	[0.06, 0.10]	[0.06, 0.11]	[-15.6647, 13.44]
Hippocampus	6.22 (0.45%)	3.22 (0.23%)	3.00 (0.22%)	6.8204
	[0.48, 0.65]	[0.24, 0.33]	[0.23, 0.32]	[-6.3681, 13.15]
Amygdala	1.44 (0.10%)	0.75 (0.05%)	0.68 (0.05%)	9.5541
	[0.09, 0.14]	[0.05, 0.07]	[0.04, 0.07]	[-12.8508, 18.80]
Accumbens	0.45 (0.03%)	0.21 (0.02%)	0.24 (0.02%)	-14.9153
	[0.02, 0.06]	[0.01, 0.03]	[0.01, 0.03]	[-36.4826, 14.33]

Figure 3.9 An example of first page of a result report in pdf format generated from volBrain system

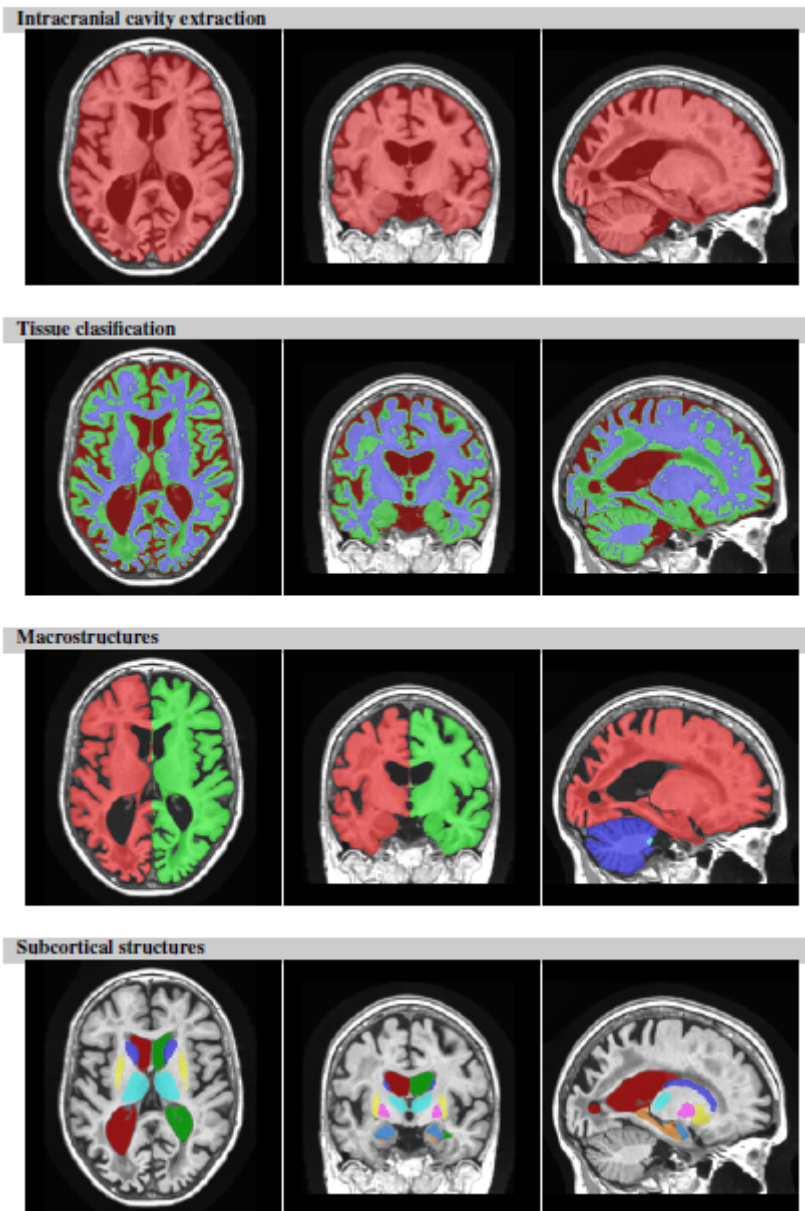


Figure 3.10 An example of second page of a result report in pdf format generated from volBrain system

3.2 Methods

3.2.1 Feature Selection

Feature selection is the process of the reduction of dimensionality of the data. It selects a subset of the features by assessing them according to certain criteria. Reduction of the number of features is carried out by removing redundant, irrelevant or noisy data. As a result of that processes, speeding up the data mining algorithm, improving performance such as accuracy and result comprehensibility are expected effects. The feature selection

methods can be broadly divided into three categories as filter, wrapper, and embedded models. Filter methods rely on general characteristics of the data for the selection of the feature subset process. Besides, evaluation of the selected feature subsets is done independently of a learning algorithm. Wrapper methods need a pre-defined learning algorithm as an evaluation criterion to evaluate the selected feature subsets. Embedded models acquire the feature relevance from the objective of the learning model, analytically, after incorporation of variable selection as a part of the training stage (Janabi & Kadhim, 2018) (Zhao et al., 2010).

Feature selection is an active research topic and has been applied to many fields such as genomic analysis (Inza et al., 2004), image retrieval (Gonzalez & Woods, 1993) (Swets & Weng, 1995), text mining (Forman, 2003) (Liu et al., 2003), intrusion detection (Lee et al., 2000), image processing/computer vision (Mustra et al., 2012), bioinformatics (Dessi, Pascariello, & Pes, 2003) (busamra, 2013), fault diagnosis (Liu et al., 2014).

Some swarm intelligence algorithms have been also implemented as a feature selector. A comprehensive and informative comparison study for swarm algorithms including Particle Swarm Optimization (PSO), Ant Colony Optimization (ACO), Artificial Fish Swarm Algorithms (AFSA), ABC, Firefly Algorithms (FA) and Bat Algorithms (BA) for feature selection is carried out by Basir & Ahmad in 2014.

In the thesis, ABC algorithm has been used as feature selector and Information Gain (IG), Gain Ratio (GR) and Correlation Based Feature Selection (CFS) methods have been also used for feature selection. CFS which is a feature subset selection algorithm calculates the correlations of all features and classes. The best feature subset found by employing correlation based heuristic evaluation method is selected. IG that selects the feature by creating a decision tree was proposed by J. R. Quinlan. Likewise, GR uses tree structure and it is an extension to IG. GR cope with the bias of IG by applying normalization to IG (Janabi & Kadhim, 2018).

ABC algorithm is been coded using JAVA programming language and NetBeans IDE (<https://netbeans.org/>). Results from IG, GR and CFS are obtained from Weka software. Performance of these methods are compared at the results and discussion chapter.

3.2.2 Artificial bee colony (ABC) algorithm as feature selector

ABC is a nature-inspired swarm intelligence based algorithm proposed by Karaboga. The algorithm simulates the foraging behavior of honey bees by dividing the bees into three

categories. Categories of the bees are employed bees, onlooker bees, and scout bees. In the algorithm, bees foods are considered an optimal solution and the bees are looking for it. In ABC, half of the population consists of employed bee and another half is onlooker bees. The employed bees whose food source is exhausted are turned into scout bees. So it can be understood that each employed bee assigned for one food source. An onlooker bee makes the decision about choosing a food source by waiting in the dance area. The employed bee goes to the food sources which is visited by itself previously. A random search is performed by the scout bee (Karaboga & Basturk, 2007). The algorithm's main steps are given in Table 3.5.

Table 3.5 Main steps of ABC (Karaboga & Basturk, 2007)

Initialize
REPEAT
(a) The employed bees are placed on the food sources in the memory
(b) The onlooker bees are placed on the food source in the memory
(c) Send the scouts to the search area for discovering new food sources.
UNTIL (requirements are met)

Explanation of the ABC steps can be done such that (Akay & Karaboğa, 2012);

In the beginning, the algorithm produces food sources randomly. Those food sources are potential solutions for the problem. This process is done by using Equation 1.

$$X_{i,j} = X_j^{min} + rand(0,1)(X_j^{max} - X_j^{min}) \quad (1)$$

where $i=1,2,3...SN$ and $j=1,2,3...D$ and D stands for the number of optimization parameter while SN stands for the number of food sources. Then created probable solutions, called food sources, are employed for search processes.

As abovementioned, employed bee numbers equal to the number of food sources, so employed bees are sent to the food sources by assigning one to one. Employed bees explore the food sources and find their neighborhoods by applying the modification to it, then assess their quality. Neighborhood exploration is done by Equation 2.

$$V_i = X_{ij} + \varphi_{ij}(X_{ij} - X_{kj}) \quad (2)$$

In this equation, j is a random integer between $[1,D]$ range. k is a element of $\{1,2,3,\dots,SM\}$ and randomly chosen. φ_{ij} is a real random number that has uniform distribution and in the range $[-1,1]$.

After exploration, the fitness value for the minimization problem can be assigned to the V_i by Equation 3.

$$fitness_i = \begin{cases} \frac{1}{1 + f_i}, & f_i \geq 0 \\ 1 + abs(f_i), & f_i < 0 \end{cases} \quad (3)$$

In the equation, f_i is cost value. If the related problem is a maximization problem, fitness function can be considered as the cost function directly. A greedy search is applied between x_i and v_i and the one has the better fitness value is selected. The new one is memorized and the old one is forgotten.

After all, the information including fitness value of food sources and the position of the food sources are shared with the onlooker bees. Onlooker bees choose a food source by assessing fitness related probability as shown in Equation 4. Positive feedback feature of the algorithm can be seen in this stage.

$$p_i = \frac{fitness_i}{\sum_{i=1}^{SN} fitness_i} \quad (4)$$

If the probability value, produced by Equation 4, is greater than the random number generated by the algorithm for each source, a modification is done by using Equation 2 by onlooker bee. Then another greedy selection is done between the old one and the modified one. If the solution is improved, the related counter is incremented by 1, otherwise, the counter is set to 0.

After all employed bees and onlooker bees finished their tasks, exhausted sources are checked whether there is any. If the counter is greater than the limit parameter of ABC, the food source is considered exhausted. Subsequently, the newly created food source by scout bees is replaced. A pre-defined maximum cycle number, error tolerance or fitness value can be used as termination criteria. Flowchart of ABC algorithm can be seen in Figure 3.11.

