FLORIDA STATE UNIVERSITY

COLLEGE OF ART & SCIENCE

PATTERN RECOGNITION IN MEDICAL IMAGING:

SUPERVISED LEARNING OF FMRI AND MRI DATA

By

AMIRHESSAM TAHMASSEBI

A Dissertation submitted to the
Department of Scientific Computing
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

2018

Amirhessam Tahmassebi defended this dissertation on July 6, 2018.
The members of the supervisory committee were:

Anke Meyer-Baese
Professor Directing Dissertation

Simon Y. Foo
University Representative

Katja Pinker-Domenig
Committee Member

Peter Beerli
Committee Member

Dennis Slice
Committee Member

The Graduate School has verified and approved the above-named committee members, and certifies that the dissertation has been approved in accordance with university requirements.

To my love, Persia
To my parents, who always suspected I'd end up here

ACKNOWLEDGMENTS

Foremost, I would like to express my sincere gratitude to my advisor Dr. Anke Meyer-Baese for the continuous support of my PhD study and research, for her patience, motivation, enthusiasm, and immense knowledge. Her guidance helped me in all the time of research. I could not have imagined having a better advisor and mentor for my PhD study. She taught me a big lesson in that we should compete and still be honest people. Additionally, I would like to thank the rest of my doctoral committee: Dr. Katja Pinker-Domenig, Dr. Peter Beerli, Dr. Dennis Slice, and Dr. Simon Y. Foo for their encouragement, insightful comments, and hard questions. I would also want to thank Dr. Gordon Erlebacher, the chair of the Department of Scientific Computing for his valuable advice during my doctoral study. In addition to this, I would like to thank Eitan Lees for being such a honest and supportive friend since the first day at Florida State. My sincere thanks also goes to my dear friend, Dr. Amir H. Gandomi, for supporting me throughout my life abroad. He led me through my doctoral study and helped me to find my scientific career. I had the chance to publish numerous scientific papers with him and I am sure that it continues forever. Last but not the least, I would like to thank my parents for supporting me throughout my life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS & ABBREVIATIONS

The following short list of symbols and abbreviations are used throughout the document.

MRI	Magnetic Resonance Imaging
CAD	Computer Aided Diagnosis
NMR	Nuclear Magnetic Resonance
T_1	Longitudinal or Spin-Lattice Relation Time
T_2	Transverse or Spin-Spin Relation Time
$\hat{M}$	Magnetization Vector
$\hat{B}$	Magnetic Field
M_z	Longitudinal Magnetization Vector
M_{xy}	Transverse Magnetization Vector
fMRI	Functional Magnetic Resonance Imaging
BOLD	Blood Oxygen-Level Dependent
HRF	Hemodynamic Response Function
EEG	Electroencephalogram
MEG	Magnetoencephalography
DCE	Dynamic Contrast-Enhanced
DWI	Diffusion-Weighted Imaging
ADC	Apparent Diffusion Coefficient
TR	Repetition Time
TE	Echo Time
ROI	Region of Interest
NAC	N-acetylcysteine
MNI	Montreal Neuro-logical Institute
SPM	Statistical Parametric Mapping
FSL	FMRIB Software Library
FWHM	Gaussian Full Width Half Maximum
SNR	Signal-to-Noise Ratio
ICA	Independent Component Analysis
PCA	Principal Component Analysis
SVD	Singular Value Decomposition
MAE	Mean-Absolute-Error
MSE	Mean-Squared-Error
SVM	Support Vector Machine
LR	Logistic Regression
DT	Decision Tree
GP	Genetic Programming
P	Probability
RBF	Radial Basis Function

SLIQ	Supervised Learning In Ques
SPRINT	Scalable Parallelizable Induction of Decision Trees
IDE3	Iterative Dichotomizer 3
CART	Classification And Regression Tree
NB	Naive-Bayes
AdaBoost	Adaptive Boosting
XGBoost	eXtreme Gradient Boosting
GBM	Gradient Boosting Machine
GBRT	Gradient Boosted Regression Tree
HPC	High Performance Computing
LDA	Linear Discriminant Analysis
QDA	Quadratic Discriminant Analysis
RF	Random Forest
SGD	Stochastic Gradient Descent
LABC	Locally Advanced Breast Cancer
pCR	Pathological Complete Response
PPV	Positive Predictive Value
NPV	Negative Predictive Value
FOV	Field of View
TA	Acquisition Time
BI-RADS	Breast Imaging Reporting and Data System
NME	Non-Mass Enhancing
MTT	Mean Transit Time
RFE	Recursive Feature Elimination
ONF	Optimum Number of Feature
RCB	Residual Cancer Burden
RFS	Recurrence Free Survival
DSS	Disease-Specific Death
TN	Triple Negative

ABSTRACT

Machine learning algorithms along with magnetic resonance imaging (MRI) provides promising techniques to overcome the drawbacks of the current clinical screening techniques. In this study the resting-state functional magnetic resonance imaging (fMRI) to see the level of activity in a patient's brain and dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) to explore the level of improvement of neo-adjuvant chemotherapy in patients with locally advanced breast cancer were considered. As the first project, we considered fMRI of patients before and after they underwent a double-blind smoking cessation treatment. For the first time, this study aims at developing new theory-driven biomarkers by implementing and evaluating novel techniques from resting-state scans that can be used in relapse prediction in nicotine-dependent patients and future treatment efficacy. In this regards, two classes of patients have been studied, one took the drug N-acetylcysteine and the other took a placebo. Our goal was to classify the patients as treatment or non-treatment, based on their fMRI scans. The image slices of brain are used as the variable. We have applied different voxel selection schemes and data reduction algorithms on all images. Then, we compared several multivariate classifiers and deep learning algorithms and also investigated how the different data reductions affect classification performance. For the second part, we have employed multi-parametric magnetic resonance imaging (mpMRI) using different morphological and functional MRI parameters such as T_2-weighted, dynamic contrast-enhanced (DCE) MRI, and diffusion weighted imaging (DWI) has emerged as the method of choice for the early response assessments to NAC. Although, mpMRI is superior to conventional mammography for predicting treatment response, and evaluating residual disease, yet there is still room for improvement. In the past decade, the field of medical imaging analysis has grown exponentially, with an increased numbers of pattern recognition tools, and an increase in data sizes. These advances have heralded the field of radiomics. Radiomics allows the high-throughput extraction of the quantitative features that result in the conversion of images into mineable data, and the subsequent analysis of the data for an improved decision support with response monitoring during NAC being no exception. In this study. we determined the importance and ranking of the extracted parameters from mpMRI using T_2-weighted, DCE, and DWI for prediction of pCR and patient outcomes with respect to metastases and disease-specific death employing different machine learning algorithms.

CHAPTER 1

INTRODUCTION

Medical imaging has become one of the most essential visualization and interpretation techniques in biology, psychology, and medicine. In the past few years, numerous innovations were developed to help medical scientists in terms of improving the quality of the visualization to obtain better quantitative measurements to produce novel scientific hypotheses along with medical diagnoses. Medical imaging techniques, mostly noninvasive, play an important role in several disciplines such as medicine and psychology. The four main medical imaging signals are: (1) x-ray transmission, (2) gamma ray transmission, (3) ultrasound echoes, and (4) magnetic resonance induction [2, 4, 5, 6, 7].

One of the most important medical imaging approaches is the analysis of the images made by magnetic resonance imaging (MRI). This has recently shown that it could be one of the bedrocks of medical imaging which would play a vital role in the improvement of the biomedical imaging and its interpretations. In addition to this, pattern recognition in the medical imaging demands novel techniques as well. The patterns might appear to one as complex and totally black-box (like the brain). Thus, extracting enough features from the MRI and also functional MRI (fMRI) data requires mathematical algorithms with the help of computers to increase the speed and accuracy and is called computer aided diagnosis (CAD).

1.1 Magnetic Resonance Imaging

Magnetic resonance imaging represents a non-invasive imaging method used to render images of the inside of the body. During the past 30 years, it became one of the key bio-imaging modalities in medicine [2]. Nuclear magnetic resonance (NMR) is the basis of the physical principles behind the MRI signals. In fact, the change of the nuclear magnetism due to the hydrogen atoms in the fat and water of the human body would make this signal. The contrast of this image would totally depend on longitudinal or spin-lattice relation time (T_1) and transverse or spin-spin relation time (T_2). T_2 is tissue-dependent which would produce contrast in MR images. If one were to take into

Figure 1.1: Philips Intera Achieva 3T MRI Scanner.

account the difference in the observed intensities of different tissues, T_1-wighted and T_2weighted images can be constructed. The details are presented in the following [2].

In MR imaging, a specific spin system (hydrogen atoms) within a small volume of tissue or a voxel can represent the macroscopic magnetization. The spin system can be magnetized by the presence of the magnetic field $\vec{B_0}$. This can be modeled by a bulk magnetization vector $\vec{M}$ which has an equilibrium value $\vec{M_0}$ with the same direction as $\vec{B_0}$. The bulk magnetization vector $\vec{M}$ depends on the 3-dimensional spatial coordinates also the time of recording these coordinates ($\vec{M} = \vec{M}(\vec{r}, t)$). The value of an MRI image at a given voxel is characterized by two essential factors: (1) the tissue properties (T_1 and T_2 relaxation parameters and proton density which is basically the number of targeted nuclei per unit volume) and the scanner imaging protocol. As shown in Figure 1.2 the magnetization vector $\vec{M}$ has two components: (1) the longitudinal magnetization vector ($M_z(t)$), (2) the transverse magnetization vector ($M_{xy}(t)$). The transverse magnetization vector is a complex quantity which combine two orthogonal components as shown in the following implementation [1]:

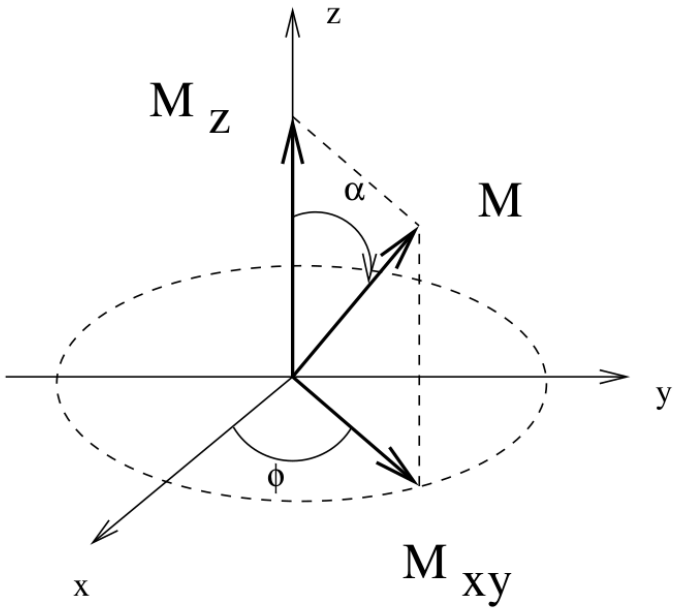


Figure 1.2: The magnetization vector M precesses about the z-axis [1].

$$M_{xy}(t) = M_x(t) + iM_y(t) \qquad \phi = \arctan(\frac{M_x}{M_y}) \tag{1.1}$$

RF signals can inject perturbation to the spin system. The excitation will push the bulk magnetization $\vec{M}(t)$ at an angle α towards the xy-plane as shown in Figure 1.2. The equilibrium state process of the magnetization vector $M_z(t)$ is explained in the following implementation:

$$M_z(t) = M_0[1 - \exp(-\frac{t}{T_1})] \tag{1.2}$$

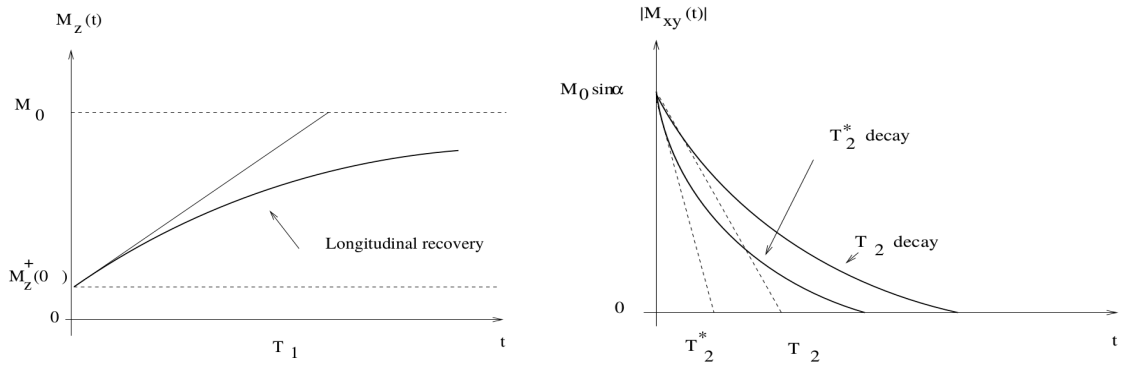
Similarly, the equilibrium state process of the magnetization vector $M_{xy}(t)$ is explained in the following implementation:

$$M_{xy}(t) = M_{x_0y_0}[1 - \exp(-\frac{t}{T_2})] \tag{1.3}$$

Figure 1.3 illustrates the decay associated with the external fields. relationship between the three transverse relaxation components is implemented as following:

$$\frac{1}{T_2^*} = \frac{1}{T_2} + \frac{1}{T_2'} \tag{1.4}$$

where $T_2^* < T_2$ due to the local perturbations in the static field B_0.



(a) Transverse (b) Longitudinal

Figure 1.3: Transverse and longitudinal relaxation [1].

Functional magnetic resonance imaging represents a novel non-invasive technique for the study of cognitive functions of the brain. In fMRI, the changes in blood flow in the brain would be detected. Specifically, when a particular part of the brain is more active, the blood flow in that region would be increased. This would cause more oxygen in that particular region to bring nutrients to the hard-working cells. fMRI would track the variations of the blood flow to detect the active part of the brain. An MRI machine as shown in Figure 1.1 contains a giant magnet. Magnetic fields of the nuclei in oxygen-rich blood are flipped due to the combination of a strong magnetic field and radio waves. This produces a detailed map of the regions where the ratio of flow of oxygen-rich blood to the brain is high which explains the high activity areas of the brain and known as BOLD signal. The BOLD signal is generally modeled as the convolution of the stimulus function with Hemodynamic Response Function (HRF) [8, 9, 10, 11, 12, 13].

The energy due to an influx of oxygenated blood to a local area of neuronal activity produces the BOLD signal. Oxygenated hemoglobin has a diamagnetic effect. However, hemoglobin would show paramagnetic characteristics once it is deoxygenated. The MRI machine produces a magnetic field which aligns the randomly oriented atomic nuclei within the direction of the magnetic field

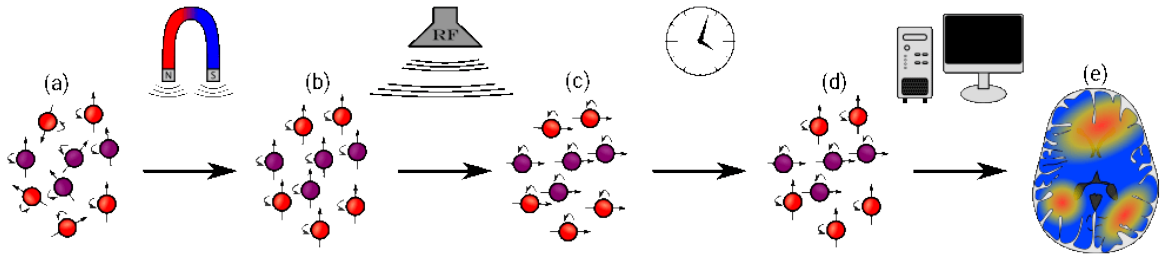


Figure 1.4: Schematic work structure of fMRI. (a) stochastic movements of atoms in the brain, (b) alignment of the atoms' spins due to the magnetic field, (c) spins got knocked by radio frequency pulse, (d) recovery of the spins to stage (b) and recording this transition time based on the brain tissue, (e) mapping the neural activity based on the recorded transion time using computer.[1]

[14, 15]. Hemodynamic responses include two main effects which are spatial and are handled by vasculature and temporal that project delays caused by the neural activities [2]. Subsequently, fMRI advantages include (1) recording the brain signals noninvasively and with zero danger in terms of radiation, (2) perfect resoultions for both spatial and temporal scans, (3) and contains the ability of being combined with the other techniques such as electroencephalogram (EEG), and magnetoencephalography (MEG) to study the brain [2]. Figure 1.4 illustrates the details of how fMRI works.

In MRI of the breast, detection, differentiation and characterization of lesions is facilitated by the intra-venous application of contrast agent. Dynamic contrast-enhanced (DCE) MRI allows the simultaneous assessment of lesion morphology and enhancement kinetics and it has been demonstrated that combining this morphologic and functional information is essential for an accurate diagnosis [16, 17]. To reduce false-positive findings, additional functional imaging approaches, such as diffusion-weighted imaging (DWI), were developed and successfully introduced into the clinical routine [18]. DWI is an MRI parameter that provides information about the local diffusivity of water in body tissue, which is typically restricted in malignancies. This method is based on molecular diffusion, or Brownian motion, which is the random motion of water molecules as a result of agitation by thermal energy. In an isotropic medium, water molecules tend to move in all directions equally and the signal attenuation of an MRI voxel can be measured using the diffusion coefficient.

[1] http://sitn.hms.harvard.edu

Figure 1.5: Four-channel double-tuned $^{31}P/^{1}H$ breast coil (Stark Contrast, MRI Coils Research, Erlangen, Germany).

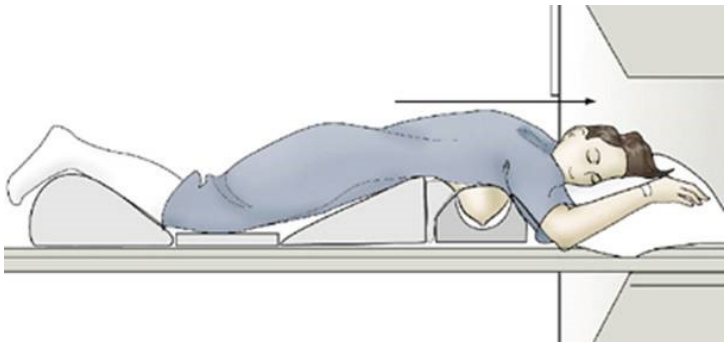


Figure 1.6: Illustration of patient positioning for MRI of the breast.[2]

However, as in breast tissue cell membranes and other physiological barriers restrict water diffusion, DWI is quantified using the apparent diffusion coefficient (ADC), assuming that diffusion in body

tissue is free. For the evaluation of water diffusivity, ADC values are displayed in parametric maps illustrating the varying degrees of diffusion in In malignant lesions, diffusivity is even more hindered due to higher cell density and micro-structural changes, resulting in lower ADC values [19].

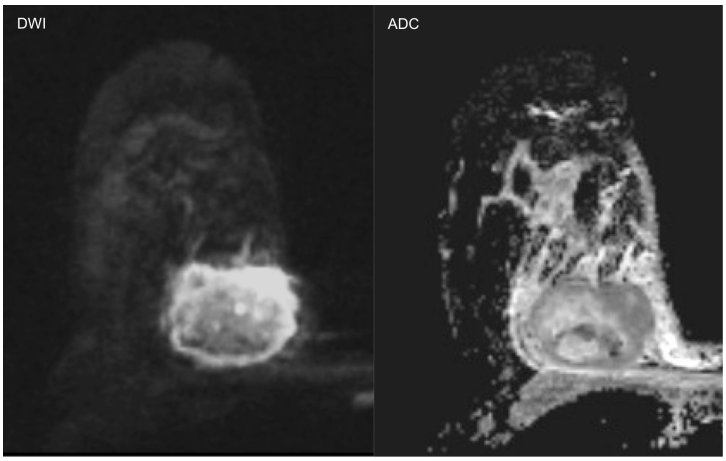


Figure 1.7: DWI and ADC map of a meta-plastic breast cancer.

T_2-weighted MRI (T_2W MRI) is based on a long repetition time (TR) setting, decreasing the effect of T_1 signal in nuclear magnetic resonance (NMR), signal measured, and enhancement the T_2 effect of tissues by increasing the values of the echo time (TE). Based on the characteristics of transverse relaxation, T_2 values can be computed as the following implementation:

$$M_{xy}(t) = M_{xy}(0)\exp(-\frac{t}{T_2}) \tag{1.5}$$

where $M_{xy}(0)$ is the initial value of $M_{xy}(t)$ and T2 is the relaxation time [20].

The most commons steps one must follow in terms of analysis and interpretation of medical images are presented in Figure 1.8. After image formation, increase the visualization precision and resolution the image should preprocessed which includes image registration and transformation. Additionally, segmentations and filtering would be employed to deblur the images and increase the visibility of the edges of the object. Thus, shape modeling and feature extraction based on region or voxel (pixel) would be the next stage. The last step would be prediction and classification using

[2]http://mrprotocols.com/oldsite/MRI/Chest/breast_mass_mri.htm

machine learning algorithms. The predictions includes texture characterization, decision making, and a separation of normal and abnormal tissues that ultimately lead to a CAD pipeline [21, 22, 23].

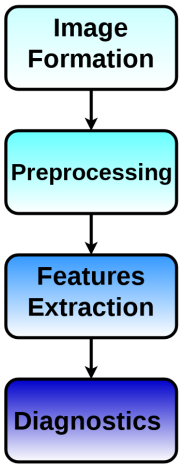


Figure 1.8: Model for diagnostic system using medial images [2].

A typical CAD pipeline contains three layers: (1) a data layer which includes the database in terms of storage and distribution, (2) an application layer like a web server to manage the users and provide them with visualizations, and (3) a presentation layer for the users to have access to graphics remotely. In this study, the application layer which contains the CAD workstations will be considered in depth. The CAD workstation contains different stages such as image preprocessing, definition of region of interests (ROIs), feature extraction and dimensionality reduction, and classification of ROIs. In addition to this, two softwares for both regression and classification tasks will be proposed as future works [24, 25, 26].

In chapters 2 and 3, the use of machine learning in medical imaging will be presented in two different case studies: 1) fMRI data for brain, 2) and MRI data for breast. For each case study, background and previous works will be explained as an introduction. As the current research, the data acquisition and data preprocessing will be presented as well as the data reduction methods along with different machine learning algorithms will be presented. Additionally, the proposed research will be presented as well. Chapter 4 includes the summary of the current research along with the proposed goals for the future works. The manuscript ends with a bibliography.

CHAPTER 2

BRAIN: FMRI SMOKING CESSATION CLASSIFICATION

2.1 Background & Previous Works

Smoking cigarettes leads to illnesses such as heart disease, strokes and cancer. Smoking is the leading cause of preventable mortality in the United States with around 50% of lifelong smokers dying from one of the illnesses mentioned earlier [27]. What drives people to continue smoking cigarettes is the nicotine dependency that smoking causes them to have. The nicotine released by the tobacco increases the neurotransmitters Dopamine in the brain. Dopamine plays an important role in the addiction center of the brain (mesolimbic system) and it controls better than demands pleasure, reward, and addiction. Insomnia, tremors and quivering, lightheadedness, high blood pressure, heart attack, and decreasing in bone density are just a few symptoms that nicotine could cause. This dependency drives them to compulsively have to smoke in order to keep the withdrawal effects associated with smoking cessation away. Mesolimbic dopamine reward has traditionally been the region of interest for the neuro-biological research in the area of drug addiction [28]. In the recent literature published in this area, it has shown that glutamate played an important role for cocaine-dependent subjects in terms of continuation of cocaine use or even relapse after quiting [3]. Figure 2.1 presents a schematic view of brain circuitry involved in learning, memory, and addiction. It clearly illustrates the firing path of glutamatergic projections from pre-frontal cortex to nucleus accumbens. Developing a cessation treatment that will reduce a patient's dependency on nicotine as well as reduce the effects of withdrawal could help millions of people quit a dangerous habit. As one of the potential glutamatergic substances, N-acetylcysteine (NAC) can be used [28]. NAC ($C_5H_9NO_3S$, mw:163.19) is a derivative of the amino acid cysteine prodrug which is approved as a mucolytic agent and an acetaminophen antidote. NAC restores the basal level of glutamate in the accumbens which may reduce the drug seeking behavior [29]. Previously, Schmaal et al. have shown that NAC appears to be a new potential treatment for nicotine dependence [28]. NAC has been approved by FDA in the United States and use to be treated for acetaminophen overdose

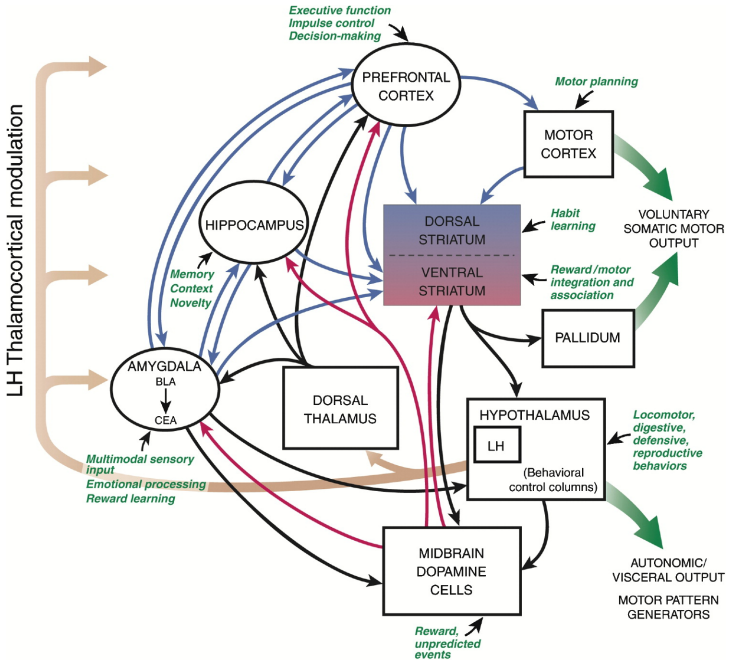


Figure 2.1: Schematic view of brain circuitry involved in learning, memory, and addiction. The essential neurotransmitter is the glutamate which its pathways are shown in blue, dopamine pathways in red, and bright tan lines illustrates the direct and indirect projections from hypothalamus to neocortex and fore-brain limbic structures [3].

and sold over-the-counter in the United States. It has been shown that NAC restores the basal level of glutamate in the accumbens [29] and would reduce the drug seeking behavior. Previous studies have investigated the effect of NAC treatment in cocaine-dependence which showed that nicotine-dependent behavior is related to brain networks [29]. In addition to this, the relapse in cocaine-dependent rats were studied in 2003 [30]. In 2007, Schubert et al. found no abnormal pattern detected in the Glutamate concentrations level for chronic tobacco smokers employing single voxel proton magnetic resonance spectroscopy at 3 Tesla [31]. Although the changes in the Glu levels due to the NAC induction have not been proven yet, NAC has shown different benefits in several parts including pathological gambling [32], number of cigarettes smoked [33], and marijuana users [34].

Dynamic functional connectivity of the brain changes over time [35]. Moreover, in 2015 it was found that genetic variants influence human brain structures, especially subcortical brain regions

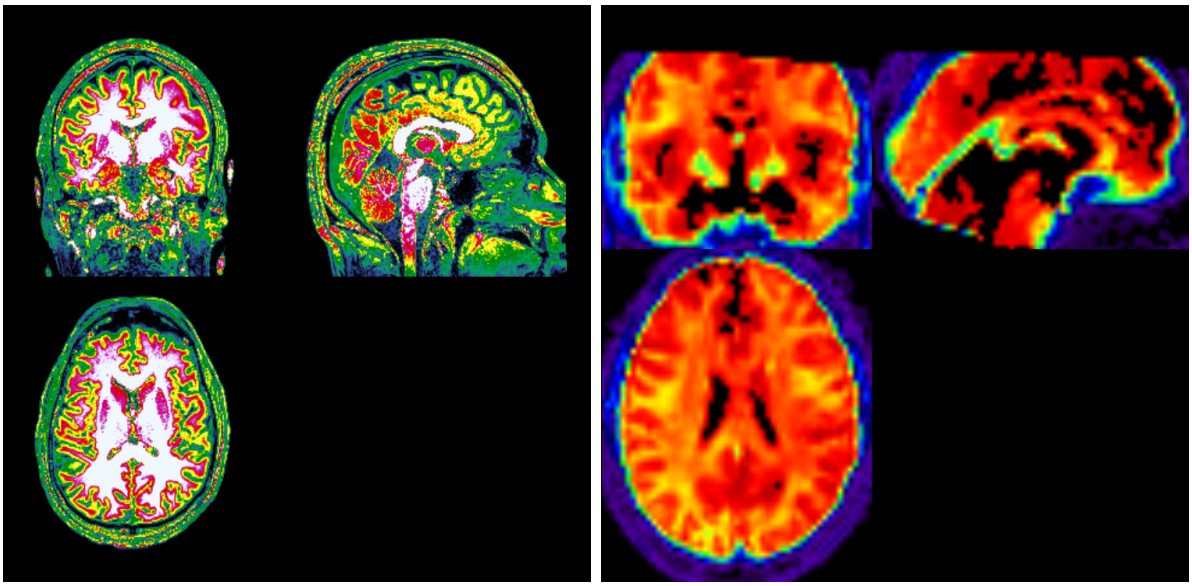
which coordinate movement, learning, memory and motivation [36, 37, 38]. Many activities such as thinking, learning, and quitting a habit would cause these changes. For instance, the association of different spatial patterns of neural with thinking about different semantic categories of pictures and words have been shown by brain imaging studies. In this regard, Mitchell et. al. [39] presented a computational model to link fMRI activities and thinking about arbitrary words. Employing machine learning algorithms and statistical inference methods are commonly used to implement computational models to find relations between fMRI data and related tasks [39, 40, 41, 42, 43, 44, 45]. Finding the relationships among brain connections would be possible by using functional magnetic resonance imaging (fMRI) which would help to scan a patient's brain in resting state with a minimized amount of error and noise. Previously, Smith et. al. have investigated the functional connectivity of nicotine-dependent patients using the Broadmann area in the brain [46]. However, they could not report a classification accuracy better than 50% using linear support vector machines. This would suggest to employ new methods for extracting region of interests and non-linear multivariate machine learning algorithms for classifications.

In this study, as a novel approach to choose region on interests, three different voxel selection schemes (masks) with three essential dimensionality reduction methods have been applied to extract features to analyze data from a smoking cessation treatment, where subjects take a drug to reduce their nicotine dependence while still being allowed to smoke in order to keep off the effects of withdrawal. This is the preferred method as more people are likely to try it if they do not have to quit smoking immediately. The goal is to reduce the nicotine dependency to the point that it is easier for the subject to stop. Relapse is more probable in smokers than aiming to quit smoking [47]. The purpose of this paper is to prove that there is a difference in the resting-state [48] functional magnetic resonance imaging (fMRI) images of a smoker that undergoes this smoking cessation treatment compared to a smoker that receives a placebo.

Here is the outline of this chapter: first the details of data acquisition and data preprocessing will be explained. Consequently, several feature extraction methods, various machine learning and deep learning algorithms will be discussed. Finally, the results and discussion of the model validation and relapse prediction in subjects will be presented.

2.2 Data Acquisition

The main goal of this study was to determine whether or not the drug N-acetylcysteine (NAC) would decrease nicotine dependency. NAC may have an effect on relapse in smoking cessation [47, 49]. In this regard, 39 regular smokers participated this treatment study at the Spinoza Center[2] of the University of Amsterdam. 19 heavy smokers who wanted to quit, took the drug NAC (class 1) and the other 20 subjects took a placebo (class 0) for two weeks. Anatomical and functional scans of their brains were taken at baseline, and after two weeks of NAC treatment. Then, the relapse data were assessed at six months past NAC treatment.



(a) Anatomical (b) Functional

Figure 2.2: Anatomical and functional slices of the brain.

The Spinoza Center of University of Amsterdam is equipped with a 3.0 T Intera MRI scanner (Philips Health care, Best, The Netherlands) with a 32-channel SENSE head coil to obtain MRI data (Figure 1.1). The subjects were asked to keep their eyes closed, stay relaxed, and stay awake during the scan (resting state). Two hundred 3-dimensional functional images of the subjects' brains of size $80 \times 80 \times 37$ with a voxel size of 3 mm^3 with 2.3 seconds as repetition time were

[2]https://spinozacentre.nl/

taken due to the sensitivity to blood oxygen level-dependent (BOLD) contrast by the gradient-echo planar sequence. In addition to this, the 3-dimensional anatomical data of size $240 \times 240 \times 220$ with a voxel size of 1 mm^3 have been acquired. Figure 2.2 shows slices of the brain from one patient in all three axes for both the anatomical (Figure 2.2a) and the functional (Figure 2.2b) representations.

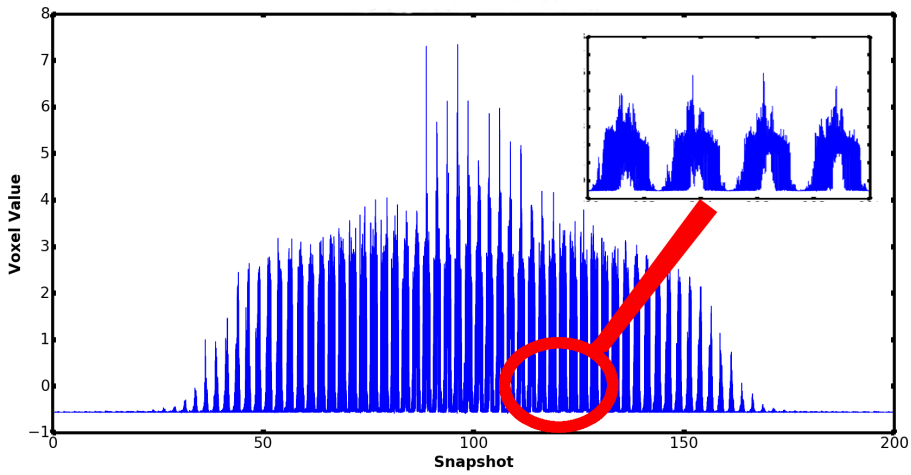


Figure 2.3: BOLD signal during the time domain series.

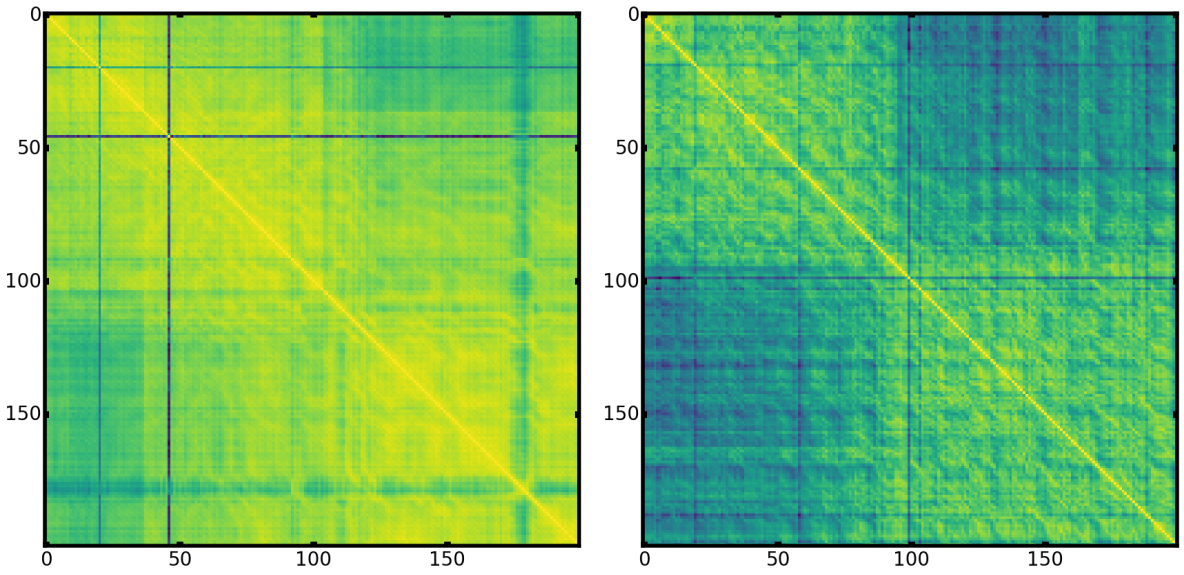
2.3 Data Preprocessing

The fMRI data was given in NIFTI (Neuroimaging Informatics Technology Initiative) formats which contains spatio-temporal slices. Due to the long process of the scans, possible movements of the subject, and physiological noise, this resulted in subject-dependent artifacts in the data [50].

2.3.1 Initial Trial

In the initial trial, imaging data were analyzed using Statistical Parametric Mapping (SPM12). Functional images of each subject were realigned and unwarped, co-registered with the structural MRI image, and segmented for normalization to a Montreal Neuro-logical Institute (MNI) template. Finally, images were smoothed using a 4 mm full-width at half maximum Gaussian kernel. The registered functional MRI volumes with the Montreal Neurological Institute template were divided into 116 regions according to the automated anatomical labeling atlas [51, 52]. The atlas divides the

cerebrum into 90 regions (45 in each hemisphere) and divides the cerebellum into 26 regions (nine in each cerebellar hemisphere and eight in the vermis). Regional mean time series were obtained for each subject by averaging the functional MRI time series over all voxels in each of the 116 regions. The residuals of the regression represent the set of regional mean time series that will be used for functional connectivity analyses. Figure 2.4 presents the raw and preprocessed data after the initial preprocessing stage.