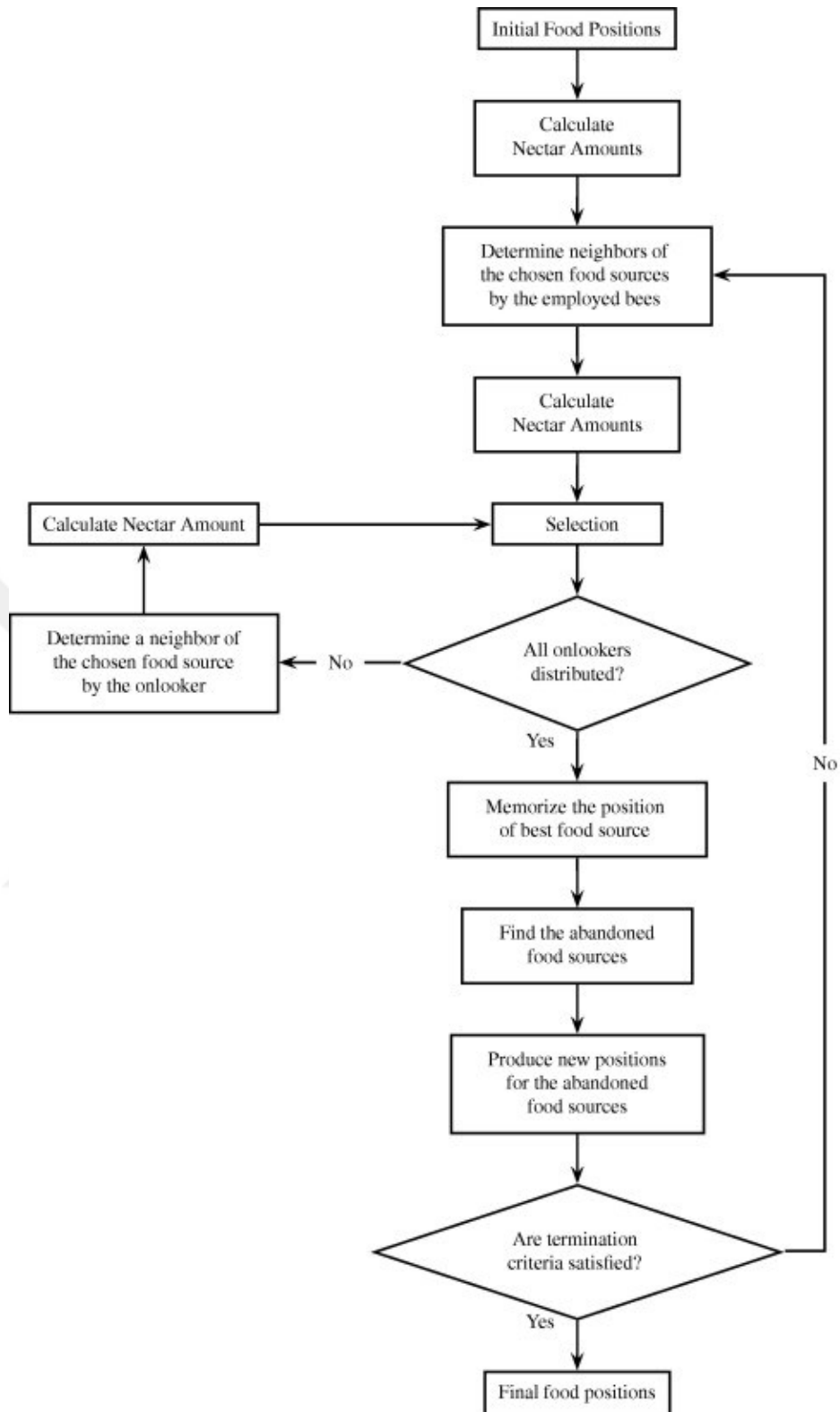


Figure 3.11 Flowchart of ABC algorithm (Akay & Karaboğa, 2012)

For using ABC algorithm as a feature selector, some changes must be applied. The algorithm is turned into binary version. In the algorithm, each probable solution which is food source is taken into account as a probable optimal feature subset for feature selection. These food sources represented as bit vectors have N-size where N is the total number of

feature. If the value is 1 in the position of the related feature, that means the corresponding feature is a part of the feature subset. If the related value is 0, then the feature is not included feature subset.

F1	F2	F3	F4	F5	F6	F7	F8	F9	F10
1	0	0	0	1	0	1	1	0	0

Figure 3.12 An example of bit vector for feature selection

From the example of bit vector version of food sources in Figure 3.12, it can be said that F1, F5, F7 and F8 features are selected for the feature subset.

Steps of the ABC algorithm for feature selection are shown in Table 3.6.

Table 3.6 Steps of ABC for feature selection

Steps	Processes
1.	The bit vectors that represent probable feature subsets with N-size is created, randomly. Then, the bit vectors are sent to classifier to get their fitness value.
2.	Each employed bee is assigned to bit vectors. The neighborhoods of the bit vectors are found with their fitness values.
3.	Greedy selection is applied between the original and its neighbor food sources.
4.	All onlookers are distributed to exploit food sources.
5.	Best food sources are memorized. Exhausted sources are replaced with new ones by scout bees, if any.
6.	If the termination criteria are met, the memorized best food source is the optimal solution. Otherwise, the algorithm continues by executing step 2.

To be able to obtain the fitness value of the food sources which represents the quality of the sources, classifiers are used as a cost function in the feature selector version of the ABC algorithm. F-measure result is considered as fitness value.

3.2.2.1 Classifiers used in ABC

In the algorithm, BayesNet, KNN, Random Tree, NaiveBayes, J48, Decision Table, SMO, Bagging, Random Forest, OneR classifiers have been used to get f-measure of features subnet and use it as fitness value.

BayesNet is also called Bayesian model, Bayesian network, decision network, Bayes network, belief network, or probabilistic directed acyclic graphical model. The classifier belongs to probabilistic graphical models' community. Nodes in the graph depict

a random variable and the edges between the nodes represent probabilistic dependencies among the related random variables (Ben-Gal, 2007).

In this classifier, examples are classified according to their nearest neighbors using KNN. It is considered Lazy Learning or called Example-Based Classification or Case-Based Classification because of that it employs the training examples directly for the classification process. Manhattan distance, Euclidean distance, and Chebyshev distance are three of the distance method to find neighbors (Cunningham & Delany, 2007).

Random Tree is a supervised classifier and an ensemble learning algorithm. For the creation of a decision tree, a bagging idea is used in the algorithm to construct a random set of data. The algorithm has been presented by Leo Breiman and Adele Cutler to deal with both classification and regression problems (Mishra & Ratha, 2016).

Naive Bayes is based on Bayes' theorem with naive independence assumptions among the features. It works even on small datasets, it is fast and not sensitive for irrelevant features (Unicamp, 2011).

J48 is developed by Ross Quinlan in 1993 as an extension of his earlier ID3 algorithm. The algorithm is used to generate a decision tree using the concept of information entropy for classification (Wikipedia, 2019, Quinlan, 1993).

Decision Table is a tabular form of the organized conditional logic. The conditional logic represents a bundle of actions to take as a result of tests that are referred by the conditional logic. In short, the actions are determined to depend on given conditions (DT Rules, 2012).

SMO is an algorithm proposed by Platt to solve the problems what arise during the training of SVM. It separates the problems into sub-problems using Osuna's theorem to secure the convergence. SMO breaks problems into a series of smallest possible sub-problems (Platt, 1998).

Bagging algorithm which stands for Bootstrap Aggregation selects samples from the original population to the aim of estimation of models or statistics. The process is to create random samples with replacements. The final model that produced by the algorithm is an aggregated model of all sample models that are trained on each of the bootstrap samples and this makes the bagging process an ensambling process (DnI Institute, 2016).

Random Forest algorithm proposed by Leo Breiman in 2000's (Breiman, s. 2001) is designed to create a predictor ensemble with a series of decision trees which grows using randomly selected subspaces of the dataset. The parts of the ensemble are tree-structured

predictors and each of the tree-structured predictor is built with the phenomena of randomness (Biau, 2012).

OneR which is a rule-based learner is a simple classification algorithm that creates just one rule for each predictor in the data. The rule that smallest total error belongs to is selected. The appearance of each class is count, most frequently appear class is found, the rule is assigned to the class for the value of the predictors, each predictor's total error of the rule is calculated and smallest total error is selected (Nasa & Suman, 2012).

3.2.3 Deep Learning

Deep learning is a method that allow send raw data to machine and find out the appropriate views needed for classification and detection (LeCun, Bengio, & Hinton, Deep Learning, 2015).

Deep learning is a version of Artificial Neural Network (ANN). ANN mimics human biological nervous system structurally and conceptually. ANN consists of the perceptrons which is one of the earliest neural network (Rezzak, Naz, & Zaib, 2017).

The perceptron contains at least one input, a bias, an activation function and one output. Inputs are received by the perceptron, they are multiplied by some weights and sent to activation function (Portilla, 2017). A view of the perceptron is shown in Figure 3.13.

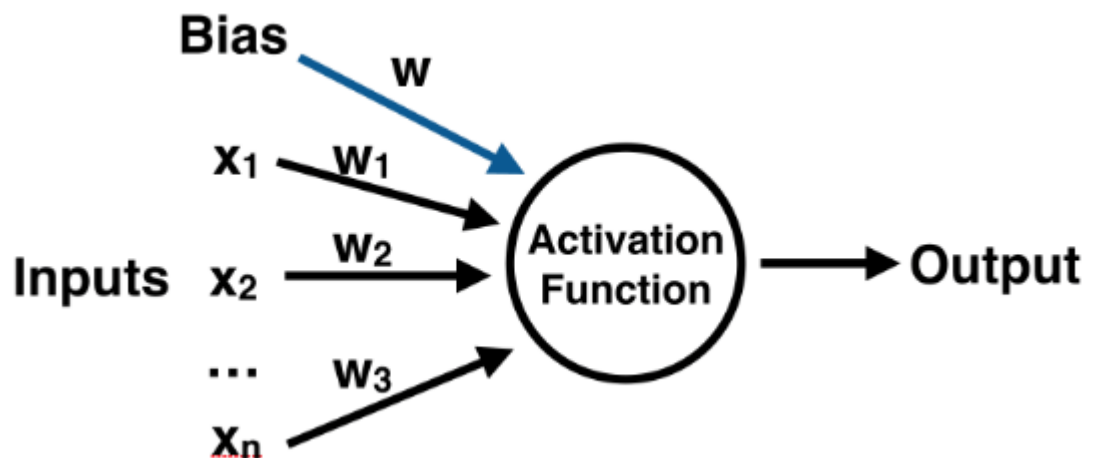


Figure 3.13 Representation of the perceptron

3.2.3.1 Activation Functions

Activation functions are mathematical equations that determine the output of the node with the given set of inputs. There are many activation functions divided into two groups as linear and non-linear function.

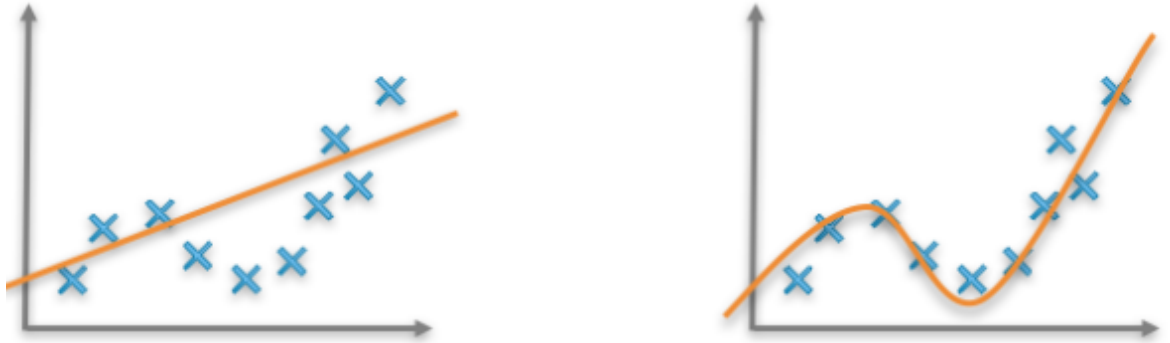


Figure 3.14 Linear (left) and non-linear (right) models

Straight line functions are linear functions and they are not used in neural networks because of that they can not capture complex patterns generally. Sigmoid function, hyperbolic tangent function (tanh), rectified linear unit (ReLU), leaky ReLU, softmax function can be count as non-linear function.

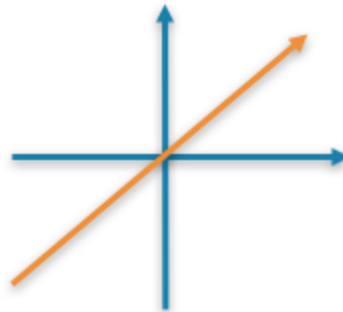


Figure 3.15 Linear function (Ronaghan, 2018)

Linear functions' range is $-\infty$ to $+\infty$.

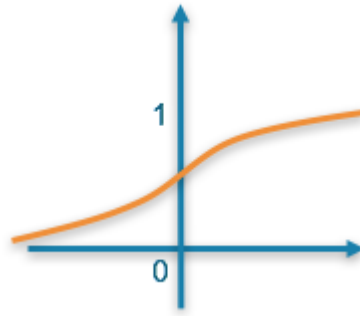


Figure 3.16 Sigmoid function (Ronaghan, 2018)

Range of sigmoid function is 0 to 1 and its formula is $sigmoid(z) = \frac{1}{1 + e^{-z}}$.

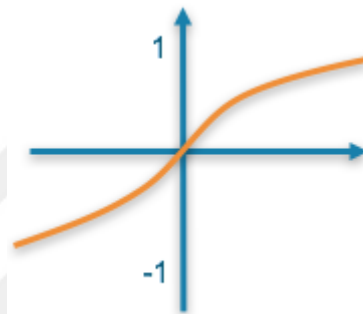


Figure 3.17 Hyperbolic Tangent Function (Ronaghan, 2018)

Tanh function's range is -1 to 1 and its formula is $tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}}$.