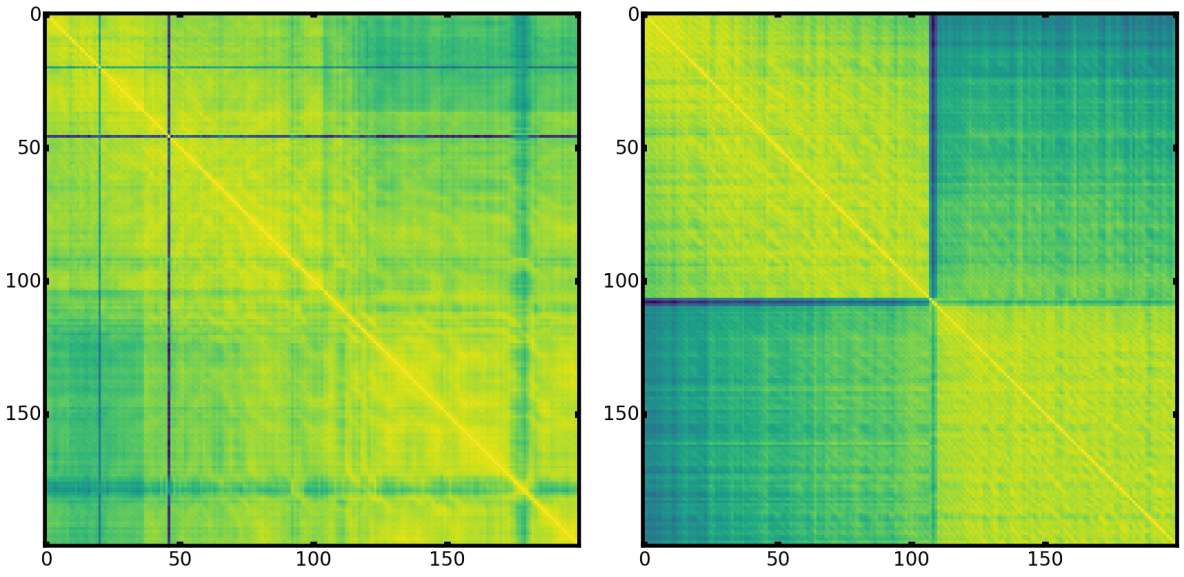
(a) Raw (b) Preprocessed

Figure 2.4: Raw and initial preprocessed data.

2.3.2 Final Trial

In the final trial, the fMRI data was analyzed using a pipeline, a combination of Statistical Parametric Mapping (SPM12) and FMRIB Software Library (FSL). The scientific goal of the preprocessing phase is to increase the BOLD contrast to noise. It began with motion correction, since the movement of the subject for 1% of the voxel size would make 1% change in the signal. This change can be greater than the BOLD signal that is going to be extracted as a feature. It might also cause a loss in correspondence between a voxel and anatomical location. The final voxel would not be as the same previous voxel. This is important due to the sensitivity of the statistical

analysis of the residual noise in the image series. Then the images underwent segmentation and realignment. Correction only for in-plane rotations and translation of the head within the image were applied. The first image was selected as the reference image. The other images of that slice were aligned with the reference image.



(a) Raw (b) Preprocessed

Figure 2.5: Raw and final preprocessed data.

The approach, including two dimensional rotations and translations were applied to the second image. As a stopping criteria, further translations were continued and rotation for realignment of the subsequent images until the sum of the square difference between voxels is minimized [53, 54]. This approach was quite fast and there were not any problems related to convergence. In addition to this, scanners might acquire slices in an interleaved fashion to avoid interfering with neighboring ones. Thus, temporal slice timing correction is needed as well as shifting the slices back in order. For example, the middle of the sequence is not necessarily the middle of the brain. Smoothing the temporal domain might also increase the signal-to-noise ratio. In this regard, general linear model along with a Gaussian to approximate the haemodynamic response function and smooth the

voxel time course was employed. All the types of filtering of temporal parts could be carried out as Fourier domain [55].

Regarding spatial normalization and spatial smoothing a Gaussian Full Width Half Maximum (FWHM) kernel was employed. A kernel of 3 mm^3 was chosen for each voxel to be replaced by the weighted average of its neighbors. High pass or low pass filters based on the frequencies would be used for temporal filtering. Improvement of the signal-to-noise ratio (SNR) using reduction in random noise is desired to detect true activations using the statistical techniques [56].

Finally, to map functional and anatomical scans into a brain template one must start analyzing the slices. Brain templates can be regarded as a subclass of brain atlases and are usually used as a references for mapping different brains of subjects in a group analysis study. The choice of template would change the statistical results based on machine learning algorithm. Thus, the data should be aligned with the template as closely as possible [57]. The images are spatially normalized to the Talairach standard coordinates [58, 59]. In addition to this, the Montreal Neurological Institute (MNI) [60] brain template was applied to the data. To remove linear trends in each session of 200 images, the function data were band-pass filtered and de-trended. Figure 2.5 presents the raw and preprocessed data after the final preprocessing stage.

2.4 Feature Extraction

As discussed, the preprocessed data include about $94,720,000$ features for each subject. To feed the images into machine learning classifiers the features matrix must be made. Dealing with big data led to employing new approaches to extract essential features for the classification tasks. In this regard, three different masks [61] were made. A mask is a 3-dimensional array of 0s and 1s, where a 1 signifies to keep the voxel in that position, and 0 indicates to ignore it in the data. The first mask is related to the parts of the brain that have high activity as shown in Figure 2.6. The highest voxel values that would account for 60% of the average values of the voxels for each subject were kept. By applying high activity mask, the feature matrix had a size around 94,720 features.

In the second mask, the limbic system, where the addiction occurs was studied [62]. The word addiction is derived from a Latin term for enslaved by or bound to. The process of addiction in brain is quite interesting. It starts by changing the brain in terms of pleasure registrations and

[3] http://learn.genetics.utah.edu

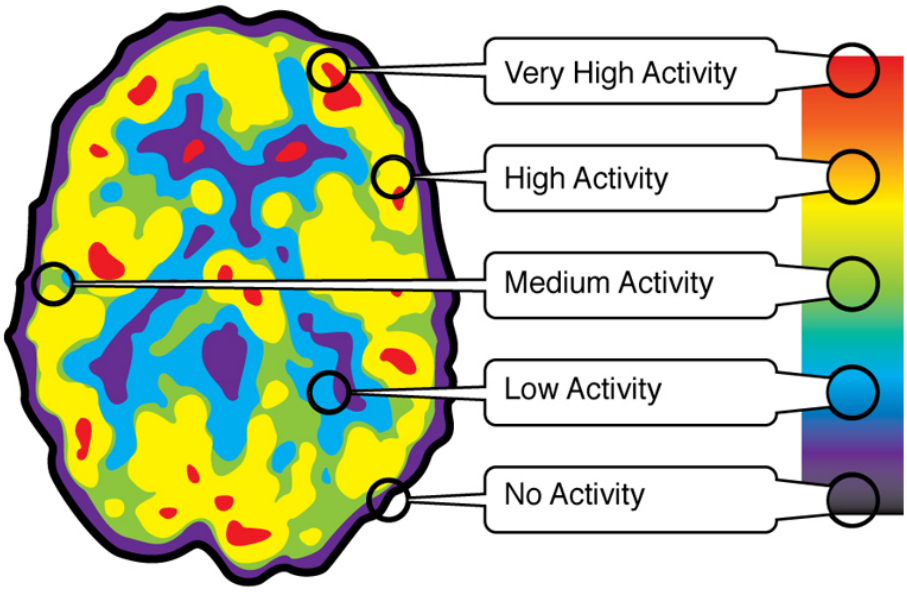


Figure 2.6: Activity color map of a brain.[3]

then changing the normal patterns in this part such as learning and motivations. It was proven that addiction and pleasure are correlated. In addition to alcohol and powerful drugs, the other pleasurable activities such as shopping, sex, and gambling could cause an addiction and corrupt the registered patterns in the brain [28].

The way that the brain treats all different kinds of pleasures such as sex, gamble, drugs, and even an attractive meal is the same. In fact, in the nucleus accumbens of the brain there are nerve cells lying underneath the cerebral cortex which produce the neurotransmitter dopamine. Dopamine is the distinct signature of pleasure. This is also called the region of the brain's pleasure center.

Like sex, and gambling, drugs including nicotine or heroin, could cause an increase in producing dopamine in the nucleus accumbens of the brain. Using neuroimaging, it was proven that the speed of dopamine release in the brain is correlated with the likelihood of using drugs or any rewarding activity which might cause an addiction. In this regard, different ways of drug usage would change this pattern. For instance, smoking a drug, eating the drug as a pill, or injecting the drug in the veins would cause different speed of dopamine release in the body. This would change the likelihood

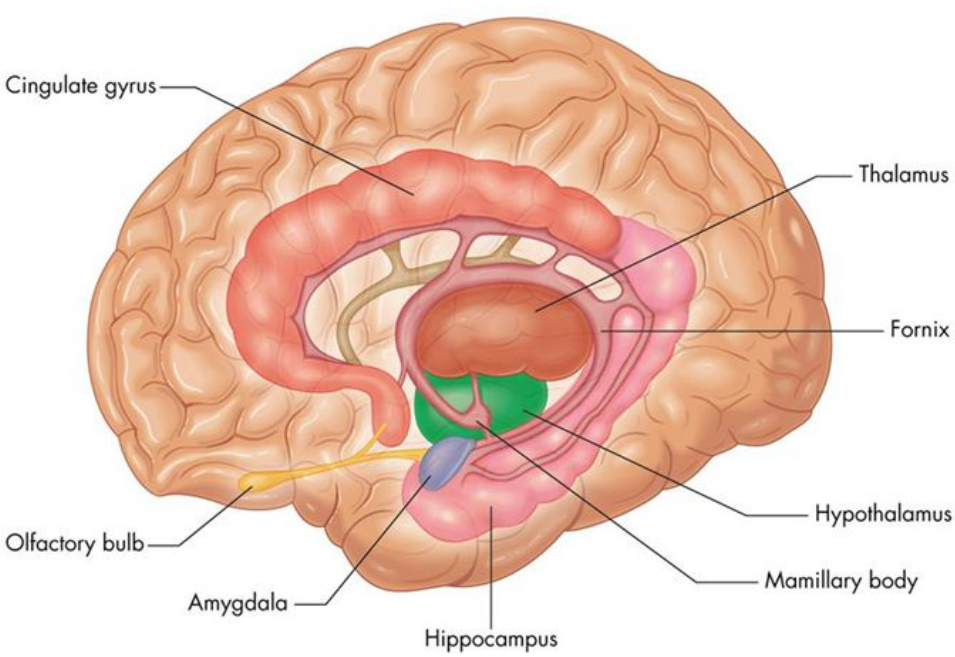


Figure 2.7: The limbic system of a brain.

of the addiction in the subject.

Figure 2.7 presents the different parts of the limbic system of a brain. Hippocampus plays an important role in learning process, memory, and emotions. Amygdala is related to the tasks such as eating, drinking, and sexual behaviors and hypothalamus monitors the blood level in terms of glucose and salt. In addition to this, it also controls the level of hormones in the blood and its pressure as well. It was previously shown using neuroimaging evidence that the frontal cortex of the brain involves drug addiction [63]. Essentially, the orbitofrontal cortex and the anterior cingulate gyrus are the regions that neuro-anatomically connected with limbic system [63]. It was shown that these parts of the limbic system are activated in the subjects who were addicted to drugs and deactivated in the subjects who quit.

Observing the average 3-dimensional image obtained earlier, a 3-dimensional rectangular box was constructed around where it was believed the average patient's limbic system was. By applying the limbic system mask, the feature matrix had a size around $25,200$ features. Next, only voxels

that had high activity inside the limbic system were observed. This was done by combining the first and the second mask. By applying this mask, the feature matrix had a size of $10,080$ features.

2.5 Data Reduction

The research problem consisted of a big data problem and was very computationally intensive. By applying masks to the pre-processed data, the feature matrix would be $39 \times 94,720,000$ which is a huge number for a feature matrix. As was stated previously, there were 200 temporal snapshots before the treatment and 200 after the treatment. Hence, the average of temporal parts was used. By applying the high activity limbic system, and high-limbic mask on the pre-processed data, to extract the features, the size of the final matrix would be $39 \times 94,720$, $39 \times 25,200$, and $39 \times 10,080$. The final matrix size was still high to be used as a feature matrix for classification and six different algorithms were employed to reduce the data and find the feature matrix to feed the classifier.

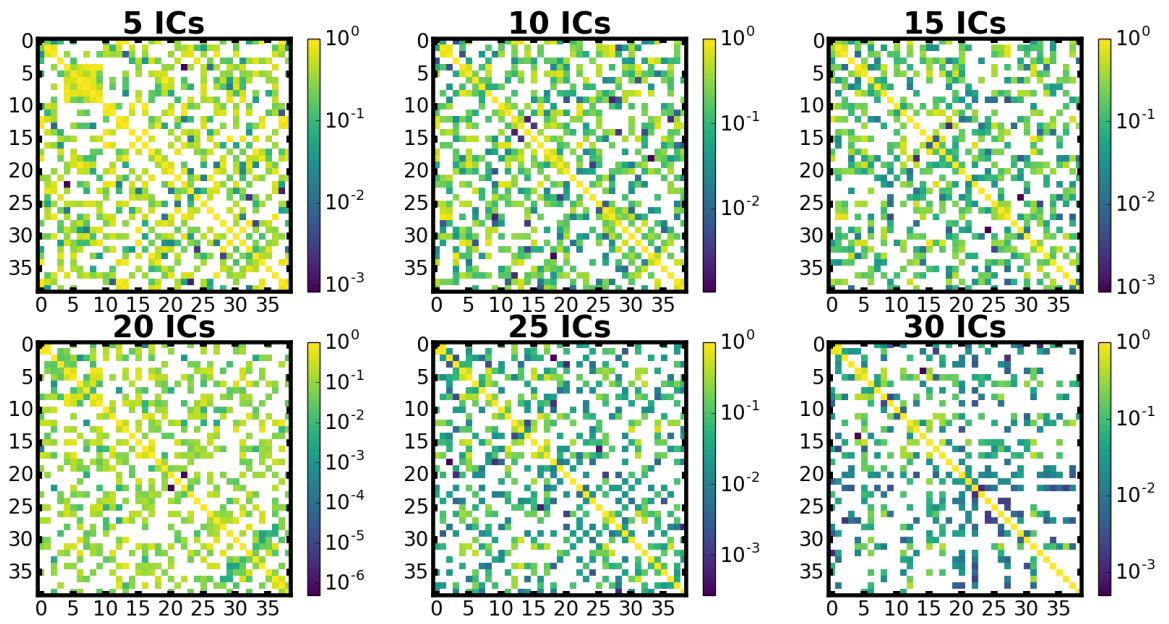


Figure 2.8: Correlation matrix for different numbers of independent components.

Algorithm 1: FastICA

Input: Number of desired components q, pre-whitened matrix $[A]_{N\times M}$

Output: Un-mixing weights $[W]_{M\times q}$, independent component matrix $[S]_{N\times q}$

$g \leftarrow$ the measure of non-Gaussianity of the projection $W^T A$;

$g' \leftarrow$ the first derivative of g;

for $i \in \{1, \ldots, q\}$ **do**

  $w_i \leftarrow$ Random vector of length N;

  while w_i *changes* **do**

    $w_i \leftarrow \frac{1}{M} A g(w_i^T A)^T - \frac{1}{M} g'(w_i^T A) I w_i$;

    $w_i \leftarrow w_i \sum_{j=1}^{i-1} w_i^T w_j w_j$;

    $w_i \leftarrow \frac{w_i}{||w_i||}$;

  end

end

$W \leftarrow \{w_1 \ldots w_q\}$;

$S \leftarrow W^T A$;

2.5.1 Independent Component Analysis (ICA)

Independent Component Analysis (ICA) [64] is a technique used to separate a multivariate signal into multiple independent non-Gaussian signals. It should be noted that ICA has been used to extract the hidden spatio-temporal structure in neuroimaging. ICA has the assumption that these underlying signals are maximally independent of each other. It uses the fact that two random variables would be uncorrelated if they are independent [65]. ICA can explain that all processes in the brain can be associated with a single time component in a voxel [64]. There are some algorithms such as infomax, JADE, and FastICA to employ for this approach [66]. In this paper, the FastICA, and topographic ICA (TopoICA) approach have been used. The FastICA is a hierarchical and symmetric approach by minimizing the mutual information. It employs non-Gaussianity by measurement of negentropy [64]. Typically, an ICA model tries to extract a feature matrix U from the full rank matrix A. Assuming there are N patients and M features for each, then the matrix size would be $[A]_{N\times M}$. When trying to find a good approximate with respect to ones sources, call W which is an unmixing matrix providing a linear decomposition of A. Thus, for extracting q features out of M features, one will have:

$$[U]_{N\times q} = [A]_{N\times M}[W]_{M\times q} \tag{2.1}$$

Now, one can feed vector U into the classifiers with a different number of features q. Here, $[5, 10, 15, 20, 25, 30, 35]$ independent components have been extracted. ICA will tell where the regions of the brain are that share similar brain activity. ICA is also limited by the number of subjects. Figure 2.8 illustrates the correlation matrix of different numbers of independent components with each other.

2.5.2 Principal Component Analysis (PCA)

The basic idea of PCA is to reach low redundancy and high information density with finding such transformed features of the original input. PCA is also referred to as Karhunen-Loeve transformation or the Hotelling transform [1]. PCA orthogonally transforms data consisting of correlated and uncorrelated variables into linearly uncorrelated variables, which are called principle components. Due to the normalization of the input data within unit interval and chosen based variance, the larger variances would have better discriminatory properties in the data.

The principal components (uncorrelated variables) are ordered so that the first principal component will have the largest variance within the data set and the last will have the least variance. In other words, PCA yields feature discrimination based on choosing a ranked approach of the variances of the dimensions. All principal components must also be orthogonal to one another, thus giving an orthogonal basis set. Let R be the correlation matrix and λ_i the corresponding ith eigenvalue of the matrix R, and Q the eigenvector matrix with q_i columns. It should be noted that Q is an orthogonal matrix ($Q^TQ = I$). In the spectral theorem one has:

$$[R]_{m\times m} = [A]_{m\times n}[A^T]_{n\times m} = \sum_{i=1}^{m} \lambda_i q_i q_i^T \tag{2.2}$$

One could rewrite a new basis by choosing eigenvectors q_i of the original data again with having C as coefficient vector which is the projection of A onto the principal directions:

$$A = \sum_{i=1}^{m} q_i c_i = QC = QQ^TA \tag{2.3}$$

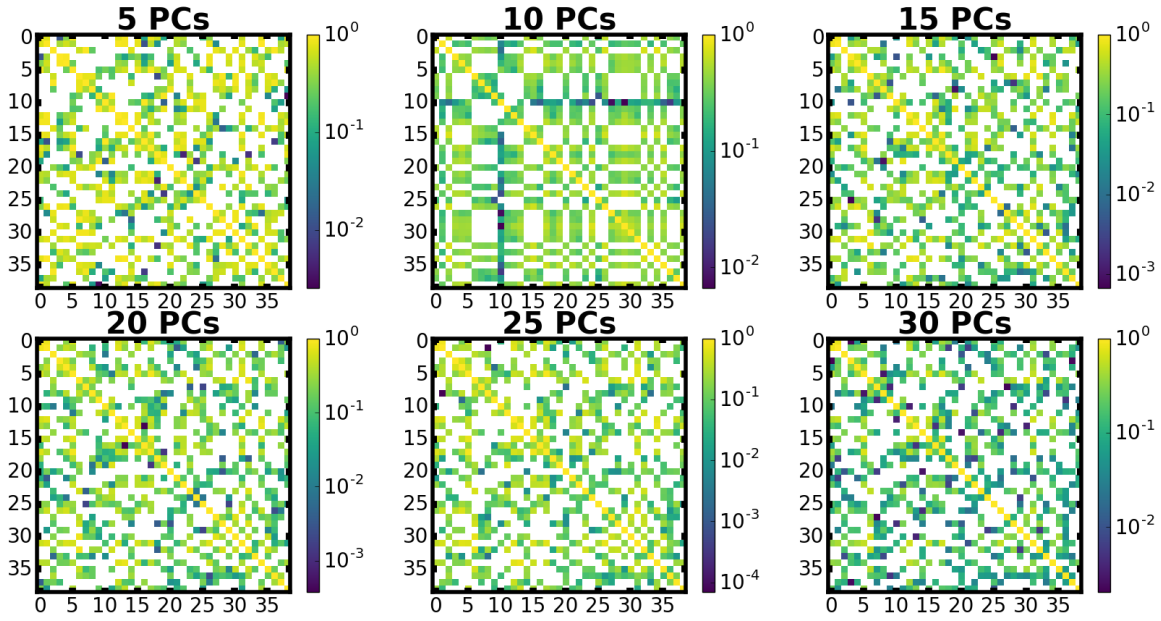


Figure 2.9: Correlation matrix for different numbers of principal components.

To understand the importance of PCA which is reducing the dimension of the data, one could have a rank k approximation A_r of the original data:

$$A_r = \sum_{i=1}^{k} q_i c_i = Q_r Q_r^T A \tag{2.4}$$

Employing all the above-mentioned equations, leads to a strategy such as subspace decomposition to find the largest eigenvalue and project the original data orthogonally onto its subspace. So, the eigenvalue closer to zero which plays a role as redundant information will be discarded [1]. The possible number of principle components is equal to or less than the number of subjects. Figure 2.9 shows the correlation matrix produced from getting the correlation of the principal components with each other. Again it is apparent that the numbers of high correlation values go down as the number of principal components go up.

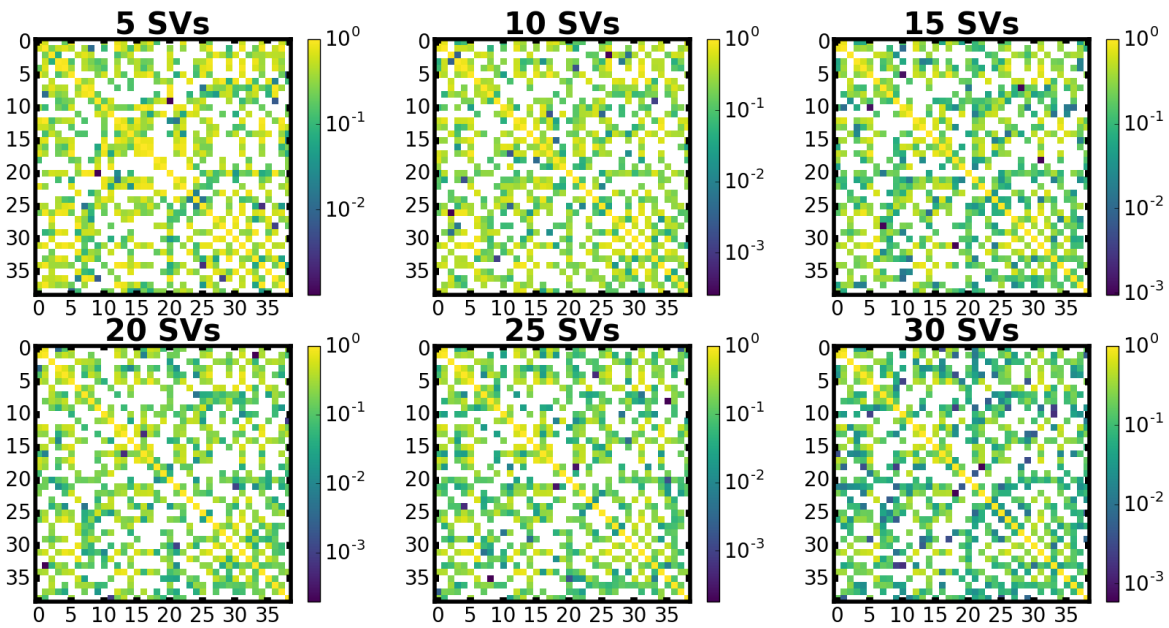


Figure 2.10: Correlation matrix for different numbers of singular values.

2.5.3 Singular Value Decomposition (SVD)

SVD is the factorization of matrix $[A]_{m\times n}$ to the form $U\Sigma V^T$, where U is a $m \times m$ unitary matrix, Σ is a $m \times n$ diagonal matrix, and V is an $n \times n$ unitary matrix. The diagonal values of Σ are the singular values of the original matrix and the columns of U and V are the left and right singular values of the original matrix respectively. When SVD is used in this paper, only the diagonal elements of the matrix Σ are used because this shows the properties of the matrix that can be used to compare with other matrices and reduces the dimension of the matrix.

$$[A]_{m\times n} = [U]_{m\times m}[\Sigma]_{m\times n}[V]^T_{n\times n} \tag{2.5}$$

Figure 2.10 is a graph of the correlation matrix of the Σ matrix with itself for different numbers of singular values. One can observe that the numbers of red (high correlated) values go down as the numbers of singular values go up.

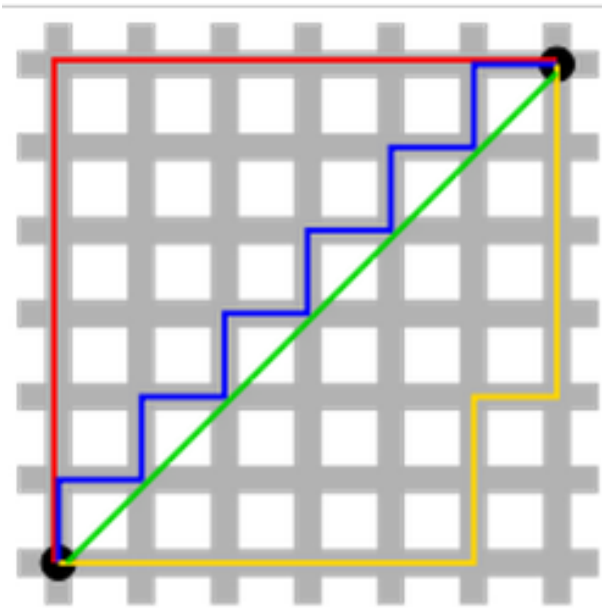


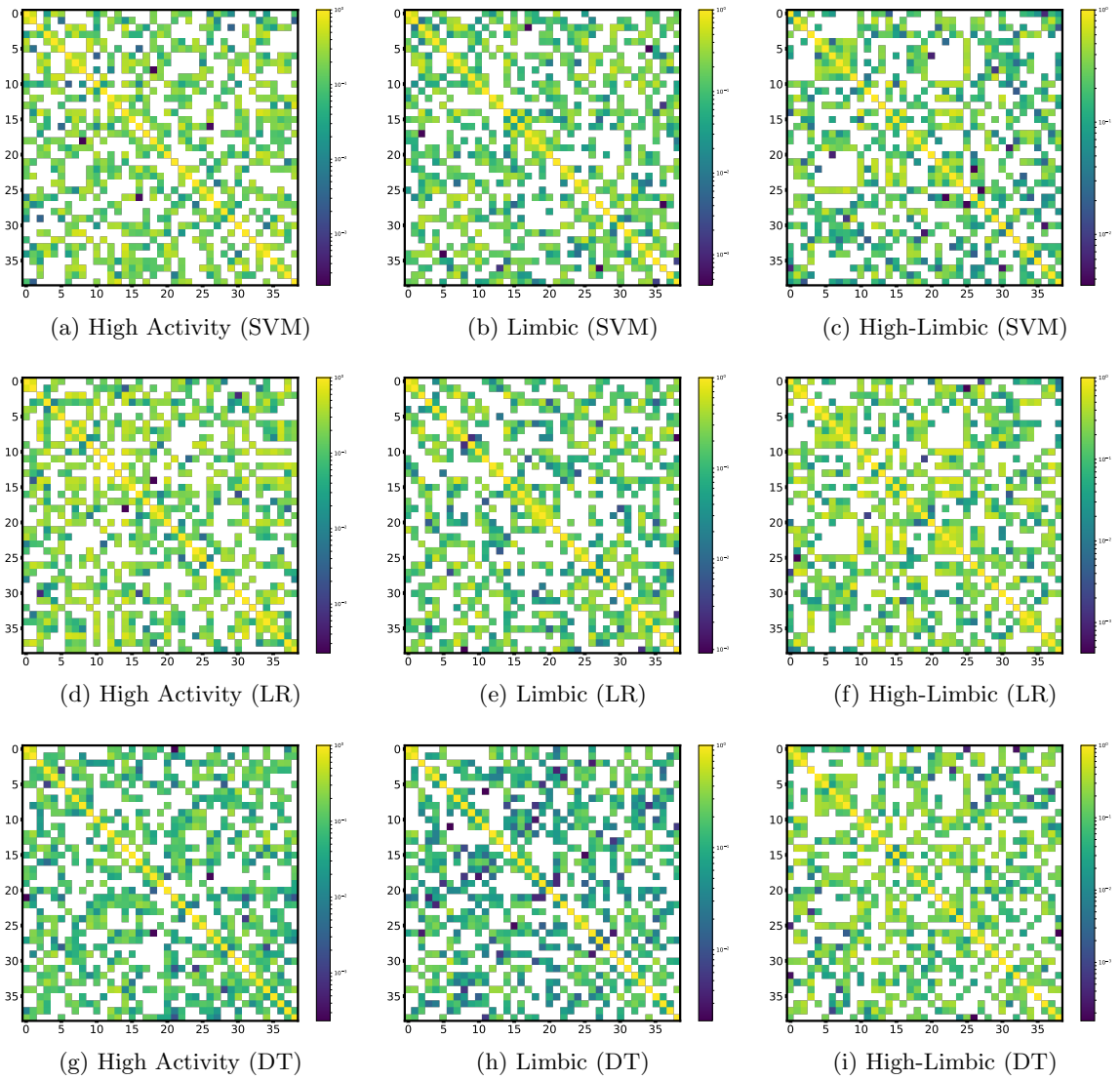
Figure 2.11: A schematic illustration of solution uniqueness of L1 and L2 regularization. The green line (L2-norm) is the unique shortest path, while the red, blue, yellow (L1-norm) are all same length (=12) for the same route.

2.5.4 Regularization

Regularization is one of famous processes previously used in the field of machine learning in order to prevent over-fitting. In fact, it is an additional term to the coefficients to fit to stay away from over-fitting with a perfect fit. The two famous regularization terms are L_1 and L_2 regularizations [67]. The difference between the L_1 and L_2 is just the difference between mean-absolute-error (MAE) and mean-squared-error (MSE). L_1 is the sum of the weights, but L_2 is the sum of the square of the weights. L_1 , L_2, and $\frac{L_1}{L_2}$ regularizations are also called Lasso, Ridge, and Elastic net, respectively. The L_1 and L_2 regularization terms for least squares are presented, as follows:

$$w^* = argmin_w \sum_j [t(x_j) - \sum_i w_i h_i(x_j)]^2 + \lambda \sum_{i=1}^{k} |w_i| \tag{2.6}$$

$$w^* = argmin_w \sum_j [t(x_j) - \sum_i w_i h_i(x_j)]^2 + \lambda \sum_{i=1}^{k} {w_i}^2 \tag{2.7}$$



(a) High Activity (SVM) (b) Limbic (SVM) (c) High-Limbic (SVM)

(d) High Activity (LR) (e) Limbic (LR) (f) High-Limbic (LR)

(g) High Activity (DT) (h) Limbic (DT) (i) High-Limbic (DT)

Figure 2.12: Correlation matrix for L_1 regularization using linear support vector machine (SVM) and logistic regression (LR) and tree-based feature extraction based on three different voxel selection schemes.

As shown in Figure 2.11, the green line (L_2-norm) is the unique shortest path, while the red, blue, yellow (L_1-norm) are all same length (=12) for the same route. By generalizing this illustration to n-dimensions, the uniqueness of the solution might be way more complicated. Therefore, L_2-norm with a unique solution will be a stable and unique solution with computational solution

efficiency. However, the L_1-norm solutions are not efficient on non-sparse cases, and the solutions might be unstable and not unique (multiple possible solutions). In addition to this, due to the sparse output of L_1 norm, it is usually used as built-in feature selection method [68]. Sparsity refers to that only very few entries in a matrix (or vector) is non-zero. As discussed, using L_1-norm would produce sparse output which contains many coefficients with zero values or very small values with few large coefficients. However, L_2-norm produces non-sparse coefficients and does not have the aforementioned property. Moreover, despite the L_1-norm, L_2-norm has an analytical solution to be computed in an efficient computational way. It should be noted that, due to the sparse property of the L_1-norm, it can be computed using sparse algorithms to increase the efficiency of the computational calculation [69, 70, 71]. The first hint that can be used in terms of choosing the right regularization method is the sample size. For example, L_1 regularized logistic regression requires a sample size that grows logarithmically in the number of irrelevant features. However, L_2 regularized logistic regression requires a sample size that grows linearly in the number of irrelevant features [67]. Figure 2.12 illustrates the correlation matrices for L_1 regularization using linear SVM and logistic regression and tree-based feature extraction based on all three voxel selection schemes that are previously discussed. As shown before, high activity, limbic system, and high-limbic voxel selection schemes extracted $94,720$, $25,200$ and $100,80$, respectively. Using regularization and tree-based feature selection, these numbers were also reduced. Figure 2.12 contains the following numbers of features including: the first row: 27, 35, and 37, the second row: 25, 23, and 32, and the the third row: 673, 749, and 797. It is interesting that with using tree-based feature selection, the process was ended up with amount of features which are almost 20 times larger that the number of features extracted by L_1 regularizations.

2.6 Machine Learning Algorithms

To run any classification task in medical imaging, one is required to employ a machine learning algorithm. There are two classification strategies for medical images. For the first strategy, measurements of a set of features from a region in an image would be employed as the feature vector. This is called region-based classification. For the second strategy, voxel-based classification (contextual or non-contextual information about every voxel) is used as a feature vector to feed into the classifier [1, 69, 70]. In this study, multivariate voxel-based analyses along with various classifiers

such as genetic programming, support vector machines, decision tree, and Gaussian Naive-Bayes were employed [72, 73, 74].

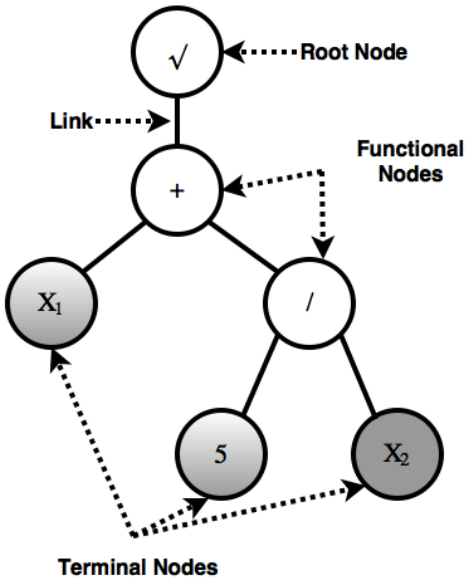


Figure 2.13: Tree representation of a GP model for ($\sqrt{X_1 + \frac{5}{X_2}}$).

2.6.1 Genetic Programming (GP)

Genetic Programming (GP) was employed due to the selection of designs applies on fitness measurement phase [75, 76]. GP was formulated as a symbolic optimization technique originally based on functional programming language as an evolutionary method [77] to use computer programs for solving a problem following the principle of Darwinian natural selection. Returning real values based on each tree and turning into class labels is the way that GP performs classification [78]. GP, instead of using one candidate, uses a group of individuals (population) and genetic operators to make new individuals (generations) guided by a function which measures the quality of each individual (fitness). In other words, having a higher probability of being selected for an individual at each generation would lead to having a better fitness measure [79].

Table 2.1: Parameters setting for Genetic Programming (GP) classifier.

Parameter	Setting
Population Size	500
Number of Generations	2000
Hall of Fame	300
Tournament Size	20
P Crossover	0.9
P Subtree Mutation	0.01
P Hoist Mutation	0.01
P Point Mutation	0.01
P Point Replace	0.05
Function Set	$add, sub, mul, div, log, neg, inv, abs$
Parsimony Coefficient	0.0005
Max Samples	0.9
Random State	0
Number of Jobs	3

It is always desired to solve a given problem in an efficient way. In this regard, the fitness function was calculated during evolution to have the most efficient guided GP [80].

$$Fitness = \frac{Number\ \ of\ \ patients\ \ classified\ \ correctly}{Number\ \ of\ \ patients\ \ used\ \ for\ \ training}$$

In the GP model to find the best mathematical formula, a crossover operator was used to select and replace the winner of the tournament with a stochastic subtree. In addition to this, to maintain the population diversity subtree mutations was added to the GP model. It also could have been done with point mutation, hoist mutation, and reproduction operators in the model. Table 2.1 lists the parameter setting used in the GP model.

Through each evolution, the developed GP model picks 300 (hall of fame) best programs from population size (500). Then, these programs compete in a tournament and just 20 of them will be considered for the next generations. In this regard, the probability of crossover, subtree, point, and hoist mutation (reproduction phase) was performed on a tournament winner. Moreover, the fitness of large programs have less probability of being selected based on the parsimony coefficient. In

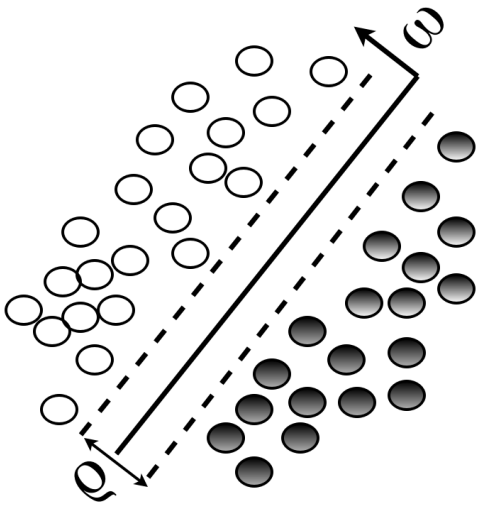


Figure 2.14: The classification of binary data using linear support vector machine with maximum margin ρ via assigning a weight vector w to the data.

other words, parsimony coefficient might decrease the computation times by controlling the depth or length of the program to earn better estimation of fitness and stay away from Bloat phenomenon. The maximum distance from its root node to the furthest leaf node is known as depth and the number of nodes in the program is known as length of the program. To decrease the cost of the evaluating the fitness of all programs, three cores (number of jobs) has been parallelized in the Python code to work on this part. It should be noted that the maximum number of generations and also perfect score have been chosen as stopping criteria to terminate the evolution early.

2.6.2 Support Vector Machine (SVM)

An SVM is a classifier that finds a hyperplane based on maximal margin rule. That is why SVM is also known as a maximal margin classifier. There are two types of linear and nonlinear support vectors. Linear SVM is the result of solving a constrained optimization problem for the quadratic objective function. The learning task was implemented as following [70]:

$$min \frac{||w||^2}{2} \tag{2.8}$$

$$subject \quad to \quad y_i(w.x_i + b) \geq 1 \qquad i = 1, 2, ..., N \tag{2.9}$$

For nonlinear separable data, one cannot apply linear decision boundaries. In this regard, nonlinear SVM is needed. It could be possible to apply the linear decision boundary to nonlinear data conditions. The nonlinear data has to be transformed into a new nonlinear space from its original coordinate space. Thus, in a new coordinate space, the linear decision boundaries could separate the samples shown as following [70]:

$$min \frac{||w||^2}{2} \tag{2.10}$$

$$subject \quad to \quad y_i(w.\phi(x_i) + b) \geq 1 \qquad i = 1, 2, ..., N \tag{2.11}$$

where $\phi(x)$ is the transformed attribute of x in a linear SVM. Curse of dimensionality is always a problem in high dimensional data classification. Employing an SVM classifier based on a method named kernel trick would help us to avoid this problem. Taking advantage of the cosine similarity measure and dot product helps us to define the kernel function in the following:

$$K(u, v) = \phi(u).\phi(v) \tag{2.12}$$

In this study, a radial basis kernel $K(x, y) = exp(\frac{-||x-y||^2}{2\sigma^2})$, polynomial degree of three kernel $K(x, y) = (x.y + 1)^3$, and sigmoid kernel $K(x, y) = Tanh(\kappa x.y - \delta)$ were employed.