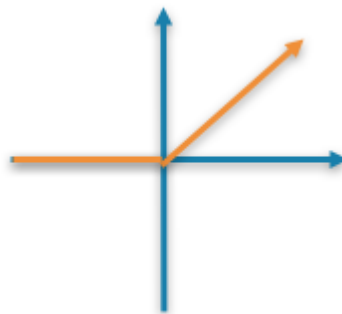


Figure 3.18 ReLU function (Ronaghan, 2018)

Range of ReLU is 0 to infinity and its formula is $relu(z) = \max(0, z)$.

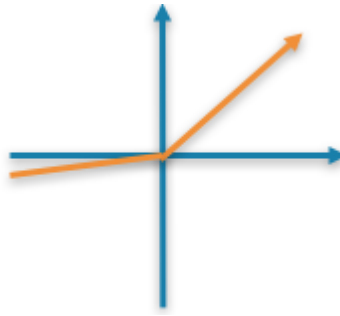


Figure 3.19 Leaky ReLU function (Ronaghan, 2018)

Leaky ReLU's range is $-\infty$ to ∞ and the formula is

$$\text{leakyrelu}(z) = \begin{cases} 0.01z, & z < 0 \\ z, & z \geq 0 \end{cases}$$

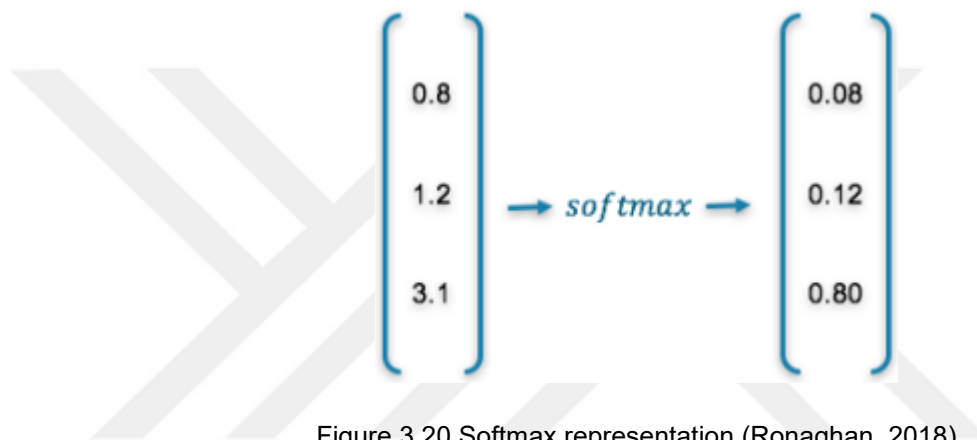


Figure 3.20 Softmax representation (Ronaghan, 2018)

Softmax function's range is between 0 and 1 and sum of all variables is 1. So, it can be used to model probability distribution. Generally, it is used in the output layer. Its formula is $\text{softmax}(z_i) = \frac{\exp(z_i)}{\sum_j \exp(z_j)}$.

A neural network consists of layers added together. Those layers are input layer that takes the input, output layer that creates results' output, and hidden layer that stays between the input and output layer and process the input. Visualization of the neural network is in Figure 3.21 (Portilla, 2017).

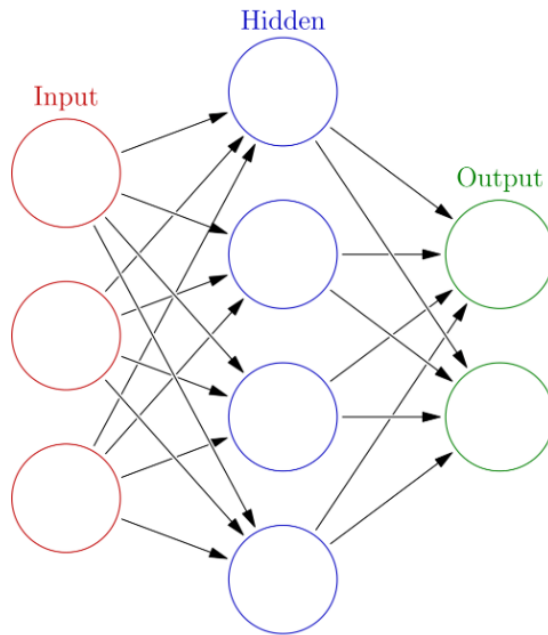


Figure 3.21 Visualization of Neural Network (Wikipedia, 2019)

To be able to solve more complex problems, more hidden layers are added to the neural network. This type of network called Deep Neural Network (DNN) takes input and perform data processing, and then the current layer's output is forwarded to the coming layer. More hidden layers help to deal with the complex problem with capturing non-linear relationships (Rezzak et al, 2017). A representation of DNN can be seen in Figure 3.22.

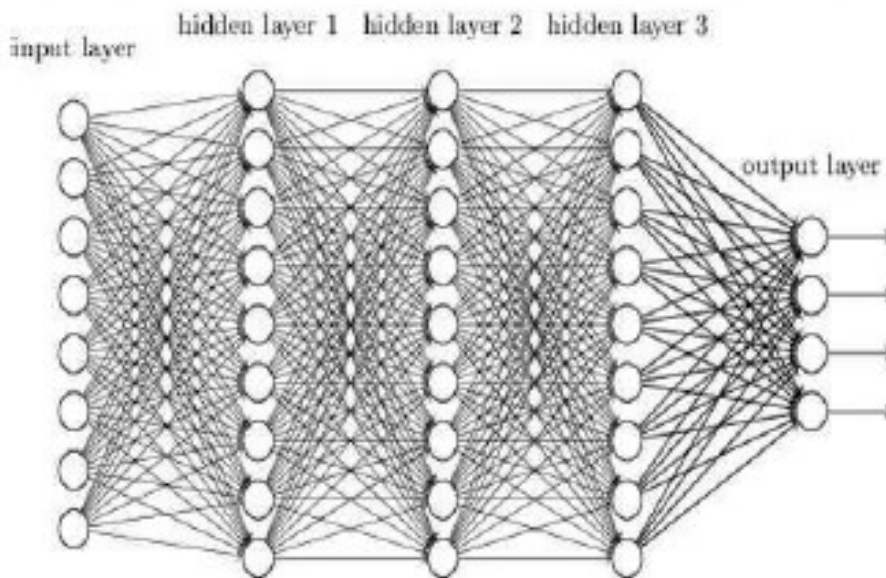


Figure 3.22 Deep Neural Network (Rezzak et al., 2017)

Different types of deep learning algorithms such as CNN, recurrent neural network (RNN), DNN, Deep Autoencoder (dA), deep Boltzmann Machine (DBM), deep belief network (DBN) are used by researchers (Rezzak et al., 2017).

Each type of NN is evolved for specific tasks and problems. Brief details, advantages, and disadvantages of types are shown in Table 3.7.

Table 3.7 Brief detail, Advantages and Disadvantages of Deep Learning models (Rezzak et al., 2017)

Type	Description	Advantage	Disadvantage
DNN	It has more than two layers to accomplish to find non-linear relationships for complex problems. It can be used for regression or classification.	It has been using widely by researchers with great accuracy.	The learning process is too much slow and they become so small. The model of the training diffuses the error to the back layers.
CNN	This is very good for 2-dimensional data. It has convolutional filters to convert 2D into 3D.	Good performance on computer vision tasks and speed of the learning model is fast.	The network needs quite a few data with the labels for the tasks.
RNN	Its powerful side is the qualification of learning of sequences.	It is the best for speech recognition, natural language processing, and character recognition related tasks.	It needs big datasets and has some problem because of gradient vanishing.
DBM	It has unidirectional connections between all hidden layers. It is based on Boltzmann's family.	The top-down feedback integrates with uncertain data to obtain more robust interference.	For the big dataset, parameter optimization is not possible.
DBN	The network used for both supervised and unsupervised machine learning.	In layers, a greedy strategy is used and it maximizes directly the likelihood.	Initialization stage is computationally expensive.
dA	It is used for unsupervised learning and evolved for the feature extraction and the dimensionality reduction. Unlike the others, the number of input and output is the same.	It does not need labeled data.	The training may suffer from vanishing and the pre-training step is needed.

3.2.3.2 Convolutional Neural Networks and Layers

CNN proposed by LeCun et al (1998) to be able to utilize in the computer vision area is a feed-forward artificial neural network. As mentioned above, a simple neural network consists of three layers as input, hidden and output. Unlike a simple neural network, CNN, which is based on simple neural networks, has some extra layers in the hidden layer. Those layers are such Convolution layer, Pooling layer, and Fully Connected layer.

Convolution layer takes the data and uses a kernel or filter to produce a feature map from the data. An example of a convolution process can be seen in Figure 3.23.

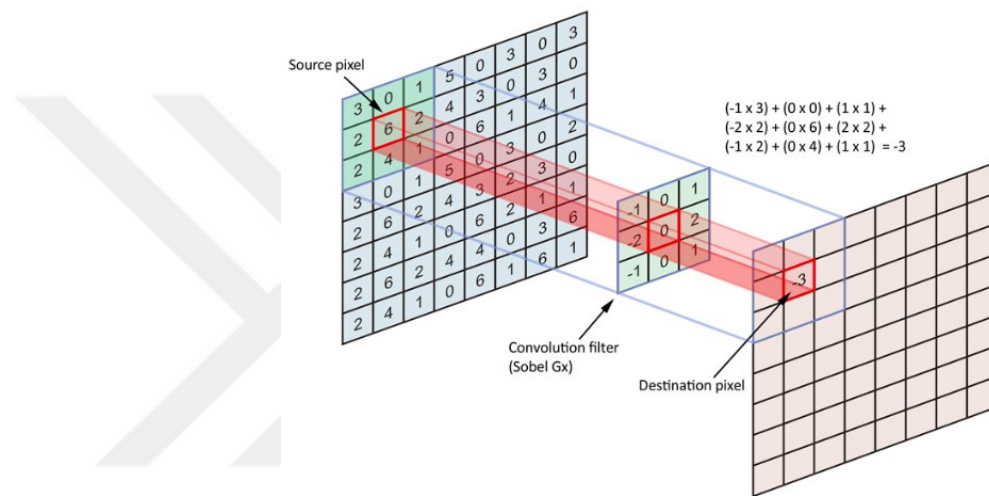


Figure 3.23 Example of convolution process

Pooling layer reduces the dimension of the feature map produced by convolutional layer by applying operations such called Max Pooling and Average Pooling which are most utilized pooling processes. Max Pooling takes maximum value while Average Pooling finds and uses the average of the values. Example of those pooling processes can be seen in Figure 24 and 25.

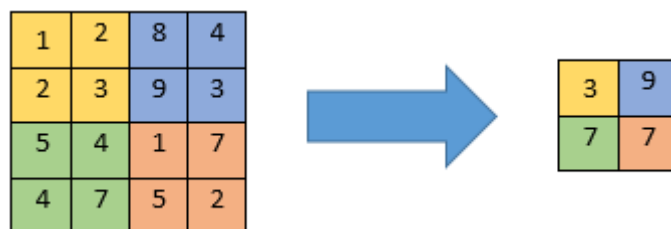


Figure 3.24 Example of Max Pooling process

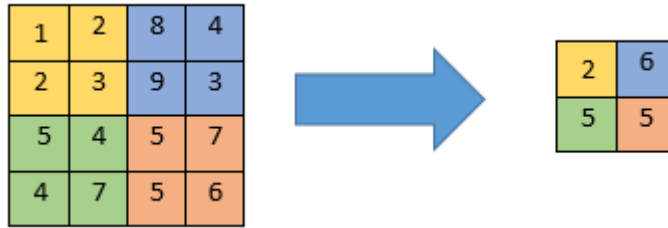


Figure 3.25 Example of Average Pooling process

Fully connected layer get the matrixes produced mentioned layers and classify the data by flattening the matrix.