The SVM classifier has the ability of being formulated as a convex optimization problem. It gives more freedom to choose an efficient optimization algorithm to find the global minimum of the objective function.

2.6.3 Decision Tree (DT)

Decision tree is a machine learning algorithm which partitions a set of input data recursively. A decision tree structure is made of a root node, (without incoming edges and also without or more outgoing edges), internal nodes which have one incoming edge and more outgoing and leaf nodes

which have one incoming and no outgoing edges [70]. Based on the volume of the data and available memory resources, decision tree algorithms can be implemented in a serial fashion such as CART (classification and regression tree) [81], C4.5 [82], and IDE3 (iterative dichotomizer 3) [83, 84] or in a parallel fashion such as SLIQ (supervised learning in ques) [85], and SPRINT (scalable parallelizable induction of decision trees) [86]. Dealing with a feature matrix of size 39×10 after data reduction and voxel selection scheme motivates one to employ a serial algorithm to implement the decision tree. In addition to this, decision trees are easy to interpret by boolean logic and can also be visualized.

The CART, by Breiman [81], builds both classification and regression trees based on binary splitting of the features selected based on Gini [87] index splitting measure. In principle, it follows the Hunt's algorithm [88]. It yields the largest information gain at each node. For a given training data set X and label vector Y, CART partitions the space recursively such that the matched instances and labels are clustered together [89]. For the D amount of data at node k, CART partitions the data into D_{Left} and D_{Right} subsets based on the splitting threshold and feature. The impurity at node k is calculated through the impurity measure Gini as $\sum_i p_{ki}(1 - p_{ki})$ with p_{ki} as the proportion of class i observations in node k [89]. For a data set of size $\{N$ samples $\times$ M features$\}$, the run time cost order to build the tree is $O(NMlog(N))$ and the query time is $O(log(N))$.

Based on the CART algorithm of how to build trees, if the splitting process continues to the point that there are few samples in each leaf of the tree, it is likely to over-fit the data. On the other hand, a small tree also might not capture the important structural information about the sample space. This problem is known as the horizon effect. Therefore, the complexity of the tree in such a way that the estimated true error is low is desired. In this regard, a reduced error pruning algorithm, which is a bottom up fashion pruning method, was employed. This improves predictive accuracy by starting at the leaves and replacing each node with its most popular class to reduce over-fitting and increase the simplicity of the tree and speed of the process. This process continues until the prediction accuracy is not affected. The optimization part was repeated 51 times for each of the data reduction methods to reach the most efficient result.

2.6.4 Naive-Bayes (NB)

Naive-Bayes is a classification technique based on Bayes Theorem with an assumption of independence among predictors to model probabilistic relationships between the feature matrix and the class labels [70]. In simple terms, a Naive-Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature. Bayes Theorem combines prior knowledge of the classes with new evidence gathered from training data [70]. First, the Naive-Bayes model builds the frequency table of the training data set. Then creates the likelihood table by calculating the probabilities. Finally, it calculates the posterior probability for each class and the class with maximum posterior probability is the result of the prediction. Naive-Bayes classifier is easy to implement, useful for big data problems, and known to outperform even highly sophisticated classifiers. In this paper, a standard algorithm for Gaussian Naive-Bayes and an optimized version of Naive-Bayes were employed.

The essential principle in Bayes method is assuming a known a-priori and then minimization of the classification error probability respectively. The class-conditional density function could be known or estimated from the available training dataset. During the Bayesian estimation, the training set conditioned density function is updated by the training set which acts as an observation to allow the conversion of the a priori information into an a-posteriori density [1].

A simple introduction can be given by considering two pattern classes: 1-NAC and 2-Placebo. To make the mathematical notations easier, they were named w_1 and w_2 respectively. Recalling the Bayes rule this can be seen as in the following implementation:

$$P(w_i|x) = \frac{p(x|w_i)P(x_i)}{\sum_{i=1}^{2} p(x|w_i)P(w_i)} \tag{2.13}$$

For the two-class patterns w_1 and w_2, the two Bayes classification rules are implemented below:

$$If \quad P(w_1|x) > P(w_2|x), \quad x \quad is \quad assigned \quad to \quad w_1 \tag{2.14}$$

$$If \quad P(w_1|x) < P(w_2|x), \quad x \quad is \quad assigned \quad to \quad w_2 \tag{2.15}$$

Considering the Gaussian probability distribution function with μ_i as the mean value and Σ_i as the covariance matrix for discriminant functions makes it more feasible to be solved.

$$p(x|w_i) = \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma_i|^{\frac{1}{2}}} exp(-\frac{1}{2}(x-\mu_i)^T\Sigma_i^{-1}(x-\mu_i)) \qquad i = 1, 2 \tag{2.16}$$

By choosing a monotonic logarithmic discriminant function this brings one to:

$$g_i(x) = -\frac{1}{2}(x-\mu_i)^T\Sigma_i^{-1}(x-\mu_i) + \ln P(w_i) - \frac{n}{2}\ln 2\pi - \frac{1}{2}\ln|\Sigma_i| \qquad i = 1, 2 \tag{2.17}$$

By calculating the mean vector and covariance matrices of the discriminant function for each class from the training data, the data can be separated by a hyperplane (if they have an equal covariance matrix) or hyperquadrics (if they have an unequal covariance matrix).

To optimize the Naive-Bayes algorithm, the bag-of-token model was employed [90]. In this way, the value of each feature k is calculated based on the non-negative number of occurrence of token k in the observation. For the estimated probability this is expressed as:

$$p(token = k|class = l) = \frac{1 + \beta_1}{N + \beta_2} \tag{2.18}$$

where β_1 is the weighted number of occurrences of token k in class l, β_2 is the total weighted number of occurrences of all tokens in class l, and N is the number of instances in the training set. Then, the classifier predicts the class label for each observation based on the estimated posterior probability presented in Equation 2.18. In this way, each observation is assigned to the class with the maximum posterior probability [91].

2.6.5 Boosting

The idea of boosting and weak learners for the first time was proposed by Schapire [92] in 1990. Boosting algorithms change the training data distribution iteratively. Thus, the base classifier would be trained to predict the exemplars that are hard to classify. This is totally a process which is different than bagging technique which is bootstraps of data according to a uniform probability distribution is aggregated. However, boosting assigns a weight to each training exemplar. The assigned weight can be changed at the end of each round of boosting adaptively. The assigned weights can be used as a sampling distribution to choose multiple bootstraps from the original training data set and can be also used as the base classifier. It should be noted that in case that

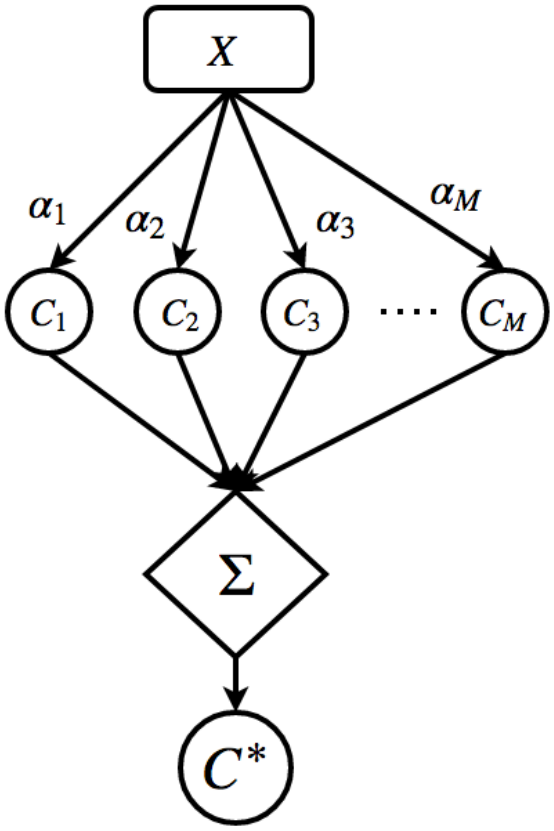


Figure 2.15: Schematic presentation of the AdaBoost algorithm. C_i indicates the base classifier, α_i indicates the importance of the base classifier C_i, and C^* indicates the best classifier after M rounds of boosting.

the weights are used as the base classifier, the classifier is likely biased towards the exemplars in the data that higher weights [70]. In this study, two variants of boosting (1) adaptive boosting (AdaBoost), and (2) extreme gradient boosting (XGBoost) were used in classification.

The first variant of boosting which is used in this study is AdaBoost. AdaBoost as shown in algorithm 2 depends on two factors: (1) base classifier, (2) error rates of weights. α_i explains the importance of the base classifier C_i. In better words, α_i gives weight to each prediction by the base classifier C_i without using any majority voting scheme. The key idea of AdaBoost convergence is minimizing an upper bound on the classification error [70, 69, 93]. Figure 2.15 shows a schematic illustration of the AdaBoost algorithm.

Algorithm 2: AdaBoost

Input: Input Variables X_i, Target Variables Y_i, i ∈ (1,N)

Output: $C^*(x)$

1 $w = \{w_j = \frac{1}{N}, j \in (1, N)\}$ Initializing the weights;

2 Initializing the number of boosting rounds as k;

3 **for** $i \in \{1, \ldots, k\}$ **do**

4 Bootstrap (X_i, Y_i) from (X, Y) according to w;

5 Train a base classifier C_i on (X_i, Y_i) ;

6 Apply C_i on all exemplars X;

7 $\epsilon_i = \frac{1}{N}[\sum_j w_j \delta(C_i(X_j) \neq Y_j)]$ Calculate the weighted error ;

8 **if** $\epsilon > 0.5$ **then**

9 $w_j = \frac{1}{N}, j \in (1, N)$ Reset the weights ;

10 Go to Line 4 ;

11 **end**

12 $\alpha_i = \frac{1}{2} \ln \frac{1-\epsilon_i}{\epsilon_i}$;

13 **if** $C_j(X_i) = Y_i$ **then**

14 $w_i^{(j+1)} = \frac{w_i^{(j)}}{\sum_i w_i^{(j+1)}} \times \exp^{-\alpha_j}$

15 **else**

16 $w_i^{(j+1)} = \frac{w_i^{(j)}}{\sum_i w_i^{(j+1)}} \times \exp^{\alpha_j}$

17 **end**

18 **end**

19 $C^*(X) = argmax_y \sum_{j=1}^{M} \alpha_j \delta(C_j(X) = Y)$;

The second variant of boosting is XGBoost [94]. Gradient tree boosting, also known as gradient boosting machine (GBM) or gradient boosted regression tree (GBRT) was originally proposed by Breiman and elaborated by Friedman in 2000 [95]. The key idea of gradient tree boosting is that it typically uses some variant of decision trees, i.e. CART algorithm, in a bounded size as the base (weak) learners and the quality of fit to each of the base classifiers can be changed with slight modifications. XGBoost uses some modifications including sparsity-aware split finding, weighted quantile sketch, and parallel structure which makes this algorithm scalable which is able to be used on high performance computing (HPC), and Apache Spark. In summary, a gradient boosting algorithm, first optimizes the loss function, makes the weak learner to predict the exemplars, and

uses an additive model to add weak learners to minimize the loss function. The type of loss function would change based on the tasks. For instance, a squared error would be suitable for regression and a logarithmic loss for classification. In addition to this, XGBoost has been implemented with the constraints applied on the additive model. For instance for the decision tree as an additive model, the number of trees, the depth of tree, number of terminal nodes, number of leaves for each tree, number of observations per split, minimum improvement to loss, and L_1 and L_2 weights (the value at each leaves) regularization can be chosen. This would improve the outcome dramatically in comparison to standard machine learning algorithms. That would be the reason that XGBoost was the choice of classifier for so many Kaggle competitions winners.

2.7 Deep Learning Algorithms

Emerging as one of the most contemporary machine learning techniques, deep learning has shown success as a reasonable solution for medical imaging problems, specifically image classification using the hierarchical architecture of multiple layers of non-linear information. Recently, most of the deep learning algorithms were designed to solve unsupervised learning problems. However, none of them has truly solved the problem and the deep learning algorithms for solving supervised problem are still more valid. It is desired to construct robust and powerful framework employing deep neural networks for supervised learning problems. The essential elements would be more layers and more neurons (units) within a layer with a specific structure [96]. In this section, some deep learning algorithms including autoencoders, and convolutional neural networks were described as new approaches to be used in this study.

2.7.1 Autoencoder

In this study, an autoencoder consisting of multiple convolutional layers was developed to learn the features of the fMRI images for each subject in order to predict the relapse. An autoencoder is a neural network which is trained to attempt to copy its input to its output using two parts: (1) an encoder function $h = f(x)$, and (2) a decoder function which reproduces a reconstruction $x' = g(h)$ of the input [96, 97]. Figure 2.16 presents the general schematic structure of an autoencoder. However, not in all cases the output of the decoder is the point of interest. In this study, it is desired that the trained autoencoder would extract some salient properties from the MRI images

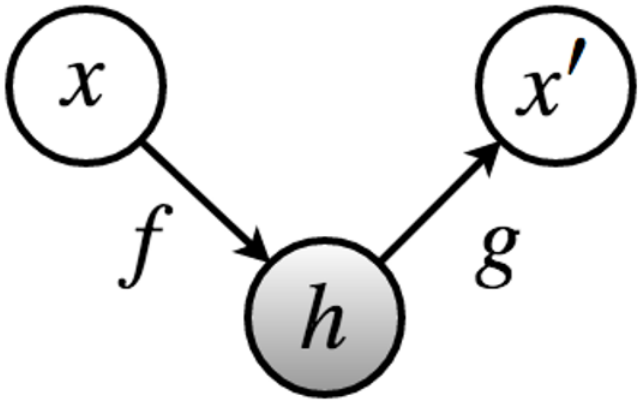


Figure 2.16: The general schematic structure of an autoencoder, mapping an input x to reconstruction x' via code h. The two essential components are: (1) encoder f which maps the input x to h, and (2) decoder which maps h to x'.

that could be used to predict the relapse in subjects. One of the possible ways is to employ undercomplete autoencoders by applying constraints on the input x to have smaller dimension. In this way, salient features can be extracted from the full dimension input ($80 \times 80 \times 37$) with a smaller dimension (i.e. $10 \times 10 \times 8$). In fact, learning an undercomplete representation forces the autoencoder to capture the most salient features of the training data [96].

The developed pipeline in this paper was written in Python employing various libraries including Keras, TensorFlow, Nipype, Nilearn, Nibabel, and Scikit-Learn [98, 99, 100, 101]. As shown in Table 2.2, the developed autoencoder contains six 2D convolutional layer with the same padding. In fact, the encoder includes the first five convolutional layer using a linear rectifier (ReLU) as the activation function and a Sigmoid function used as the activation function of the last convolutional layer (decoder). A stride size of (3×3), a pool size of (2×2), and a sample size of (2×2) were used in all convolutional, max pooling, and up sampling layers, respectively. Binary cross-entropy was used as the loss function and the Adadelta algorithm which is robust to sparsity was employed to optimize the hyper-parameters [97]. Figure 2.17 presents the flow of the developed pipeline to extract salient features using autoencoder (unsupervised phase) and build the feature matrix to feed into machine learning algorithms for classification (supervised phase).

Table 2.2: The autoencoder layer settings.

Layer Type	Kernel	Activation	Output Shape	# of Parameters
Input Image	-	-	$80 \times 80 \times 37$	0
Conv2D	[16, (3,3)]	ReLU	$80 \times 80 \times 16$	5,344
MaxPooling2D	[(2,2)]	-	$40 \times 40 \times 16$	0
Conv2D	[8, (3,3)]	ReLU	$40 \times 40 \times 8$	1,160
MaxPooling2D	[(2,2)]	-	$20 \times 20 \times 8$	0
Conv2D	[8, (3,3)]	ReLU	$20 \times 20 \times 8$	584
MaxPooling2D	[(2,2)]	-	$10 \times 10 \times 8$	0
Conv2D	[8, (3,3)]	ReLU	$10 \times 10 \times 8$	584
UpSampling2D	[(2,2)]	-	$20 \times 20 \times 8$	0
Conv2D	[8, (3,3)]	ReLU	$20 \times 20 \times 8$	584
UpSampling2D	[(2,2)]	-	$40 \times 40 \times 8$	0
UpSampling2D	[(2,2)]	-	$80 \times 80 \times 8$	0
Conv2D	[37, (3,3)]	Sigmoid	$80 \times 80 \times 37$	2,701
Total Trainable Parameters				10,957

The autoencoder was applied on both pre-treatment and post-treatment scans. The compressed images after the third MaxPooling2D layer with a size of $(10 \times 10 \times 8)$ were fed into eight similarity comparison metrics including (1) correlation coefficient (CC), (2) correlation ratio (CR), (3) L_1-norm based correlation ratio (CRL1), (4) mutual information (MI), (5) normalized mutual information (NMI), (6) Euclidean distance (ED), (7) Dice coefficient (DC), and (8) Jaccard coefficient (JC). Correlation coefficient (Pearson's correlation) measures the strength and direction of the linear relationship between the same voxels from the pre-treatment and post-treatment images. This measurement is based on the covariance of the voxels divided by the product of their standard deviations. Moreover, mutual information measures the mutual dependence between the same voxels from the pre-treatment and post-treatment images which quantifies the amount of information can be gained about the post-treatment voxel through the pre-treatment voxel and vice versa. Additionally, the Jaccard distance was also used which measures dissimilarity between the same voxels from the pre-treatment and post-treatment images. It is a completion to the Jaccard coefficient which can be calculated by subtracting the size of the intersection divided by the size of the union of the voxels from 1. In better words, it can be calculated by dividing the difference of the sizes of the union and the intersection of two voxels by the size of the union. Furthermore, Dice coefficient was also used that measures the similarity of the same voxels from the pre-treatment