Some of the most used types of CNN are VGGNet and ResNet (Rezzak et al., 2017). In this thesis, VGG16 and ResNet50 implementation have been used and results are compared.

3.2.3.3 VGG16

VGG16 is a CNN-type network proposed by Simonyan & Zisserman in 2014. They proposed the model for ImageNet Large Scale Visual Recognition Competition (ILSVRC). The model obtained 92.7% top-5 test accuracy in the ImageNet dataset that consists of over 14 million images belonging to 1000 class (Simonyan & Zisserman, 2014). Logical representation of the model can be seen in Figure 3.26 and structure of the model is shown in Figure 3.27.

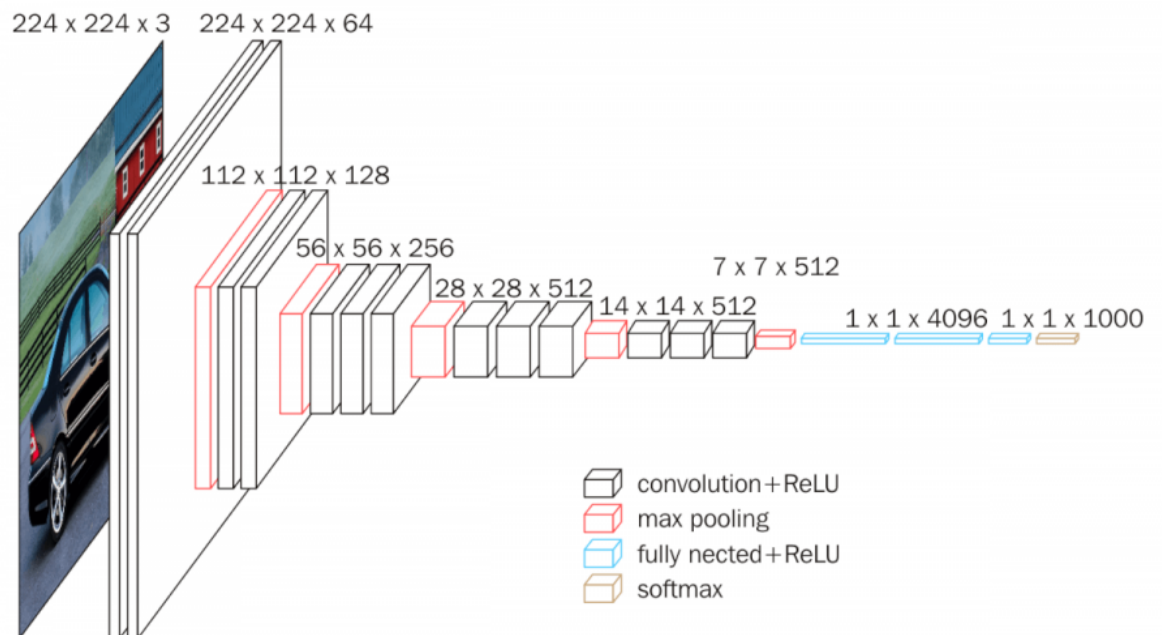


Figure 3.26 Logical representation of VGG (Neurohive, 2018)

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224×224 RGB image)					
conv3-64	conv3-64 LRN	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 conv1-256	conv3-256 conv3-256 conv3-256	conv3-256 conv3-256 conv3-256 conv3-256
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv1-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 conv1-512	conv3-512 conv3-512 conv3-512	conv3-512 conv3-512 conv3-512 conv3-512
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

Figure 3.27 VGG Architecture (Simonyan & Zisserman, 2014)

3.2.3.4 Residual network 50 (ResNet50)

ResNet network is proposed by He et al. in 2016 for ILSVRC 2015 and won the 1st place on the competition. Their network baselines are inspired by the philosophy of VGG nets, mainly. The model has lower complexity and fewer filters, according to their explanation (He et al., 2016).

The architecture of the model is shown in Figure 3.28.

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112			7×7, 64, stride 2		
				3×3 max pool, stride 2		
conv2_x	56×56	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
conv3_x	28×28	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 8$
conv4_x	14×14	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 23$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 36$
conv5_x	7×7	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$
	1×1			average pool, 1000-d fc, softmax		
FLOPs		1.8×10^9	3.6×10^9	5.8×10^9	7.6×10^9	11.3×10^9

Figure 3.28 The architecture of ResNet (He et al., 2016)

WM, GM, and CSF of MR images are generated using SPM 12 and sent to the deep learning systems. Example of those generated images are shown in Figure 3.29.

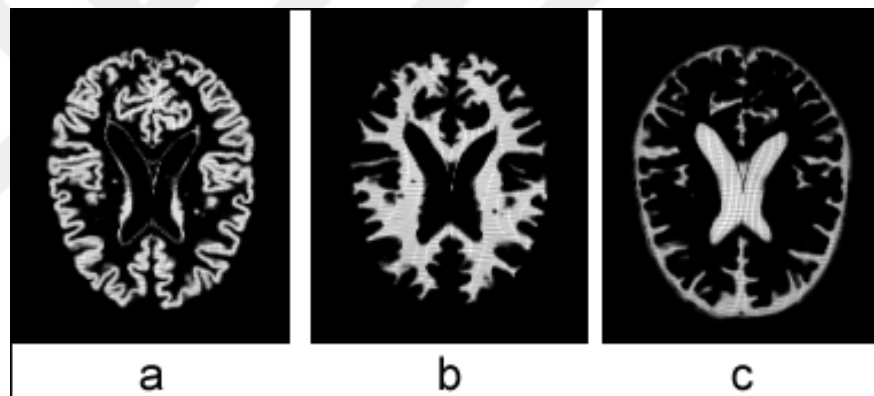


Figure 3.29 Example of the GM (a), WM (b) and CSF (c) generated using SPM 12 software

These deep learning implementations have been carried out using Python language, Komodo IDE and Keras library (Keras, 2019). Keras is a high-level neural networks API, written in Python and it has the ability to run on top of the TensorFlow (Tensorflow, 2019, Tensorflow-Github, 2019) framework which is an open source software library for numerical computation using data flow graphs and developed by Google. Tensorflow can work on the central processing unit (CPU) or on the graphics processing unit (GPU) and this affects the parallelism and computation time. In the thesis, NVIDIA GeForce 1050 Ti has been used as GPU on the computer with Intel Core I5-6400 2.70 GHz and 16 GB RAM. An example of a plotted result screen for VGG16 is as shown in Figure 3.30.

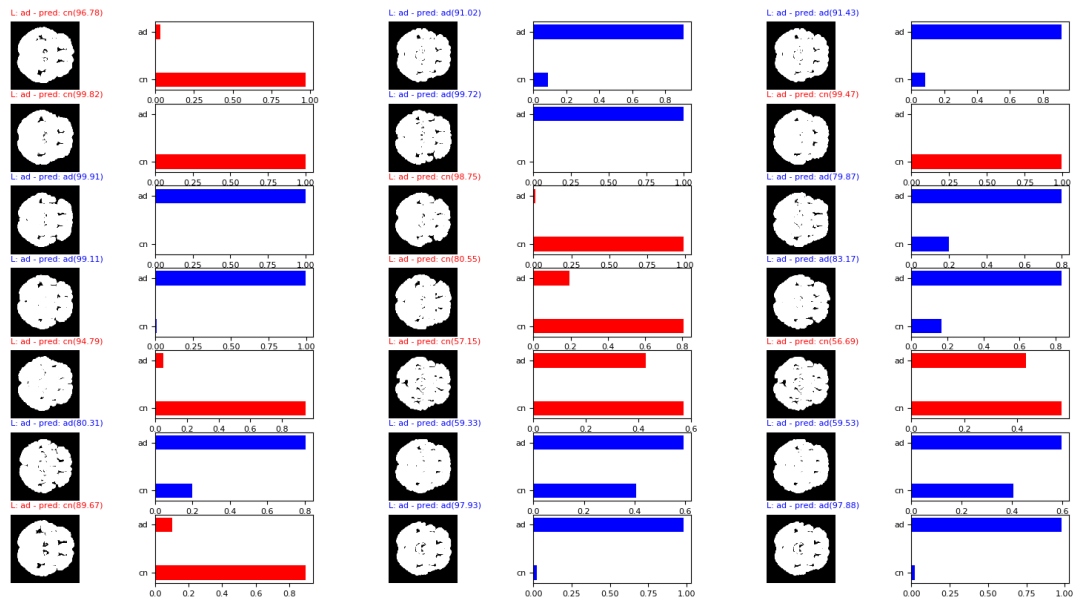


Figure 3.30 An example of plotted result screen

Results from VGG16 and ResNet50 are shown in Table 4.10 and Table 4.11. The algorithms are run for 10 times, values obtained from each of the run is in the table and the average of the runs can be seen below of the tables.

Batch size for both VGG16 and ResNet 50 is set as 16 and the learning rate is $1e-4$ as optimum. Iteration number is 50 for each run. All data have been split into training, validation, and test as 70%,20%, and 10% respectively.

3.2.4 Transfer Learning

Transfer learning is a technique to reuse a model that trained on one task for another task. It allows using different domains, tasks, and distributions in training and testing. That people can use their experience to solve new problems from previously learned motivates transfer learning (Pan & Yand, 2010). Briefly, in transfer learning, the system uses the weights that pre-trained with an excessive amount of data during the lots of time to solve new problems.

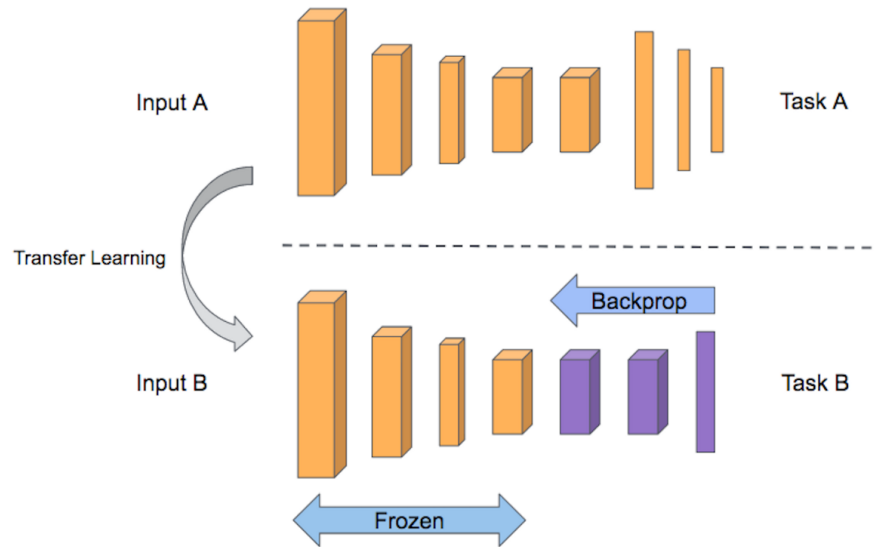


Figure 3.31 Representation of the Transfer Learning (Anonymous)

In transfer learning, firstly, a base network is trained on a base dataset and task. Then the learned features are transferred to a second target network to be trained on a target dataset and task. It tends to work if the features are suitable for both tasks (Yosinski et al., 2014).

As can be seen from Figure 3.31, all layers except the last one are frozen in the network and pre-trained weights are assigned to the frozen ones. The pre-trained weights come from the networks that trained for weeks with the distributed systems on the millions of the data such as Imagenet. Imagenet which is one of the most used weight sources for transfer learning is an ongoing image database project established to provide researchers an easily accessible image database. According to the statistics updated in 2010 lastly, there are 14.2 million images in the dataset. In the thesis, transfer learning is used in the deep learning stage and ImageNet weights are used as the source of the transfer learning (ImageNet, 2019).

4 RESULTS AND DISCUSSION

4.1 Evaluation Metrics

Performance metrics are used to be able to measure and compare the success of the algorithms. Binary classification is applied in this thesis and methods used in the experiments are measured by computing the metrics described under this title.