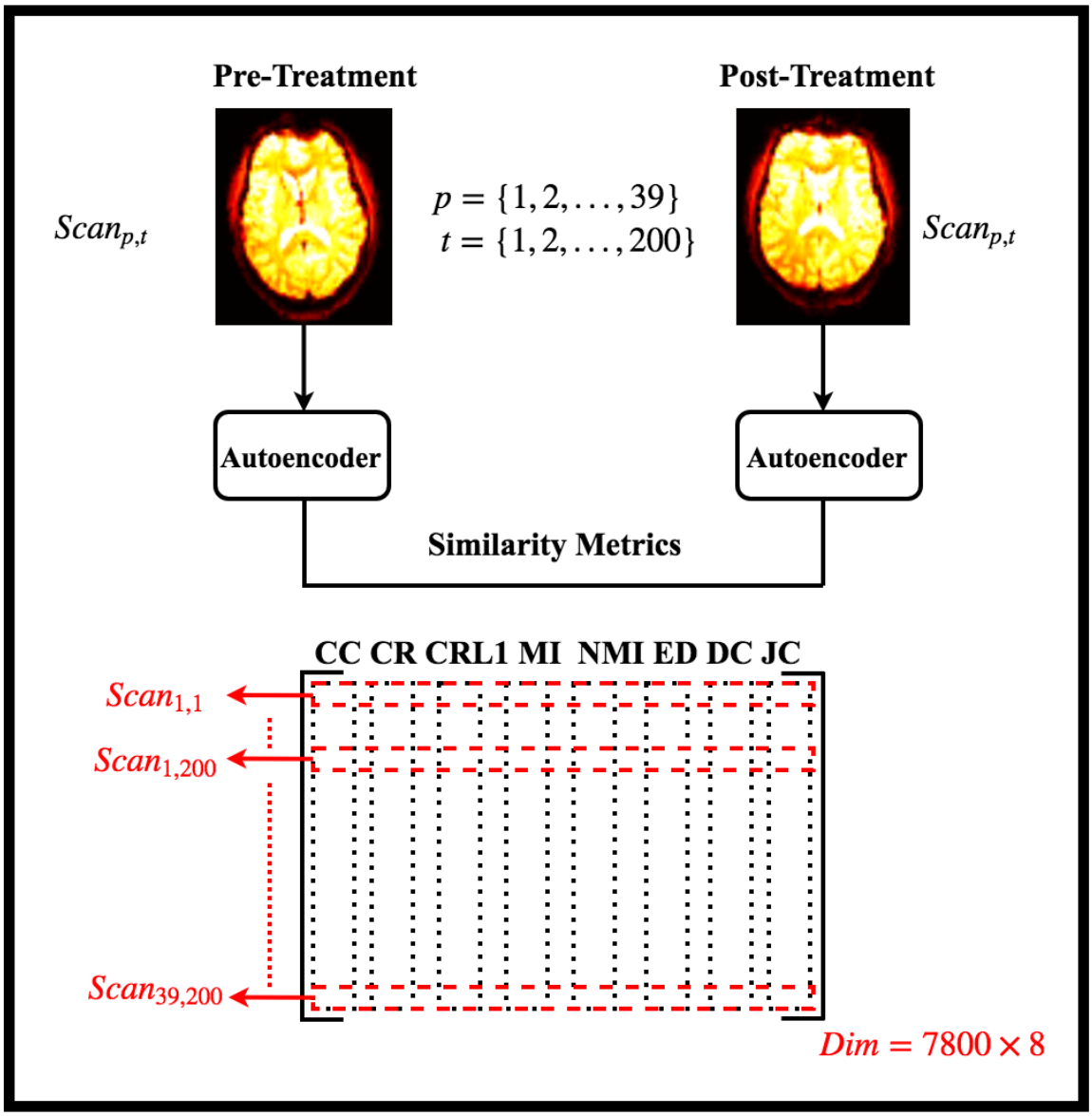


Figure 2.17: The flow of the developed pipeline to extract salient features using autoencoder (unsupervised phase) and build the feature matrix to feed into machine learning algorithms for classification (supervised phase).

and post-treatment images by computing the fraction of 2 times true positives divided by the sum of 2 times true positives, false positives, and false negatives. It is different from the Jaccard simi-

larity coefficient which only measures true positives once in both the numerator and denominator. Euclidean distance also measures how off is the voxel in the post-treatment image from the same voxel in the pre-treatment image which is used as the reference [70, 69, 71].

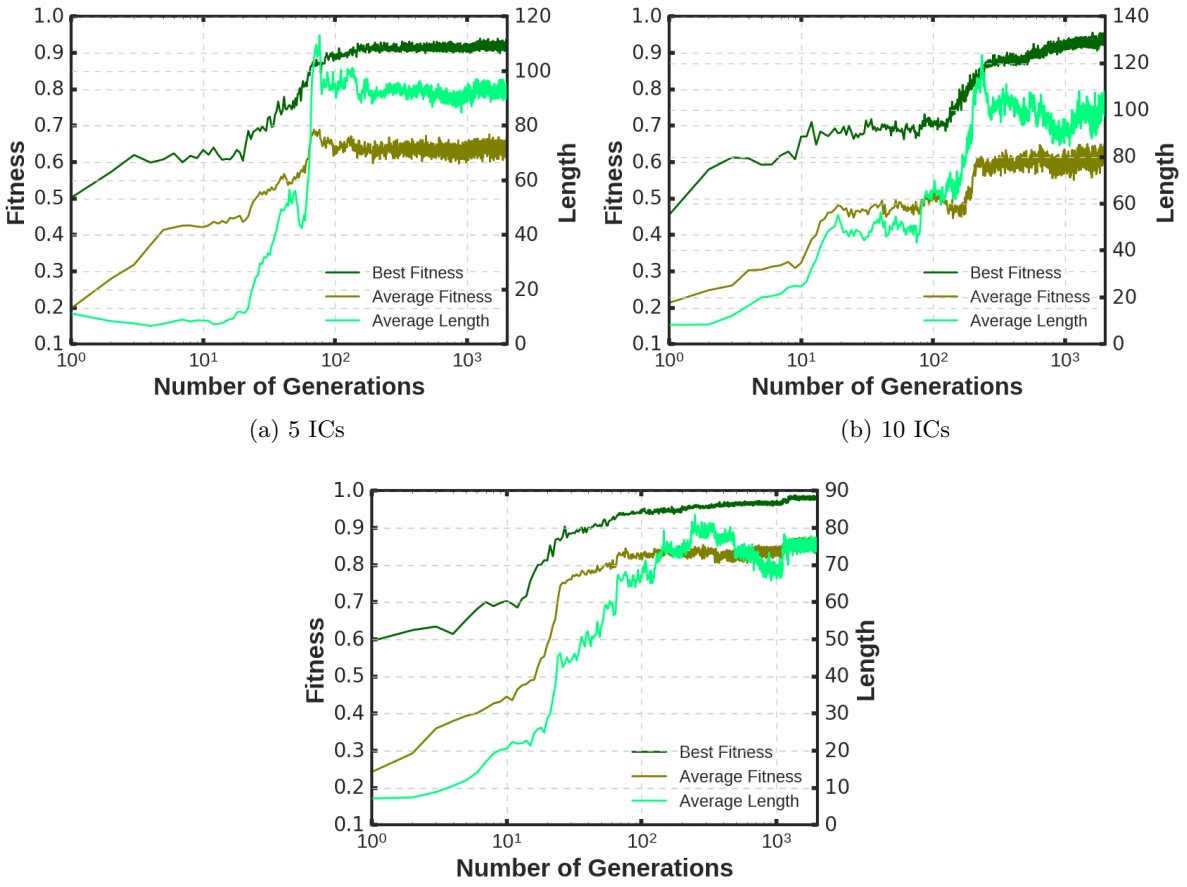
It was desired to extract salient features via comparing each of the 200 snapshots of the pre-treatment and post-treatment for each subject. This procedure was resulted in a feature matrix of size 7800×8. Then, the feature matrix was fed into robust classification algorithms along with Bayesian optimization to find the tuned hyper-parameters for each classifier. As discussed, two NIFTI image files (pre-treatment and post-treatment) were given for each subject (total 78 image files). Each NIFTI image which contains 200 snapshots requires $\sim 100MB$ on disk. However, the NIFTI format contains multiple compression layers and reading the NIFTI file of the each subject into NumPy array turned into $\sim 1.3GB$ which was led into a big data challenge ($\sim 100GB$). The training process of the autoencoder on only one subject using a normal equipment (Intel Core i7 2.2 GHz $\times$ 8 processor & 8 GB 1867 MHz DDR3 memory) took around 8 hours. Therefore, the developed pipeline was slightly changed to apply multiple computation stages in parallel. To overcome over-fitting, leave-one-out cross-validation was employed which also requires better equipment. Therefore, the developed pipeline ran on HPC. The wall-clock time was improved dramatically and the training and testing process including visualization stages were done in less than three days. It should be noted that the autoencoder model had to be trained 15600-times ($39 \times 2 \times 200$) which cost the major computational run-time of the project and it was almost impossible to be done using any normal computing equipment.

2.8 Results & Discussion

In this section, the voxel selection schemes were used in two different classification tasks: (1) validation, (2) relapse. For the model validation, we classified the subjects in terms of who received the drug NAC and who received the placebo. It would help to validate the feature extraction schemes due to the balanced numbers of subjects in each class. On the other hand, in relapse prediction the goal is to explore if the drug NAC is a factor in subjects quitting. Would it be possible to predict the relapse in the subjects based on fMRI brains scans? Due to the imbalanced numbers of subjects in each class, a model validation is required.

2.8.1 Model Validation

In this regard, four different classification methods including: (1) GP, (2) SVM, (3) DT, and (4) NB were employed to validate the voxel selection schemes.



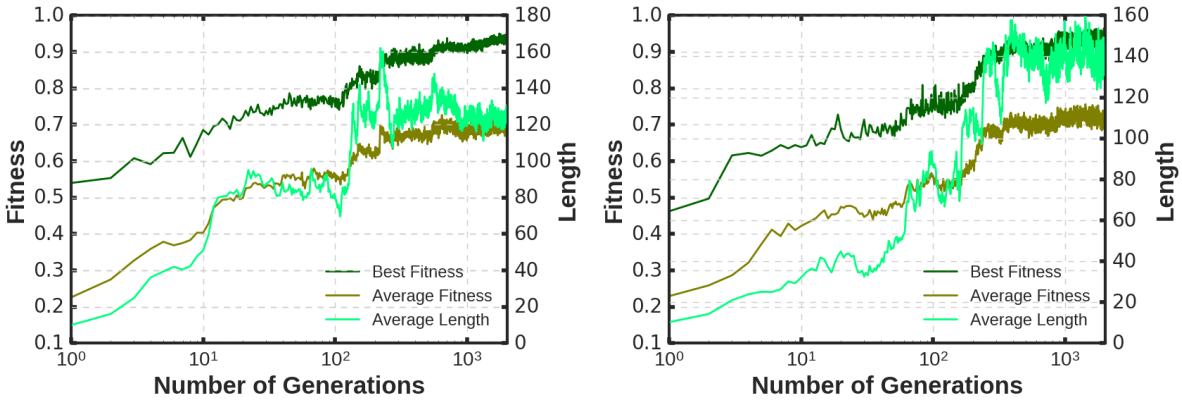
(a) 5 ICs

(b) 10 ICs

(c) 15 ICs

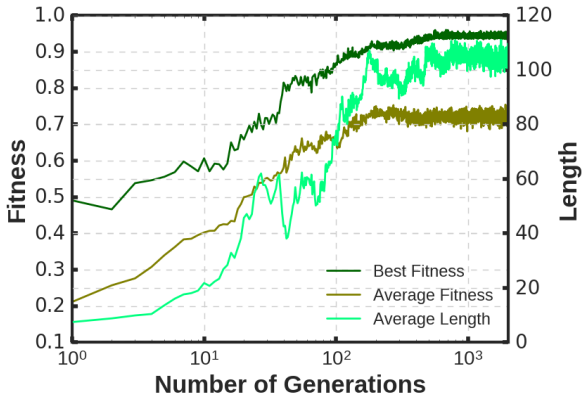
Figure 2.18: GP Evolution for 5, 10, and 15 independent components. The dark green, olive green, and light green lines present the best fitness, the average fitness, and the average length of the GP model, respectively.

In the first model, the results of the GP classifiers based on the initial trial for the preprocessing are presented. In this regard, a GP model with 2000 generations and 500 populations for classifications task with two major data reduction methods, ICA and PCA were developed in Python. Data reduction was done with 5, 10, and 15 independent components (IC) and principal compo-



(a) 5 PCs

(b) 10 PCs

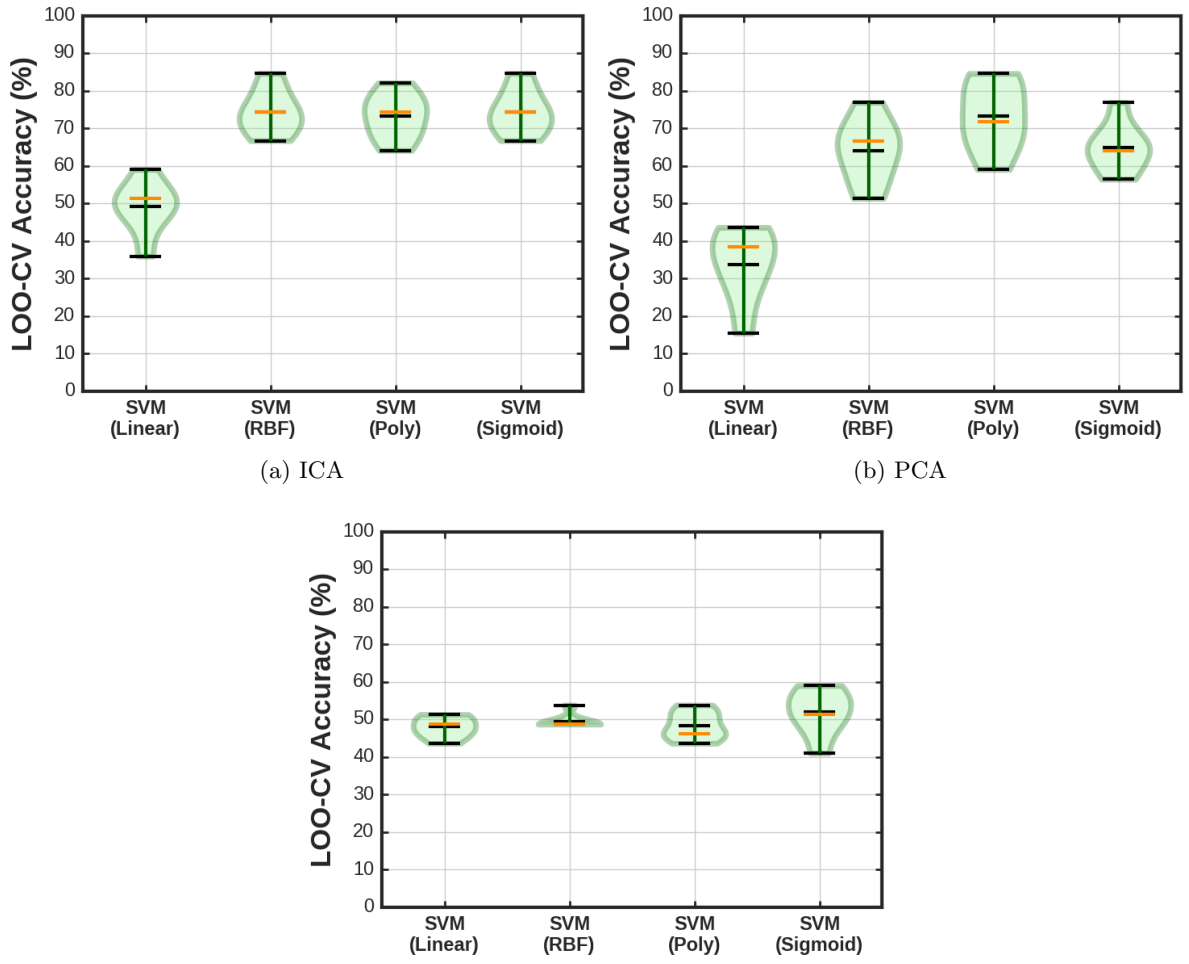


(c) 15 PCs

Figure 2.19: GP Evolution for 5, 10, and 15 principal components. The dark green, olive green, and light green lines present the best fitness, the average fitness, and the average length of the GP model, respectively.

Table 2.3: Classification accuracy for GP with different number of components of ICA and PCA data reduction methods.

Number of Components	ICA	PCA
5	64.10%	58.97%
10	64.10%	73.46%
15	68.71%	64.10%



(a) ICA

(b) PCA

(c) SVD

Figure 2.20: Leave-one-out cross-validation classification accuracy of SVM classifier with linear, RBF, polynomial degree 3, and sigmoid kernels for high activity mask. Each violin plot presents the statistics of the classification accuracy including the maximum (dark top line), minimum (dark bottom line), mean (dark middle line), and median (orange line).

nents (PC). For each one, best fitness, average fitness, and average fitness for a different number of generations were reported. The left Y axis was set to fitness, right Y-axis to length, and X-axis to generations in logarithmic scale.

As shown in Figure 2.18, as the numbers of independent components were increased, the average fitness was also increased. Assigning a higher fitness score to the classifier is the way fitness function

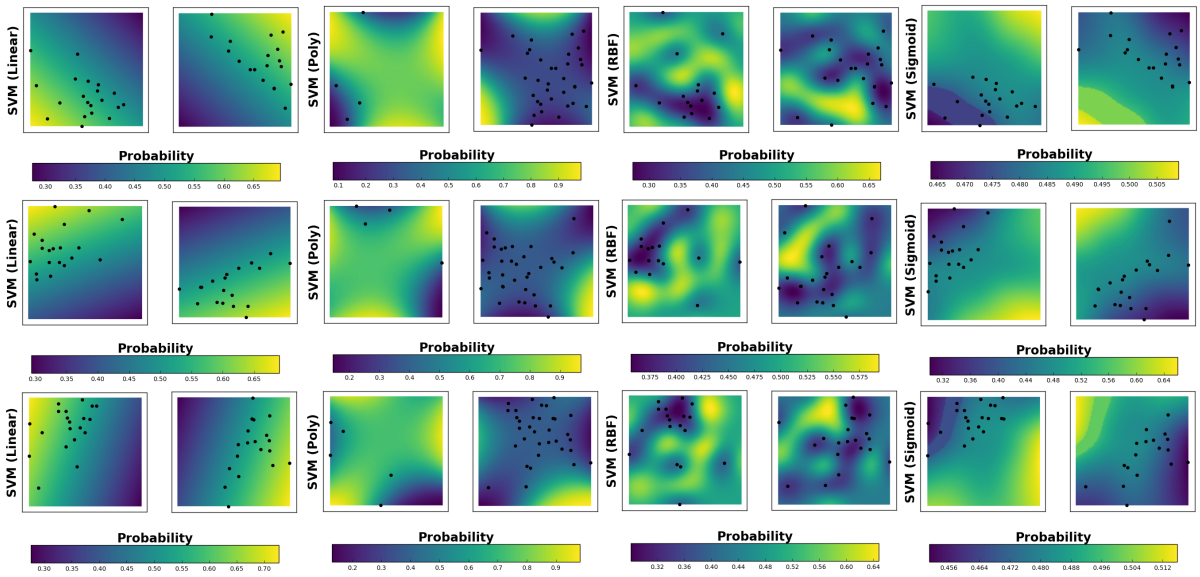
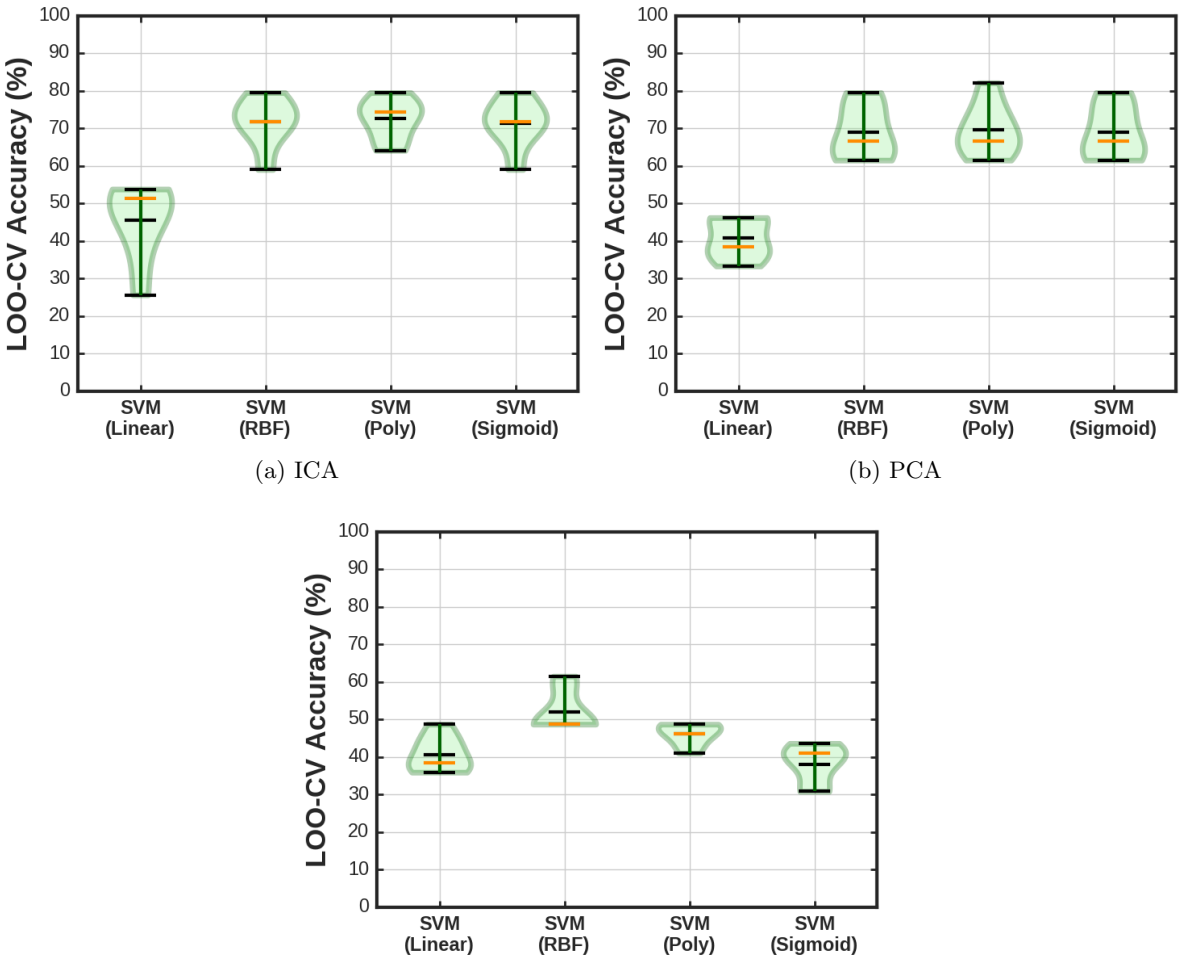