In the used dataset, there are two classes as AD or HC. Aforementioned concepts like precision, recall, etc. have been calculated the true positive (TP), true negative (TN), false positive (FP), false negative (FN). TP is correctly predicted positive value which means the actual class is HC and predicted label is HC. TN is correctly predicted negative values which means the actual label is AD and the predicted label is AD. FP means the actual label is AD but predicted label is HC while FN means the actual label is HC but predicted label is AD. Accuracy is ratio of the number of the correctly predicted labels to the total number of labels as shown in equation 5. (Powers & Ailab, 2011).

Table 4.1 Confusion Matrix

		Predicted Class	
		HC	AD
Actual Class	HC	TP	FN
	AD	FP	TN

$$Accuracy = \frac{(TP + TN)}{TP + TN + FP + FN} \tag{5}$$

The ratio of the number of correctly predicted positive labels to the number of totals predicted positive labels is called precision. Recall is found by dividing the number of correctly predicted positive labels to all labels in positive class. Recall also is called as sensitivity in the literature. Equation 6 and 7 show formulas of precision and recall respectively. F-measure which is called f1 score or f score in the literature is the weighted average of the precision and recall as shown in equation 8. (Powers & Ailab, 2011).

$$Precision = \frac{TP}{(TP + FP)} \tag{6}$$

$$Recall = \frac{TP}{(TP + FN)} \quad (7)$$

$$F - measure = 2 * \frac{(Recall * Precision)}{(Recall + Precision)} \quad (8)$$

AUC - ROC Curve is a measurement for performance of classification problem at various thresholds. It shows the situation of the capability of the model for distinguishing between classes. A higher value means better model at distinguishing between classes (Narkhede, 2018).

RMSE is prediction error's standart deviation. RMSE gives information about how the data close to the best fit line. The formula can be seen in equation 9 (Data Science Central, 2016).

$$RMSE = \sqrt{(f - o)^2} \quad (9)$$

Where f is expected value and o is actual value.

4.2 Result Tables

In this section, accuracy, f-measure, precision, recall, ROC Area and Root Mean Squared Error are given per algorithm in the tables. For IG, GR and CFS, measurements have been acquired by Weka software. As said before, ABC has been coded in Java language and measurements have been obtained by Weka's Java library. Scikit-learn library (<https://scikit-learn.org/stable/index.html>) has been used to be able to have the aforementioned measurements in deep learning with Python language.

The volumetry statistical data obtained from volBrain has 112 features consist of asymmetry, cm3 and percentage information of the brain regions. Results of the raw dataset can be seen in Table 4.2. Weka is used to produce results. K value for KNN is set to 3. 10 fold-cross validation is used for all algorithms.

Table 4.2 Result of the raw dataset with the mentioned algorithms

Raw Dataset + DM						
Algorithm	Accuracy (%)	Precision	Recall	F-Measure	ROC Area	RMSE
BayesNet	83.6991	0.841	0.837	0.837	0.902	0.3959
Naive Bayes	78.6834	0.798	0.787	0.787	0.872	0.4497
SMO	79.3103	0.794	0.793	0.793	0.792	0.4549
KNN	73.6677	0.737	0.737	0.737	0.765	0.4505
Bagging	82.4451	0.825	0.824	0.825	0.926	0.3284
OneR	80.2508	0.806	0.803	0.803	0.805	0.4444
J48	81.8182	0.818	0.818	0.818	0.821	0.4168
Decision Table	85.5799	0.856	0.856	0.856	0.898	0.3464
Random Forest	85.8934	0.859	0.859	0.859	0.926	0.3349
Random Tree	79.6238	0.798	0.796	0.797	0.797	0.4514

Random Forest algorithm has the best results for Raw Dataset + DM in means of accuracy, precision, recall, and f-measure. Bagging has the best value for RMSE while Random Forest and Bagging share the best result for the ROC area.

Before using ABC as feature selector, classical feature selection methods such as Info Gain (IG), Gain Ratio (GR) and Correlation based Feature Selection (CFS) has been used to select feature. The results with the selected features are shown in Table 4.3.

Table 4.3 Results after feature selection with Info Gain, Gain Ratio and CFS

Dataset after Feature Selection With Classical Methods + DM																		
Algorithm	Accuracy (%)			Precision			Recall			F-Measure			ROC Area			RMSE		
	IG	GR	CFS	IG	GR	CFS	IG	GR	CFS	IG	GR	CFS	IG	GR	CFS	IG	GR	CFS
BayesNet	83.386	85.266	87.147	0.843	0.857	0.873	0.834	0.853	0.871	0.834	0.853	0.872	0.903	0.919	0.936	0.395	0.374	0.347
Naive Bayes	83.386	86.207	88.715	0.837	0.842	0.887	0.834	0.862	0.887	0.834	0.862	0.887	0.889	0.919	0.933	0.392	0.336	0.317
SMO	83.699	84.012	86.207	0.838	0.842	0.863	0.837	0.840	0.862	0.837	0.840	0.862	0.837	0.841	0.862	0.404	0.400	0.371
KNN	80.878	84.012	82.445	0.809	0.843	0.828	0.809	0.840	0.824	0.809	0.840	0.825	0.843	0.865	0.870	0.402	0.384	0.381
Bagging	83.386	86.207	87.147	0.835	0.864	0.872	0.834	0.862	0.871	0.834	0.862	0.872	0.907	0.921	0.939	0.344	0.323	0.311
OneR	81.191	80.251	80.251	0.814	0.806	0.806	0.812	0.803	0.803	0.812	0.803	0.803	0.813	0.805	0.805	0.434	0.444	0.444
J48	83.699	85.580	84.953	0.837	0.856	0.849	0.837	0.856	0.850	0.837	0.856	0.849	0.821	0.840	0.833	0.377	0.368	0.374
Decision Table	79.937	85.893	87.147	0.799	0.860	0.872	0.799	0.859	0.871	0.799	0.859	0.872	0.887	0.903	0.908	0.368	0.338	0.334
Random Forest	82.759	87.147	87.461	0.828	0.872	0.875	0.828	0.871	0.875	0.828	0.872	0.875	0.901	0.923	0.942	0.354	0.328	0.305
Random Tree	76.489	75.862	81.818	0.766	0.761	0.822	0.765	0.759	0.818	0.765	0.759	0.819	0.764	0.759	0.821	0.485	0.491	0.426

It can be seen from Table 4.3 that CS shows superior performance.

Table 4.4 Features selected by IG, GR and CFS methods

Method	Selected Features No.
IG	92, 93, 97, 95, 101, 94, 104, 96, 100,41
GR	93, 97, 95, 96, 92, 101, 100, 103, 99, 94
CFS	92, 93, 97, 96, 101, 35, 2, 30, 95, 103

In Table 4.4, selected features' number by IG, GR and CFS methods can be seen. Full features names and brief explanations can be found in Appendix-B.

Aforementioned algorithms were used in the ABC algorithms and results for each algorithm are shown in Table 4.5. The number of the selected features with the related algorithm can be seen in the second column and and the row numbers of selected features for each algorithm can be seen in Table 4.6. According to those row number, name and description of each feature can be found in Appendix - B..

Table 4.5 Results after feature selection with ABC algorithm

Dataset after Feature Selection With ABC Algorithm + DM							
Algorithm	No. of S.F.	Accuracy (%)	Precision	Recall	F-Measure	ROC Area	RMSE
BayesNet	6	89.0282	0.890	0.890	0.890	0.922	0.3117
Naive Bayes	5	89.3412	0.893	0.893	0.893	0.936	0.3028
SMO	30	88.7147	0.888	0.887	0.887	0.888	0.3359
KNN	5	86.8339	0.871	0.868	0.869	0.896	0.349
Bagging	5	88.0878	0.881	0.881	0.881	0.925	0.3135
OneR	1	83.0721	0.831	0.831	0.830	0.826	0.4114
J48	5	88.0878	0.883	0.881	0.881	0.849	0.3271
Decision Table	4	88.7147	0.887	0.887	0.887	0.918	0.3232
Random Forest	30	89.3417	0.893	0.893	0.893	0.945	0.3134
Random Tree	1	83.0721	0.831	0.831	0.830	0.866	0.3664

In ABC algorithm for feature selection, Random Forest slightly better than Naive Bayes in means of accuracy, precision, recall and f-measure. But Random Forest performs its performance with 30 features while Naive Bayes does with just five features. It is worth to mention that Random Tree and OneR have 83% accuracy by selecting just one feature.

Table 4.6 Features selected by ABC feature selector algorithm

Algorithm	Selected Features No.
BayesNet	18, 35, 79, 92, 97, 107
Naive Bayes	35, 66, 92, 97, 100
SMO	2, 3, 6, 7, 13, 18, 20, 23, 26, 30, 32, 33, 34, 35, 38, 49, 50, 51, 52, 72, 73, 75, 79, 81, 92, 95, 97, 102, 107, 112
KNN	35, 95, 97, 99, 100
Bagging	3, 35, 76,, 95, 97
OneR	97
J48	23, 35, 73, 93, 103
Decision Table	35, 92, 95, 97
Random Forest	2, 3, 13, 15, 17, 18, 30, 32, 34, 35, 38, 40, 42, 47, 50, 53, 66, 72, 75, 79, 83, 89, 90, 95, 97, 100, 102, 107, 108, 112
Random Tree	97

Selected features by ABC algorithm is shown in Table 4.6. It can be seen from Table 4.6 that Hippocampal statistics which are between number 92 and 98 are the most selected features by selecting all of ten algorithms. Also, Cerebrum Asymmetry, whose number is 35, is selected by eight of ten algorithms. Cerebrum Asymmetry feature has been added to the feature list of OneR and Random Tree algorithms which have not selected related feature and new results have been obtained. The results are in Table 4.7.

Table 4.7 Result of OneR and Random Tree with added Cerebrum Asymmetry feature

Algorithm	No. of S.F.	Accuracy (%)	Precision	Recall	F-Measure	ROC Area	RMSE
OneR	2	83.0721	0.831	0.831	0.830	0.826	0.4114
Random Tree	2	82.7586	0.828	0.282	0.828	0.826	0.4152

Adding the feature to the OneR has not changed its results, but adding the feature to the Random Tree has decreased the performance.

Another taking attention result is from Naive Bayes. It has obtained almost the same performance with the best one but has used five features. In means of time, the raw dataset has been classified in 0.02 second and selected five features have been classified in 0.01 second in WEKA using Naive Bayes. In the feature selection process, each iteration for Naive Bayes has taken 3 minutes, approximately. Same way, time for deep learning has been measured. For ResNet50 and VGG16, each iteration has completed in about 4 minutes 30 seconds and 6 minutes, respectively. This comparison may give some clues for time performance but cannot give exact explication, also. Because the feature selection process has been implemented in JAVA language that runs on CPU while the codes of deep learning implementations run on GPU. That information can be found also in Table 4.8 and 4.9.

Table 4.8 Classification time information for Naive Bayes in WEKA

Dataset	Time
Raw Dataset	0.01 second
Dataset created after Feature selection	0.02 second

Table 4.9 Time information for ABC, VGG16 and ResNet50

Implementation	Time ()
ABC for feature selection	≈3 minutes per iteration
VGG16	≈4 minutes 30 seconds per iteration
ResNet50	≈6 minutes per iteration

When considered IG, GR, and CSF feature results, it is understood that Hippocampal and Amygdala statistics are selected for each of the methods.

As such found in the studies conducted by Raji et. al. (2009), Lindberg et al. (2012), Plant et al. (2010), Poulin et al. (2011), Zhang et al. (2011), Zhou et al. (2014), Sorensen et al. (2017) hippocampal statistics show a consistent connection with the disease and our results are coherent with the studies.