Figure 2.21: Decision boundaries of the linear, polynomial, radial basis function, sigmoid kernels for SVM for high activity voxel selection scheme. The first row of the figures corresponds to ICA, the second row to PCA, and the third row to SVD. For each of the sub-figures, the left one corresponds to the prediction of the subjects in the class 0 (placebo), and the right one corresponds to the prediction of the subjects in the class 1 (NAC).

classifies more samples using a smaller batch of features. Keeping this in mind, note that the best accuracy with 15 independent components is shown in Table 2.3. On the other hand, for 5, and 10 ICs, the GP model struggles with increasing the depth of the model to increase the fitness factor. One can observe that the average lengths for 5 and 10 independent components are around 90 and 100. Figures 2.19 demonstrates that this result matches with the results found in Figure 2.18. Looking closer, classification accuracy for the two different data reduction methods might change significantly for a different number of components, especially in the classification error which differs with data distribution for each method. In addition to this, as the number of principal components were increased, the length of the model decreased, and the best classification accuracy was found with 10 PCs [27, 102, 103].

In the second classifier, the results of SVM classifier with four different kernels including: (1) linear, (2) RBF, (3) polynomial degree 3, and (4) sigmoid were presented. In this regard, three different voxel selection schemes based on the high activity, limbic system, and the combination of



(a) ICA

(b) PCA

(c) SVD

Figure 2.22: Leave-one-out cross-validation classification accuracy of SVM classifier with linear, RBF, polynomial degree 3, and sigmoid kernels for limbic mask. Each violin plot presents the statistics of the classification accuracy including the maximum (dark top line), minimum (dark bottom line), mean (dark middle line), and median (orange line).

both were employed. For each classification task, leave-one-out cross-validation was employed to decrease the probability of over-fitting the model due to the low number of subjects.

Figure 2.20 illustrates the violin plots of classification accuracy along with leave-one-out cross-validation for ICA, PCA, and SVD data reduction methods based on the high activity map. As seen, SVM (RBF) with ICA and SVM (Poly) with PCA with 85% and 86% are the best. Also,

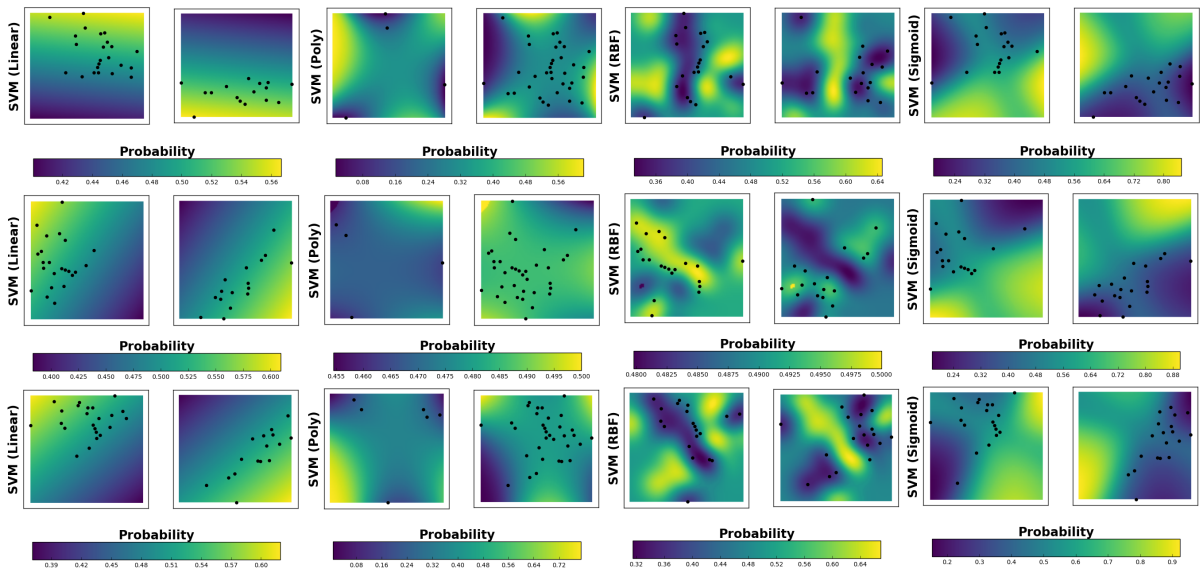
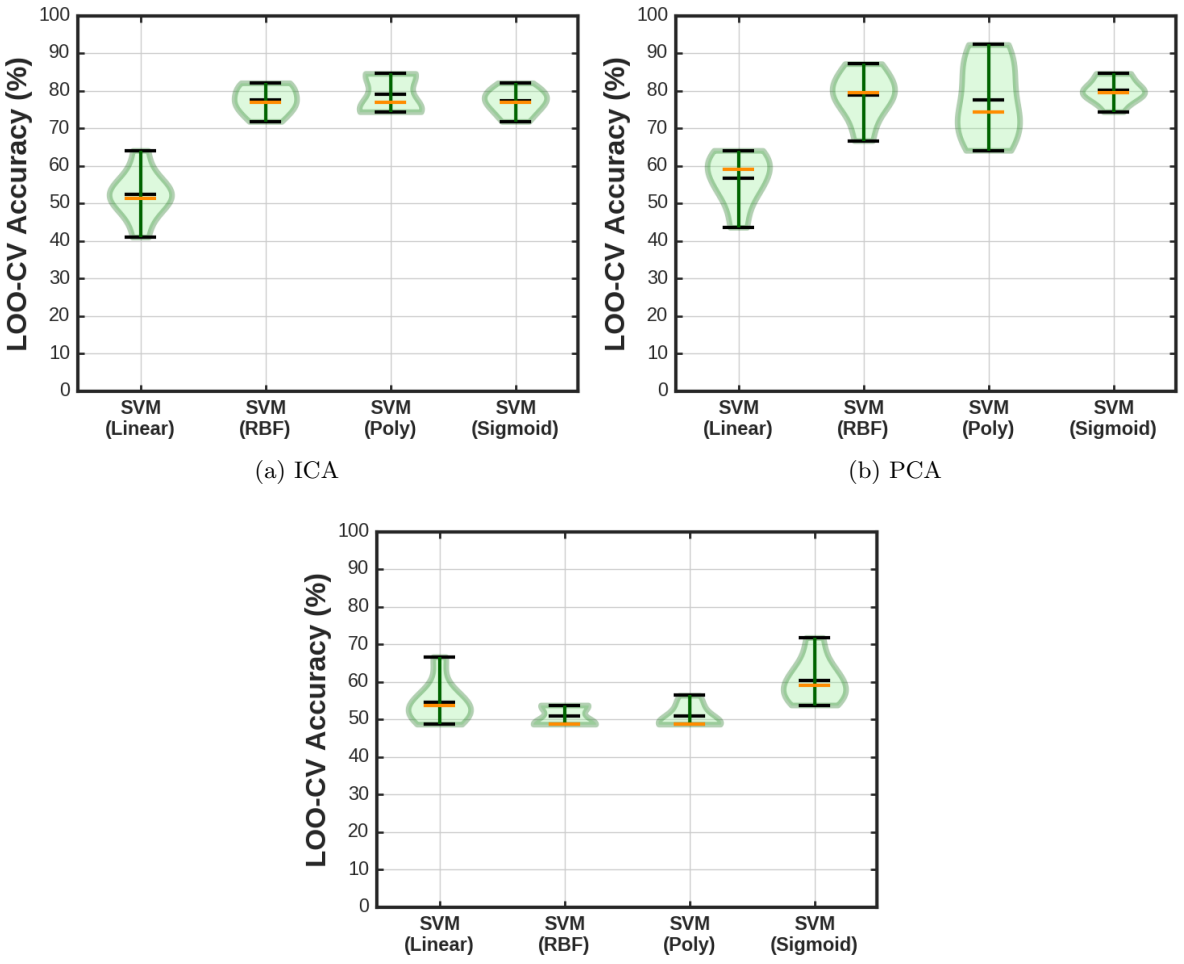


Figure 2.23: Decision boundaries of the linear, polynomial, radial basis function, sigmoid kernels for SVM for limbic system voxel selection scheme. The first row of the figures corresponds to ICA, the second row to PCA, and the third row to SVD. For each of the sub-figures, the left one corresponds to the prediction of the subjects in the class 0 (placebo), and the right one corresponds to the prediction of the subjects in the class 1 (NAC).

75% and 71% as mean accuracies have shown the best results among the employed classifiers for high activity maps. Moreover, SVM did not show a reasonable performance in the high activity map along with SVD data reduction methods. In addition to this, it can be found out from Figure 2.20 that linear SVM was not a great choice for the high activity map. To see the exact boundary decision for each of the classifiers, the decision boundaries of the classifiers for the high activity map were pictured in Figure 2.21. For each of the sub-figures, the X-axis was set to the first component, and Y-axis was set to the second component. Additionally, for each sub-figure the left plot demonstrates the decision boundaries for the class 0 (placebo) and the right one corresponds to the classification of the class 1 (NAC) as well.

Figure 2.22 demonstrates the classification accuracy of the SVM classifier with four different kernels along with ICA, PCA, and SVD data reduction methods for the voxels in limbic systems. SVM (Poly) along with PCA with 82% accuracy outperformed the other kernels along with different data reduction methods. Like Figure 2.20, linear SVM did not show a performance better than



(a) ICA

(b) PCA

(c) SVD

Figure 2.24: Leave-one-out cross-validation classification accuracy of SVM classifier with linear, RBF, polynomial degree 3, and sigmoid kernels for high-limbic mask. Each violin plot presents the statistics of the classification accuracy including the maximum (dark top line), minimum (dark bottom line), mean (dark middle line), and median (orange line).

50% which is little more than a random guess matched by the results presented by Smith et al. [14]. Figure 2.23 also presents the decision boundaries of the SVM kernels in the plane made by the first and second component for limbic voxel selection scheme.

The classification accuracy along with high-limbic mask was shown in Figure 2.24. Note that SVM with polynomial degree 3 along with PCA showed the best accuracy with 92% (max). How-

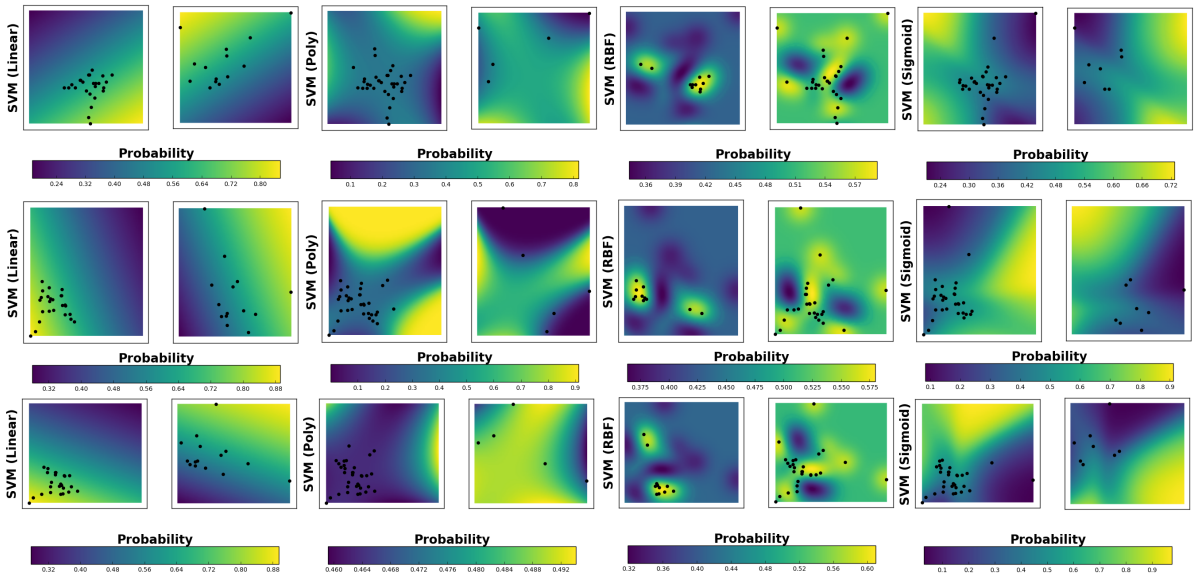
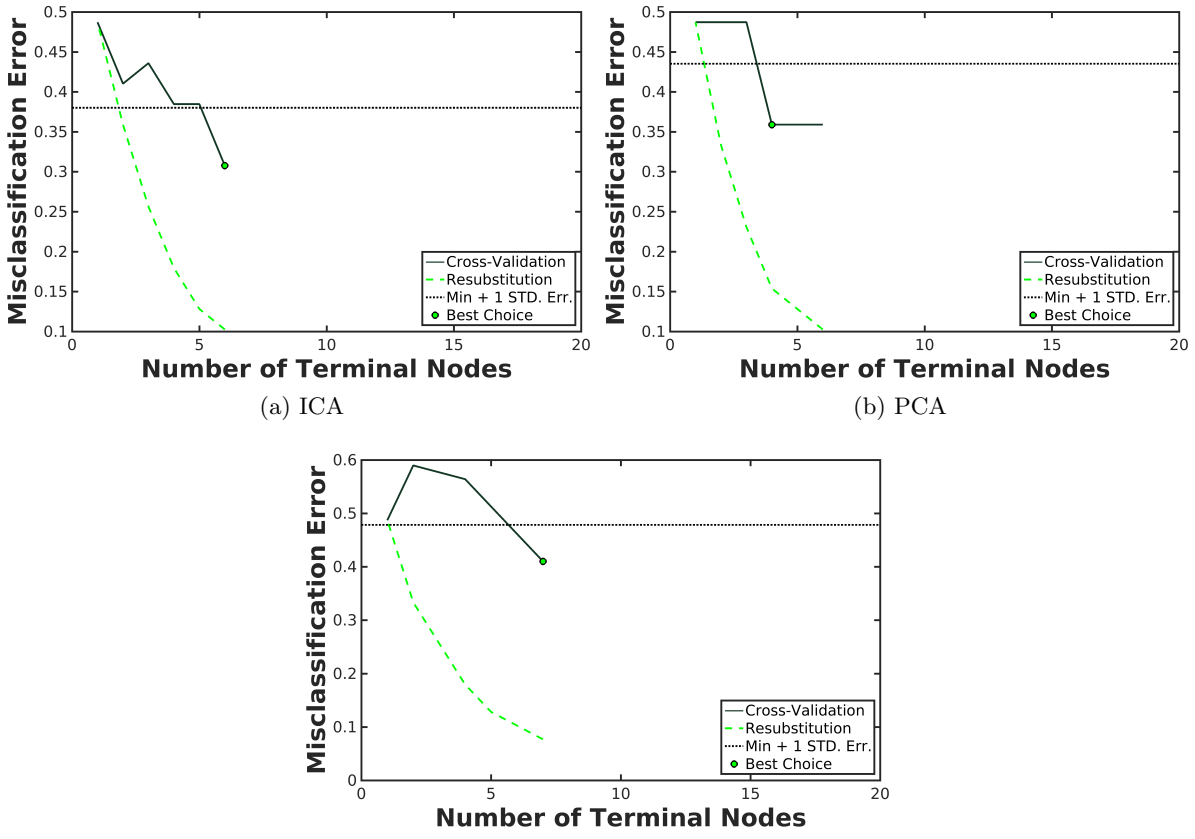


Figure 2.25: Decision boundaries of the linear, polynomial, radial basis function, sigmoid kernels for SVM for high-limbic voxel selection scheme. The first row of the figures corresponds to ICA, the second row to PCA, and the third row to SVD. For each of the sub-figures, the left one corresponds to the prediction of the subjects in the class 0 (placebo), and the right one corresponds to the prediction of the subjects in the class 1 (NAC).

ever, it should also be noted that the mean value of this classifier is 79%. This introduced a new idea in that max values in classification should not be considered. Bearing in mind that the goal is to find a reasonable value for classification in which the accuracy can be trusted, violin plots can be suggested in which the height of the violin is less leading to less standard deviations and the mean value also shows a reasonable accuracy. It should be noted that all the results are based on leave-one-out cross-validation. In addition to this, SVD along with high-limbic mask has shown better accuracies than high activity or limbic masks.