Table 4.10 VGG16 Results with GM of MRI Data

Time	Accuracy (%)	Precision	Recall	F-Measure	ROC Area	RMSE
1	62.141	0.620	0.619	0.619	0.619	0.405
2	61.710	0.621	0.617	0.615	0.617	0.402
3	72.210	0.721	0.724	0.724	0.725	0.356
4	63.244	0.632	0.632	0.632	0.632	0.410
5	59.631	0.601	0.596	0.592	0.596	0.412
6	64.262	0.647	0.643	0.640	0.643	0.401
7	60.464	0.613	0.604	0.597	0.604	0.420
8	62.132	0.621	0.621	0.621	0.621	0.410
9	62.408	0.628	0.624	0.621	0.624	0.408
10	64.444	0.643	0.643	0.643	0.643	0.399
Average	63.265	0.635	0.632	0.630	0.632	0.402

Table 4.11 ResNet50 Results with GM of MRI Data

Time	Accuracy (%)	Precision	Recall	F-Measure	ROC Area	RMSE
1	50.190	0.750	0.502	0.337	0.502	0.498
2	50.092	0.750	0.502	0.335	0.502	0.499
3	55.542	0.518	0.555	0.580	0.550	0.431
4	51.141	0.581	0.511	0.376	0.511	0.461
5	50.422	0.250	0.505	0.334	0.503	0.488
6	52.778	0.542	0.527	0.482	0.527	0.465
7	53.877	0.750	0.523	0.341	0.523	0.463
8	55.566	0.550	0.555	0.601	0.550	0.440
9	50.211	0.512	0.501	0.366	0.501	0.485
10	46.964	0.460	0.469	0.438	0.470	0.510
Average	51.678	0.566	0.515	0.419	0.514	0.474

According to VGG16 and ResNet50 results, VGG16 has superior performance with 63.265 % average accuracy while ResNet50 has 51.678 % average accuracy which is, unfortunately, a low rate for binary classification.

5 CONCLUSION

In this thesis, the classification of brain MRI data has been aimed. For this purpose, MRI data have been obtained from ADNI dataset. Those collected MRI data has been divided into two groups as AD and HC that means Alzheimer's Disease and healthy control, respectively. To be able to acquire the statistical volumetry data from the MRI, the collected and grouped data have been sent to the volBrain system which is an online brain volumetry system developed to help researchers work in this area. The system produces statistical data such as volume, asymmetry according to brain parts and returns a directory that contains segmented files in nii format and volumetry data in CSV format. The process has been repeated for each patient's MRI data and acquired files are merged. So, the dataset with 112 features has been created in arff format which is compatible with WEKA and our ABC based feature selection algorithm. Firstly, the created dataset has been classified without any feature selection. Then, three traditional feature selection methods Info Gain, Gain Ratio and Correlation-Based Feature Selection has been used to select the most relevant features. The dataset with the only selected features has been classified and the results have been noted down. Afterward, ABC feature selection algorithm which had been coded for the thesis has been employed for the feature selection process. The dataset that included only selected features by ABC have been classified and results have been saved to compare.

All the three results obtained from the raw dataset and selected features show that classical methods have better classification performance with the fewer features than the raw dataset. In the same way, ABC-based feature selection method has better classification performance than classical methods.

To have a look at the deep learning performance, two CNN-based architecture, VGG16, and ResNet50, have been implemented using Python language, Tensorflow and Keras framework and Komodo IDE. Collected and grouped raw MRI data have been segmented as GM, WM, CSF using SPM 12 software. Then those segmented MRI slices have been sent to the aforementioned deep learning implementations. The results have not been satisfying, unfortunately. Most of the deep learning studies that produced good results in this area have been conducted with the more data, more modalities like PET, MMSE, CSF, APOE, SNPs, and the experts in the medical and dementia. Those can be the reason for the unsatisfying results in the deep learning results.

Considering all the results produced from experiments, it can be said that ABC-based feature selection method has produced the best performance.

Taking into account the studies that use one modality like this thesis, the classification performance of our results are promising and can be improved.

Other findings attained from the thesis are about selected brain area in the feature selection stages. The results of the selected brain parts show that hippocampal and amygdala are the most related brain parts with the disease. Also, it can be said that Cerebrum statistics should be considered while assessing the disease. It is the principal and most anterior part of the brain in vertebrates, located in the front area of the skull and consisting of two hemispheres, left and right, separated by a fissure and it is responsible for the integration of complex sensory and neural functions and the initiation and coordination of voluntary activity in the body. Findings of the thesis about brain parts are consistent with other studies carried out to discover the most relevant brain parts.

For future works, unlike the thesis, using more modalities may be beneficial to improve the results. The more modalities and clues about the patients, the more robust results. Furthermore, having a good sample pool size is another key point of the good results. Feeding the algorithms with the increased number and quality of data can help to obtain good classification results. Because finding appropriate data is one of the most compelling problems for the studies of the medical area.

REFERENCES

- Aguilar, C., Westman, E., Muehlboeck, S., Mecocci, P., Vellas, B., Tsolaki, M., . . . Wahlund, L. (2013). Different multivariate techniques for automated classification of MRI data in Alzheimer's disease and mild cognitive impairment. *Psychiatry Research: Neuroimaging*(212), 89-98. doi:10.1016/j.psychresns.2012.11.005
- Akay, B., & Karaboğa, D. (2012). A modified ABC algorithm for real-parameter optimization. *Information Sciences*, 192, 120-142. doi:10.1016/j.ins.2010.07.015
- Alzheimer's Disease International. (2015). *World Alzheimer Report 2015*. London.
- Alzheimer's Disease International. (2018). *World Alzheimer Report 2018*. London.
- Basir, M., & Ahmad, F. (2014). Comparison on Swarm Algorithms for Feature Selections Reductions. *International Journal of Scientific & Engineering Research*, 5(8), 479-486.
- Behesti, I., & Demirel, H. (2016). Feature-ranking-based Alzheimer's disease classification from structural MRI. *Magnetic Resonance Imaging*(34), 252-263. doi:10.1016/j.mri.2015.11.009
- Behesti, I., Demirel, H., & Matsuda, H. (2017). Classification of Alzheimer's disease and prediction of mild cognitive impairment-to-Alzheimer's conversion from structural magnetic resonance imaging using feature ranking and a genetic algorithm. *Computers in Biology and Medicine*(83), 109-119. doi:10.1016/j.compbiomed.2017.02.011
- Ben-Gal, I. (2007). Bayesian Networks. *Encyclopedia of Statistics in Quality & Reliability*. doi:10.1002/9780470061572.eqr089
- Biau, G. (2012). Analysis of a Random Forest Model. *Journal of Machine Learning Research*, 13, 1063-1095.
- Breiman, L. (2011). *Random Forest*. stat.berkeley.edu: <https://www.stat.berkeley.edu/~breiman/randomforest2001.pdf> adresinden alındı
- busamra, H. (2013). A comparative study of feature selection and classification methods for gene expression data of glioma. *Procedia Computer Science*, 23, 5-14.
- Cheng, D., & Liu, M. (2017). CNNs Based Multi-Modality Classification for AD Diagnosis. *2017 10th International Congress on Image and Signal Processing, BioMedical*

Engineering and Informatics (CISP-BMEI), (s. 1-5). Shanghai. doi:10.1109/CISP-BMEI.2017.8302281

- Cunningham, P., & Delany, S. (2007). k-Nearest neighbour classifiers. *Mult. Classif. Syst.*
- Data Science Central. (2016). *RMSE: Root Mean Square Error - Statistics How To*. DataScienceCentral: <https://www.statisticshowto.datasciencecentral.com/rmse/> adresinden alındı
- Dementia Australia. (2019). *Dementia Australia, Early diagnosis of dementia*. April 2019 tarihinde <https://www.dementia.org.au/information/diagnosing-dementia/early-diagnosis-of-dementia> adresinden alındı
- Dessi, N., Pascariello, E., & Pes, B. (2003). A comparative Analysis of Biomarker Selection Techniques. *BioMed Research International*.
- Dnl Institute. (2016). *Bagging Algorithm: Concepts with Example - Dnl Intitute*. Dnl Institute - Build Data and Decision Science Experience: <http://dni-institute.in/blogs/bagging-algorithm-concepts-with-example/> adresinden alındı
- DT Rules. (2012). *Decision Table DTRules*. DTRules: <http://www.dtrules.com/newsite/?p=90> adresinden alındı
- Farooq, A., Anwar, S., Awais, M., & Rehman, S. (2017). A deep CNN based multi-class classification of Alzheimer's disease using MRI. *2017 IEEE International Conference on Imaging Systems and Techniques (IST)*, (s. 1-6). Beijing.
- Forman, G. (2003). An extensive empirical study of feature selection metrics for text classification. *Journal of Machine Learning Research*, 10, 1289-1305.
- Gonzalez, R., & Woods, R. (1993). *Digital Image Processing*. Addison-Wesley.
- Gunawardena, N., Rajapakse, R., & Kodikara, N. (2017). Applying convolutional neural networks for pre-detection of alzheimer's from structural MRI data. *2017 24th International Conference on Mechatronics and Machine Vision in Practice (M2VIP)*.
- Hand, D., Manilla, H., & Smyth, P. (2001). *Principle of Data Mining*. London: The MIT Press.
- He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep residual learning for image recognition. 770-778. doi:arXiv:1512.03385

- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. doi:10.1109/CVPR.2016.90
- Hinrichs, C., Singh, V., Xu, G., & Sterling, C. J. (2011). Predictive markers for AD in a multi-modality framework: An analysis of MCI progression in the ADNI population. *NeuroImage*(55), 574-589. doi:10.1016/j.neuroimage.2010.10.081
- ImageNet. (2019). *ImageNet*. ImageNet: <http://www.image-net.org> adresinden alındı
- Inza, I., Larranga, P., Blanco, R., & Cerrolaza, A. (2004). Filter versus wrapper gene selection approaches in dna microarray domains. *Artificial Intelligence in Medicine*, 31, 91-103.
- IPF Radiology Rounds. (2019). *Axial, Coronal and Saggital Planes*. IPF Radiology Rounds: <https://www.ipfradiologyrounds.com/hrct-primer/image-reconstruction/> adresinden alındı
- Islam, J., & Zhang, Y. (2018). Brain MRI analysis for Alzheimer's disease diagnosis using an ensemble system of deep convolutional neural networks. *Brain Informatics*, 5(2). doi:10.1186/s40708-018-0080-3
- Janabi, K., & Kadhim, R. (2018). Data Reduction Techniques: A Comparative Study for Attribute Selection Methods. *International Journal of Advanced Computer Science and Technology*, 8(1), 1-13.
- Karaboga, D., & Basturk, B. (2007). A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm. *J. Glob. Optim.*, 39, 459-471. doi:10.1007/s10898-007-9149-x
- Keras. (2019). *GitHub - keras-team/keras: Deep learning for humans*. Keras Github: <https://github.com/keras-team/keras> adresinden alındı
- Klöppel, S., Stonnington, C. M., Chu, C., Draganski, B., Scakill, R. I., Rohrer, J. D., . . . Frackowiak, R. S. (2008). Automatic Classification of MR Scans in Alzheimer's Disease. *Brain: a journal of neurology*(131), 681-689. doi:10.1093/brain/awm319
- Klöppel, S., Stonnington, C., Barnes, J., Chen, F., Chu, C., Good, C., . . . Frackowiak, R. (2008). Accuracy of dementia diagnosis - a direct comparison between radiologists and computerized method. *Brain*, 131, 2969-2974. doi:10.1093/brain/awn239
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep Learning. *Nature*, 521, 434-444. doi:10.1038/nature14539

- LeCun, Y., Bottou, L., Bengio, Y., & Haffnet, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of IEEE*, 86(11), 2278-2324.
- Lee, W., Stolfo, S., & Mok, K. (2000). Adaptive intrusion detection: A data mining approach. *AI Review*, 14(6), 533-567.
- Lin, W., Tong, T., Gao, Q., Guo, D., Du, X. Y., Guo, G., . . . Alzheimer's Disease Neuroimaging Initiative. (2018). Convolutional Neural Networks-Based MRI Image Analysis for the Alzheimer's Disease Prediction From Mild Cognitive Impairment. *Front. Neurosci.*, 12:777. doi:10.3389/fnins.2018.00777
- Lindberg, O., Walterfang, M., Looi, J., Malykhin, N., Östberg, P., Zandbelt, B., . . . Wahlund, L. (2012). Shape analysis of the hippocampus in Alzheimer's disease and subtypes of frontotemporal lobar degeneration. *J Alzheimers Dis.*, 355-365. doi:10.3233/JAD-2012-112210
- Liu, C., Jiang, D., & Yang, W. (2014). Global geometric similarity scheme for feature selection in fault diagnosis. *Expert Systems with Applications*, 41(8), 3585-3595.
- Liu, T., Liu, S., & Chen, Z. (2003). An evaluation on feature selection for text clustering. *20th International Conference on Machine Learning (ICML-2003)* (s. 488-495). Washington DC, USA: AAAI Press.
- Long, X., Chen, L., Jiang, C., Zhang, L., & Alzheimer's Disease Neuroimaging Initiative. (2017). Prediction and classification of Alzheimer disease basen on quantification of MRI deformation. *PLoS ONE*, 12(3). doi:10.1371/journal.pone.0173372
- Luo, S., Li, X., & Li, J. (2017). Auromatic Alzheimer's Disease Recognition from MRI Data Using Deep Learning Method. *Journal of Applied Mathematics and Physics*, 5, 1892-1898. doi:10.4236/jamp.2017.59159
- Manjon, J., & Coupe, P. (2016). volBrain: an online MRI brain volumetry system. *Frontiers in Neuroinformatics*. doi:10.3389/fninf.2016.00030
- Maroco, J., Silva, D., Rodrigues, A., Guerreiro, M., Santana, I., & Mendonca, A. (2011). Data mining methods in the prediction of Dementia: A real-data comparison of the accuracy, sensitivity and specificity of linear discriminant analysis, logistic regression, neural networks, support vector machines, classification trees and random forests. *BMC Research Notes*, 4(299). doi:10.1186/1756-0500-4-299
- Mishra, A., & Ratha, B. (2016). Study of Random Tree and Random Forest Data Mining Algorithms for Microarray Data Analysis. *International Journal on Advanced Electrical and Computer Engineering (IJAECE)*, 3(4), 5-7.

- Mustra, M., Grgic, M., & Delac, K. (2012). Breast density classification using multiple feature selection. *Automatika*, 53, 1289-1305.
- Narkhede, S. (2018). *Understanding AUC - ROC Curve -Towards Data Science*. Towards Data Science - Medium: <https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5> adresinden alındı
- Nasa, C., & Suman. (2012). Evaluation of Different Classification Techniques for WEB Data. *International Journal of Computer Applications*, 52(9), 34-40.
- National Institute of Biomedical Imaging and Bioengineering. (2019). *Magnetic Resonance Imaging (MRI) | National Institute of Biomedical Imaging and Bioengineering*. National Institute of Biomedical Imaging and Bioengineering: <https://www.nibib.nih.gov/science-education/science-topics/magnetic-resonance-imaging-mri> adresinden alındı
- Neurohive. (2018). *VGG16 - Convolutional Network for Classification and Detection*. Neurohive - Neural Networks: <https://neurohive.io/en/popular-networks/vgg16/> adresinden alındı
- Ota, K., Oishi, N., Ito, K., Fukuyama, H., & Group, t. S.-J. (2014). A comparison of three brain atlases for MCI prediction. *Journal of Neuroscience Methods*(221), 139-150. doi:10.1016/j.jneumeth.2013.10.003
- Pan, S., & Yand, Q. (2010). A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10), 1345-1359.
- Peng, J., Zhu, X., Wang, Y., An, L., & Shen, D. (2019). Structured sparsity regularized multiple kernel learning for Alzheimer's disease diagnosis. *Pattern Recognition*, 88, 370-382. doi:10.1016/j.patcog.2018.11.027
- Plant, C., Teipel, S. J., Oswald, A., Böhm, C., Meindl, T., Mourao-Miranda, J., . . . Ewers, M. (2010). Automated detection of brain atrophy patterns based on MRI for the prediction of Alzheimer's disease. *NeuroImage*(50), 162-174. doi:10.1016/j.neuroimage.2009.11.046
- Platt, J. (1998). *Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machine*.
- Portilla, J. (2017). *A beginner's Guide to Neural Networks in Python | SpringBoard Blog*. springboard.com: <https://www.springboard.com/blog/beginners-guide-neural-network-in-python-scikit-learn-0-18/> adresinden alındı

- Poulin, S. P., Dautoff, R., Morris, J. C., Barrett, L. F., & Dickerson, B. C. (2011). Amygdala atrophy is prominent in early Alzheimer's disease and relates to symptom severity. *Psychiatry Research*(194), 7-13. doi:10.1016/j.psychres.2011.06.014
- Powers, D., & Ailab. (2011). Evaluation: From precision, recall and F-measure to ROC, informedness, markedness & correlation. *Journal of Machine Learning Technologies*, 2(1), 37-63. doi:10.9735/2229-3981
- Quinlan, J. (1993). *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers.
- Raji, C., Lopez, O., Kuller, L., Carmichael, O., & Becker, J. (2009). Age, Alzheimer disease and brain structure. *Neurology*, 1899-1905. doi:10.1212/WNL.0b013e3181c3f293
- Rezzak, M., Naz, S., & Zaib, A. (2017). Deep Learning for Medical Image Processing: Overview, Challenges and Future. *In Classification in BioApps*, 323-350.
- Ronaghan, S. (2018). *Deep Learning: Overview of neurons and activation functions*. medium.com: <https://medium.com/@srnghn/deep-learning-overview-of-neurons-and-activation-functions-1d98286cf1e4> adresinden alındı
- Sarraf, S., & Tofighi, G. (2016). DeepAD: Alzheimer's Disease Classification via Deep Convolutional Neural Networks using MRI and fMRI. *bioRxiv*. doi:10.1101/070441
- Sarraf, S., & Tofighi, G. (2016). DeepAD: Alzheimer's Disease Classification via Deep Convolutional Neural Networks using MRI and fMRI. *bioRxiv*. doi:10.1101/070441
- Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. arXiv 1409.1556 adresinden alındı
- Sorensen, L., Igel, C., Pai, A., Balas, I., Anker, C., Lillholm, M., & Nielsen, M. (2017). Differential diagnosis of mild cognitive impairment and Alzheimer's disease using structural MRI cortical thickness, hippocampal shape, hippocampal texture, and volumetry. *NeuroImage: Clinical*, 13, 470-482. doi:10.1016/j.nicl.2016.11.025
- Suk, H., & Shen, D. (2013). Deep Learning-Based Feature Representation for AD/MCI Classification. *Medical Image Comput. Comput. Assist. Interv.*, 16, 583-590.
- Swets, D., & Weng, J. (1995). Efficient content-based image retrieval using automatic feature selection. *IEEE International Symposium On Computer Vision*, 85-90.
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., . . . Rabinovich, A. (2015). Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition., (s. 1-9).

- Tensorflow. (2019). *Tensorflow*. Tensorflow: <https://www.tensorflow.org/> adresinden alındı
- Tensorflow-Github. (2019). *Github-tensorflow/tensorflow: An open source machine learning framework for everyone*. Tensorflow-Github: <https://github.com/tensorflow/tensorflow> adresinden alındı
- Unicamp. (2011). *Naive Bayes Classifier*. <https://www.ic.unicamp.br/~rocha/teaching/2011s2/mc906/aulas/naive-bayes-classifier.pdf> adresinden alındı
- Wikipedia. (2019). *Artificial Neural Network*. wikipedia.org: https://en.wikipedia.org/wiki/Artificial_neural_network adresinden alındı
- volBrain. (2019, May). *volBrain: Automated MRI Brain Volumetry System*. volBrain Users: <http://volbrain.upv.es/users.php> adresinden alındı
- volBrain. (tarih yok). *volBrain:Automated MRI Brain Volumetry System*. volBrain: <http://www.volbrain.upv.es> adresinden alındı
- Walia, R., Aggarwal, A., & Wadhwa, Y. (2013). Investigating EEG Characteristics of a Pregnant Woman. *International Journal of Emerging Research in Management & Technology*, 2(12), 14-19.
- Westman, E., Muehlboeck, J., & Simmons, A. (2012). Combining MRI and CSF measures for classification of Alzheimer's disease and prediction of mild cognitive impairment conversion. *NeuroImage*(55), 229-238. doi:10.1016/j.neuroimage.2012.04.056
- Wikipedia. (2019). *Alzheimer's Disease Neuroimaging Initiative*. en.wikipedia.org: https://en.wikipedia.org/wiki/Alzheimer%27s_Disease_Neuroimaging_Initiative#Enrollment_of_participants adresinden alındı
- Wikipedia. (2019). *C4.5 Algorithm - Wikipedia*. Wikipedia: https://en.wikipedia.org/wiki/C4.5_algorithm adresinden alındı
- Wikipedia. (2019). *Statistical parametric mapping*. Wikipedia: https://en.wikipedia.org/wiki/Statistical_parametric_mapping adresinden alındı
- Wikipedia. (2019). *Weka (machine learning)*. Wikipedia: [https://en.wikipedia.org/wiki/Weka_\(machine_learning\)](https://en.wikipedia.org/wiki/Weka_(machine_learning)) adresinden alındı
- World Health Organization. (2019, May). *Dementia*. www.who.int: <https://www.who.int/news-room/fact-sheets/detail/dementia> adresinden alındı

- Yosinski, J., Clune, J., Bengio, Y., & Lipson, H. (2014). How transferable are features in deep neural networks? *Advances in Neural Information Processing Systems* 27, 3320-3328. <https://arxiv.org/pdf/1411.1792.pdf> adresinden alındı
- Zhang, D., Wang, Y., Zhou, L., Hong, Y., & Dinggang, S. (2011). Multimodal classification of Alzheimer's disease and mild cognitive impairment. *NeuroImage*(55), 856-867. doi:10.1016/j.neuroimage.2011.01.008
- Zhao, Z., Morstatter, F., Sharma, S., Alelyani, S., Anand, A., & Liu, H. (2010). Advancing Feature Selection. *ASU feature selection repository*, 1-28.
- Zhou, Q., Goryawala, M., Cabrerizo, M., Wang, J., Barker, W., Loewenstein, D. A., . . . Adjouadi, M. (2014). An Optimal Decisional Space for the Classification of Alzheimer's Disease and Mild Cognitive Impairment. *IEEE Transactions on Biomedical Engineering*, 61(8), 2245-2253. doi:10.1109/TBME.2014.2310709

APPENDIX - A

Literature review for Data Mining

Authors	Dataset	Subject of Study	Used Modalities	Method	Result
Klöppel et al. (2008)	Three private datasets	AD-HC Classification	MRI	SVM	95%-93%-81% for the datasets, respectively.
Plant et al. (2010)	Private dataset	AD-HC Classification and Finding most related brain regions	MRI	SVM- Bayesian- Voting Feature Intervals	92% for Bayesian as best Prefrontal cortex, parietal lobe and hippocampal regions
Poulin et al. (2011)	ADNI and OASIS	Finding most related brain regions	MRI	Variance and Chi-Square Analysis	Amygdala atrophy and Hippocampal atrophy
Zhang et al. (2011)	ADNI	AD-HC classification	MRI + FDG-PET + CSF	SVM	86.2% (MRI) 90.6% (MRI + FDG-PET) 93.2% (All Modalities)
Hinrichs et al. (2011)	ADNI	AD-HC Classification	MRI + FDG-PET + CSF, Apolipoprotein E + Cognitive Scores	Multi Kernel Learning	87.6% (MRI+PET) 92.4% (All modalities)
Westman et al. (2012)	ADNI	MCI-HC and AD-HC Classification	MRI+CSF	OPLS (Orthogonal partial least squares)	<u>MCI-HC</u> 77.6% (MRI+CSF) 71.8% (MRI) and 70.3% (CSF) <u>AD-HC</u>

					91.8% (MRI+CSF) 87% (MRI) and 81.6% (CSF)
Aguilar et al. (2013)	ADNI	AD-HC Classification	MRI	DecisionTree, ANN, SVM and OPLS	81.9%, 84.9%, 83.6%, 84.5% for DT, ANN, SVM and OPLS, respectively.
Ota et al. (2014)	Diagnosis of Early Alzheimer's Disease - Japan (SEAD-J) dataset	Comparison of Brain atlases to see their affect on the AD-MCI classification performance.	MRI	SVM with Radial Bases Function	77.9% with LPBA40.
Zhou et al. (2014)	Private dataset	AD-HC Classification and Finding important region	MRI + MMSE	SVM	78.2% (MRI) 92.4% (MRI+MMSE) Right-left hippocampus and left amygdala
Sorensen et al. (2017)	ADNI, Imaging arm of the Australian Imaging Biomarkers and Lifestyle flagship study of ageing (AIBL) and CADDementia Challenge	AD-MCI-HC Classification and Finding important region	MRI	Linear Discriminant Analysis	62.7% for ADNI and AIBL 63% for CADDementia Hippocampal volume, ventricular volume and hippocampal texture

Behesti & Demirel (2016)	ADNI	AD-HC Classification	MRI	SVM (Fisher Criterion for feature selection)	96.3%
Behesti & Demirel (2017)	ADNI	AD-HC Classification	MRI	SVM (Genetic Algorithm for feature selection)	93%
Long et al.	ADNI	AD-HC Classification	MRI	SVM	96.5%
Peng et al. (2019)	ADNI	AD-HC Classification	MRI+PET+Single Nucleotide Polymorphisms (SNPs)	Regularized MKL	Dataset I 94.5% (All combinations) Dataset II 96.1% (All combinations)

Literature Review for Deep Learning

Authors	Dataset	Subject of Study	Used Modalities	Method	Result
Suk & Shen (2013)	ADNI	AD-HC Classification	MRI+PET+CSF	Stacked Auto Encoder and Multi Kernel - SVM	95.9%
Sarraf & Tofighi (2016)	ADNI	AD-HC Classification	fMRI	LeNet and GoogleNet	99% and 98.8%
Farooq et al. (2017)	ADNI	AD-MCI-Late MCI-HC Classification	MRI	GoogleNet, ResNet18 and ResNet152	98.8%, 98.01% and 98.14% respectively.
Gunawardena et al. (2017)	ADNI	AD-HC and AD-MCI-HC Classification.	MRI	SVM and Custom CNN architecture	95% sensitivity and 71.4% specificity for SVM 96% sensitivity and

					98% specificity for CNN
Luo et al. (2017)	ADNI	AD-HC Classification	MRI	Custom CNN architecture	69% sensitivity and 98% specificity
Cheng & Liu (2017)	ADNI	AD-HC Classification	MRI + PET	Custom CNN architecture	85.4% (MRI) 87.1% (PET) 89.6% (MRI+PET)
Islam & Zhang (2018)	OASIS	Non-demented / Very mild / mild / moderate Classification	MRI	Custom CNN architecture	0.97, 1.00, 0.67, 0.50, 0.94 as precision for Non-demented / Very mild / mild / moderate and average/total.
Lin et al. (2018)	ADNI	AD-MCI and AD-HC classification	MRI	Custom CNN architecture	79.90% for AD-MCI and 88.79% for AD-HC

APPENDIX – B

List and brief explanation of the features generated from volBrain System

Feature No.	Feature name	Description
1	Sex	Patient's sex
2	Age	Patient's age
3	Scale Factor	It is a value used in normalization and changes according to MR machine.
4	SNR	Signal-noise-ratio in mri
5	mSNR	
6	QC	Quality Control in mri
7	Tissue WM cm3	Tissue's cm^3 in white matter
8	Tissue WM %	Tissue's percentage in white matter
9	Tissue GM cm3	Tissue's cm^3 in gray matter
10	Tissue GM %	Tissue's percentage in gray matter
11	Tissue CSF cm3	Tissue's cm^3 in CSF
12	Tissue CSF %	Tissue's percentage in CSF
13	Tissue Brain cm3	Brain Tissue's cm^3 in overall
14	Tissue Brain %	Brain Tissue's percentage in overall
15	Tissue IC cm3	Tissue's cm^3 in the Intracranial Cavity (IC) which refers to the space within skull
16	Tissue IC %	Tissue's percentage in ICI
17	Cerebrum Total cm3	Total cm^3 of Cerebrum which is largest part of the brain and includes parts which are about learning, memory, language, communication.
18	Cerebrum Total %	Total percentage of Cerebrum
17	Cerebrum T GM cm3	Total cm^3 of Cerebrum in gray matter
20	Cerebrum T GM %	Total percentage of Cerebrum in gray matter
21	Cerebrum T WM cm3	Total cm^3 of Cerebrum in white matter
22	Cerebrum T WM %	Total percentage of Cerebrum in white matter
23	Cerebrum Right cm3	Total cm^3 of right of Cerebrum
24	Cerebrum Right %	Total percentage of Right of Cerebrum

25	Cerebrum R GM cm3	Total cm^3 of right of Cerebrum in gray matter
26	Cerebrum R GM %	Total percentage of right of Cerebrum in gray matter
27	Cerebrum R WM cm3	Total cm^3 of right of Cerebrum in white matter
28	Cerebrum R WM %	Total percentage of right of Cerebrum in white matter
28	Cerebrum Left cm3	Total cm^3 of left of Cerebrum
30	Cerebrum Left %	Total percentage of left of Cerebrum
31	Cerebrum L GM cm3	Total cm^3 of left of Cerebrum in gray matter
32	Cerebrum L GM %	Total percentage of left of Cerebrum in gray matter
33	Cerebrum L WM cm3	Total cm^3 of left of Cerebrum in white matter
34	Cerebrum L WM %	Total percentage of left of Cerebrum in white matter
35	Cerebrum Assymetry	Assymetry of Cerebrum between left and right lobes
36	Cerebellum Total cm3	Total cm^3 of Cerebellum which receives information from our sensory system and regulates motor movements.
37	Cerebellum Total %	Total percentage of Cerebellum
38	Cerebellum T GM cm3	Total cm^3 of Cerebellum in gray matter
39	Cerebellum T GM %	Total percentage of Cerebellum in gray matter
40	Cerebellum T WM cm3	Total cm^3 of Cerebellum in white matter
41	Cerebellum T WM %	Total percentage of Cerebellum in white matter
42	Cerebellum Right cm3	Total cm^3 of right of Cerebellum
43	Cerebellum Right %	Total percentage of Right of Cerebellum
44	Cerebellum R GM cm3	Total cm^3 of right of Cerebellum in gray matter
45	Cerebellum R GM %	Total percentage of right of Cerebellum in gray matter
46	Cerebellum R WM cm3	Total cm^3 of right of Cerebellum in white matter

47	Cerebellum R WM %	Total percentage of right of Cerebellum in white matter
48	Cerebellum Left cm3	Total cm^3 of left of Cerebellum
49	Cerebellum Left %	Total percentage of left of Cerebellum
50	Cerebellum L GM cm3	Total cm^3 of left of Cerebellum in gray matter
51	Cerebellum L GM %	Total percentage of left of Cerebellum in gray matter
52	Cerebellum L WM cm3	Total cm^3 of left of Cerebellum in white matter
53	Cerebellum L WM %	Total percentage of left of Cerebellum in white matter
54	Cerebellum Assymetry	Assymetry of Cerebellum between left and right lobes
55	Brainstem cm3	cm^3 of Brainstem which is posterior part of the brain.
56	Brainstem %	Percentage of Brainstem
57	Lateral ventricles Total cm3	Total cm^3 of Lateral ventricles which are two largest cavities of ventricular system and has CSF
58	Lateral ventricles Total %	Total percentage of Lateral ventricles
59	Lateral ventricles Right cm3	cm^3 of right of Lateral ventricles
60	Lateral ventricles Right %	Percentage of right of Lateral ventricles
61	Lateral ventricles Left cm3	cm^3 of left of Lateral ventricles
62	Lateral ventricles Left %	Percentage of left of Lateral ventricles
63	Lateral ventricles Asymmetry	Assymetry of Lateral ventricles between left and right lobes
64	Caudate Total cm3	Total cm^3 of Caudate
65	Caudate Total %	Total percentage of Caudate
66	Caudate Right cm3	cm^3 of right of Caudate
67	Caudate Right %	Percentage of right of Caudate
68	Caudate Left cm3	cm^3 of left of Caudate
69	Caudate Left %	Percentage of left of Caudate
70	Caudate Asymmetry	Assymetry of Caudate between left and right lobes

71	Putamen Total cm ³	Total <i>cm</i> ³ of Putamen which interconnected with many structures and work with many types of motor behaviors
72	Putamen Total %	Total percentage of Putamen
73	Putamen Right cm ³	<i>cm</i> ³ of right of Putamen
74	Putamen Right %	Percentage of right of Putamen
75	Putamen Left cm ³	<i>cm</i> ³ of left of Putamen
76	Putamen Left %	Percentage of left of Putamen
77	Putamen Asymmetry	Assymetry of Putamen between left and right lobes
78	Thalamus Total cm ³	Total <i>cm</i> ³ of Thalamus which interconnected with many structures and work with many types of motor behaviors
79	Thalamus Total %	Total percentage of Thalamus
80	Thalamus Right cm ³	<i>cm</i> ³ of right of Thalamus
81	Thalamus Right %	Percentage of right of Thalamus
82	Thalamus Left cm ³	<i>cm</i> ³ of left of Thalamus
83	Thalamus Left %	Percentage of left of Thalamus
84	Thalamus Asymmetry	Assymetry of Thalamus between left and right lobes
85	Globus Pallidus Total cm ³	Total <i>cm</i> ³ of Globus Pallidus which regulates movements.
86	Globus Pallidus Total %	Total percentage of Globus Pallidus
87	Globus Pallidus Right cm ³	<i>cm</i> ³ of right of Globus Pallidus
88	Globus Pallidus Right %	Percentage of right of Globus Pallidus
89	Globus Pallidus Left cm ³	<i>cm</i> ³ of left of Globus Pallidus
90	Globus Pallidus Left %	Percentage of left of Globus Pallidus
91	Globus Pallidus Asymmetry	Assymetry of Globus Pallidus between left and right lobes
92	Hippocampus Total cm ³	Total <i>cm</i> ³ of Hippocampus which responds with the functions of feeling and reacting.
93	Hippocampus Total %	Total percentage of Hippocampus
94	Hippocampus Right cm ³	<i>cm</i> ³ of right of Hippocampus
95	Hippocampus Right %	Percentage of right of Hippocampus
96	Hippocampus Left cm ³	<i>cm</i> ³ of left of Hippocampus

97	Hippocampus Left %	Percentage of left of Hippocampus
98	Hippocampus Asymmetry	Assymetry of Hippocampus between left and right lobes
99	Amygdala Total cm3	Total cm^3 of Amygdala which represents the fear system.
100	Amygdala Total %	Total percentage of Amygdala
101	Amygdala Right cm3	cm^3 of right of Amygdala
102	Amygdala Right %	Percentage of right of Amygdala
103	Amygdala Left cm3	cm^3 of left of Amygdala
104	Amygdala Left %	Percentage of left of Amygdala
105	Amygdala Asymmetry	Assymetry of Amygdala between left and right lobes
106	Accumbens Total cm3	Total cm^3 of Accumbens which has role in fear, aggression and addiction.
107	Accumbens Total %	Total percentage of Accumbens
108	Accumbens Right cm3	cm^3 of right of Accumbens
109	Accumbens Right %	Percentage of right of Accumbens
110	Accumbens Left cm3	cm^3 of left of Accumbens
111	Accumbens Left %	Percentage of left of Accumbens
112	Accumbens Asymmetry	Assymetry of Accumbens between left and right lobes
113	Class	Whether patient is AD or not.

VITA

Ümit KILIÇ was born in 1992, in Adana, Turkey. He received the BSc degree in Çukurova University, Adana, Turkey in 2016. He is currently MSc student in the Department of Nanotechnology and Engineering Science at the Adana Science and Technology University. He also works as Research Assistant in the Department of Computer Engineering at the same university. His research interests are Data Mining, Machine Learning, Artificial Intelligence, Deep Learning, and their applications.