In the third part of the results, the CART algorithm was employed for the classification task for ICA, PCA, and SVD data reduction methods with 10 components and high activity voxel selection scheme. In this part, 10-folds cross-validation was employed. As seen in Figure 2.26, the misclassification error of the CART were plotted for ICA, PCA, and SVD data reduction methods. The solid dark green line corresponds to cross-validation (testing error), the dashed lawn green presents the resubstitution (training error), and the dashed black line shows the minimum error



(a) ICA

(b) PCA

(c) SVD

Figure 2.26: Misclassification error for the CART for different numbers of terminal nodes with ICA, PCA, and SVD data reduction methods.

plus one standard deviation of the error as a baseline for classification accuracy. The green circle presents the best choice of the classification accuracy. As shown in Figure 2.26a, the misclassification error and resubstitution decreased through the number of terminal nodes. After 6 terminal nodes, the best classification accuracy was found based on the bottom-to-top pruning. For PCA as shown in Figure 2.26b, the resubstitution error decreases through terminal nodes as expected and the cross-validation error reached a constant value after 4 terminal nodes. This bring about an idea that with ICA better accuracy with higher complexity was able to be achieved. However, with PCA there was reasonable accuracy with less complexity. On the other hand, for SVD as shown in Figure 2.26c the cross-validation went up after 2 terminal nodes and then it started decreasing

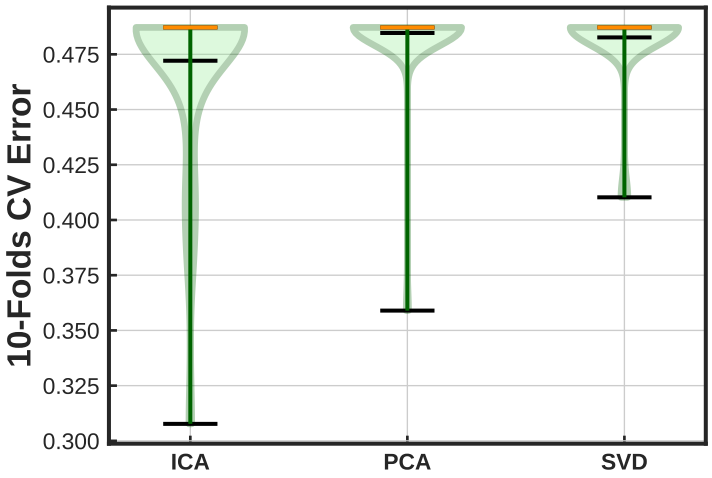
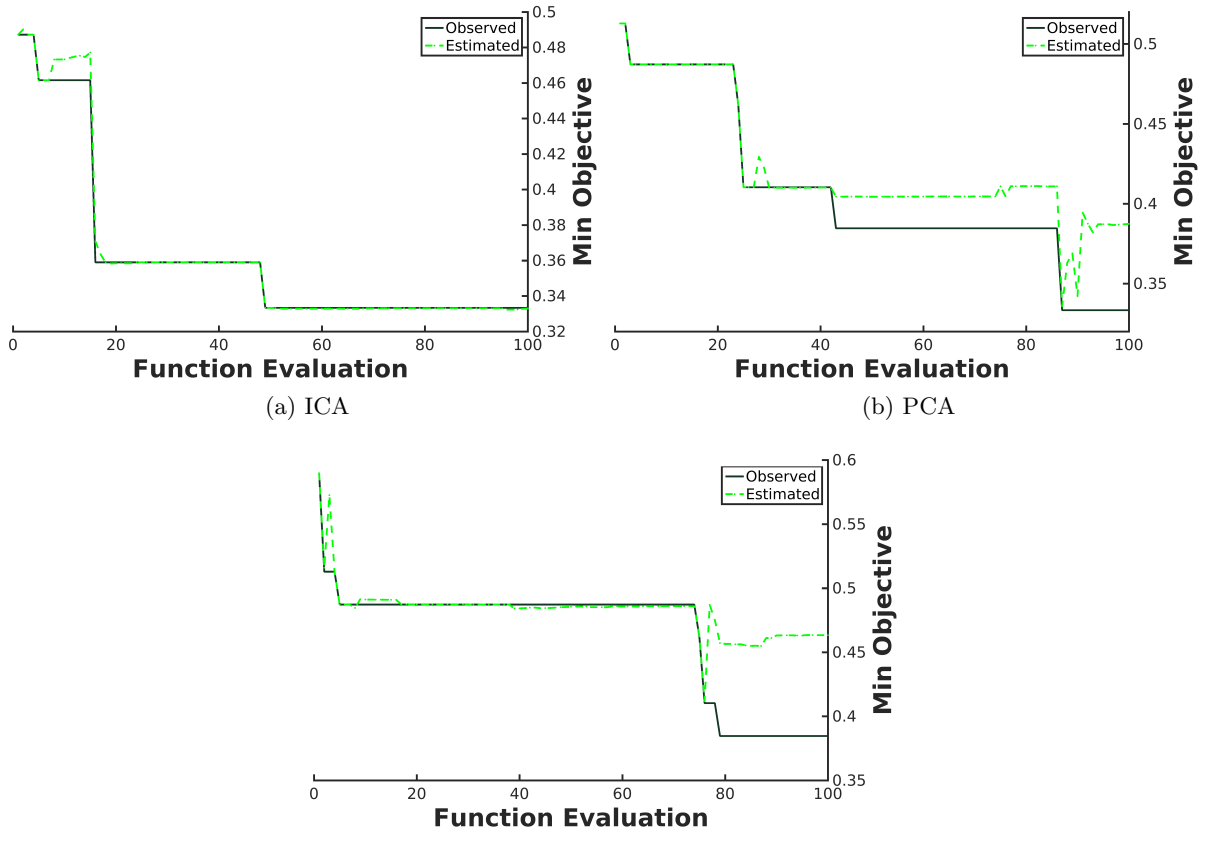


Figure 2.27: 10-folds cross-validation error for the CART classifier with ICA, PCA, and SVD data reduction methods. The top, bottom, and middle dark lines present maximum, minimum, and mean values of 51 times run, respectively. The orange line presents the median value.

and reached a stable point with 7 terminal nodes like ICA. The best misclassification errors were reported by ICA with 0.307, by PCA with 0.358, and by SVD with 0.410. The pruning process of the CART was done for 51 times and results have been shown in Figure 2.32.

One can observe the lack of deviation in the pruning process of the CART using ICA. However, the minimum error gained by PCA is around 0.255 which has a big deviation from mean value and median value of the 51 times run. SVD showed less deviation and the mean and median values are also the same. As discussed, the best accuracy was reported by the CART decision tree with zero deviation and 7 terminal nodes.

As the last algorithm, an optimized implementation of Naive-Bayes has been employed for the classification task. Figure 2.28 demonstrates the convergence rates of the Naive-Bayes for ICA, PCA, and SVD data reduction methods. In addition to this, Figure 2.29 shows the 10-folds cross validation error for the optimized Naive-Bayes classifier for different 51 times run with ICA, PCA, and SVD data reduction methods. As shown, the lowest mean value for error was received employing SVD as the data reduction method with 0.31. On the other hand, the ICA results show a symmetric deviation from the mean value. This result matched the result presented in Figure 2.28a, and the estimated values matched the observed value through the function evaluations.



(a) ICA

(b) PCA

(c) SVD

Figure 2.28: Convergence of the optimized Naive-Bayes classifier for ICA, PCA, and SVD data reduction methods.

To conclude, the best classification actuaries based on the high activity mask were presented in Figure 2.30 for ICA, PCA, and SVD data reduction methods. It is apparent that the ONB along with ICA, and the CART DT along with SVD outperformed the other machine learning classifiers.

2.8.2 Relapse Prediction

In this section, the validated model (previously presented in section 2.8.1) was employed to predict relapse in heavy smoker subjects. Similar to optimization procedure presented section 2.8.1, the CART model underwent an optimization procedure using the reduced error pruning algorithm using 10-folds cross-validation with 51 runs to achieve the best estimation of the error.

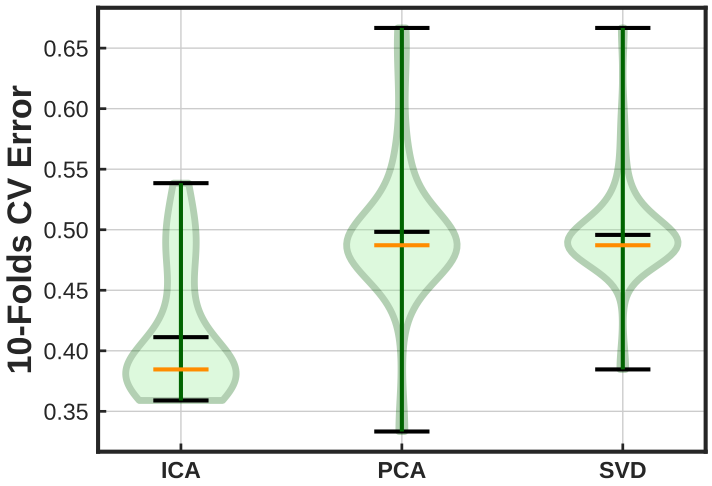


Figure 2.29: 10-folds cross-validation error for the ONB classifier with ICA, PCA, and SVD data reduction methods. The top, bottom, and middle dark lines present maximum, minimum, and mean values of 51 times run, respectively. The orange line presents the median value.

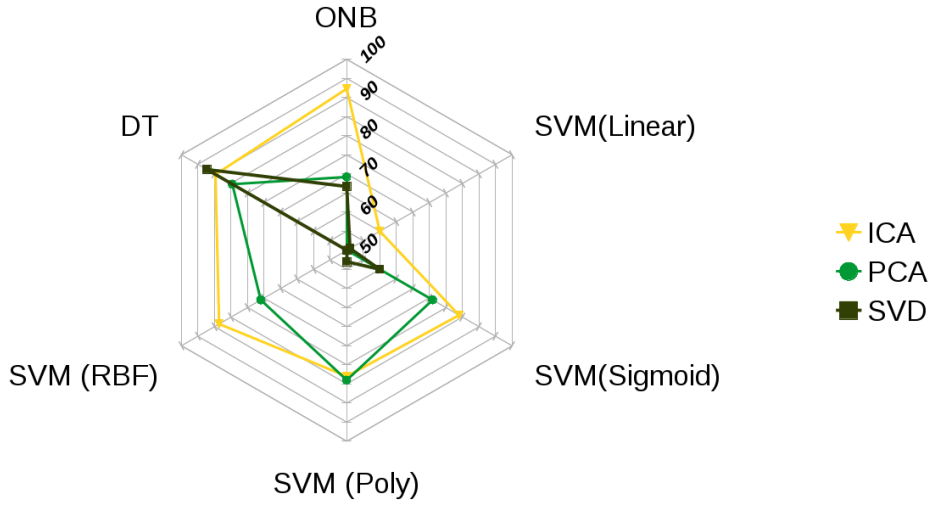
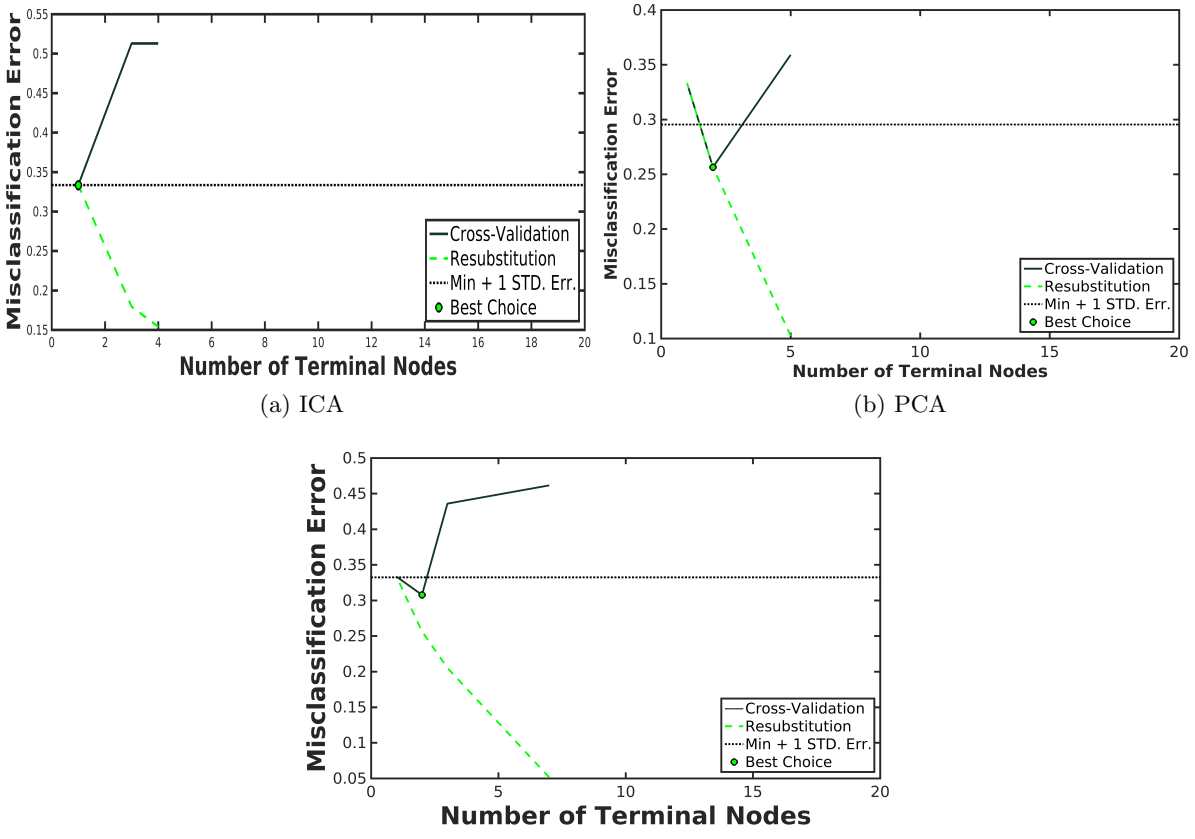


Figure 2.30: Radar plot of the best scores for the optimized Naive-Bayes (ONB), the CART decision tree (DT), SVM with four different kernels linear, polynomial degree three, radial basis function, and sigmoid for high activity mask.



(a) ICA

(b) PCA

(c) SVD

Figure 2.31: Misclassification error for the CART for different numbers of terminal nodes with ICA, PCA, and SVD data reduction methods.

Figure 2.31 shows the misclassification error for the CART using several data reduction method. This is a perfect example of bias-variance trade-off phenomenon [93]. As seen, the resubstitution error (training error of the built tree for prediction of the subjects in terms of relapse and non-relapse) decreased as the number of terminal nodes increased. SVD tree with 7, PCA tree with 5, and ICA tree with 4 terminal nodes have been made. As shown, the PCA tree has the largest margin with the defined metric line and the lowest error rate with 0.256 with respect to the other data reduction algorithms including ICA and SVD. It can be concluded that so far the PCA algorithm along with the CART showed the best performance.

In addition to this, Figure 2.33 shows the optimization of the Naive-Bayes algorithm after 51

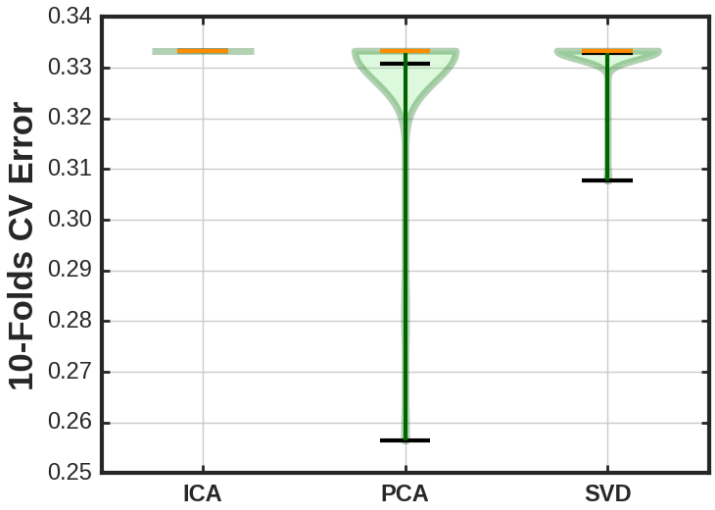
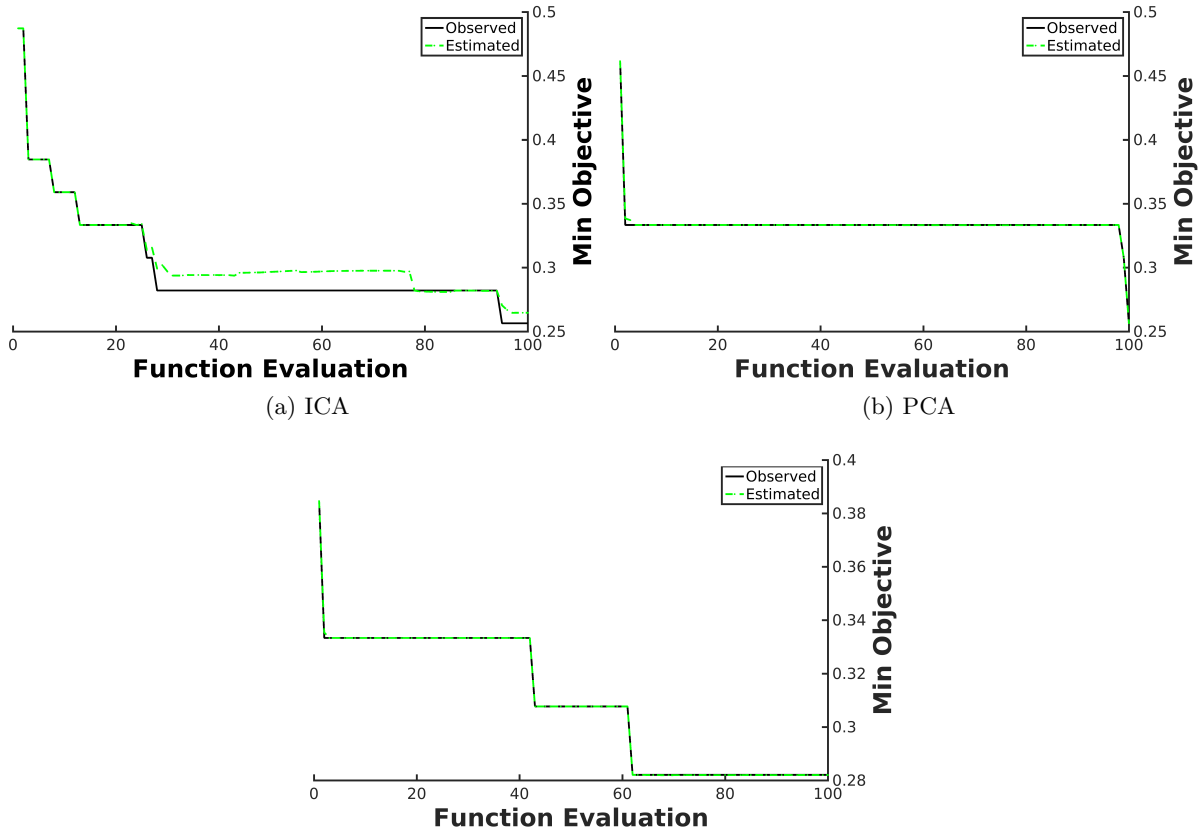


Figure 2.32: 10-folds cross-validation error for the CART classifier with ICA, PCA, and SVD data reduction methods. The top, bottom, and middle dark lines present maximum, minimum, and mean values of 51 times run, respectively. The orange line presents the median value.

runs. Therefore, the classifier was able to predict the subjects in the non-relapse even better with a reasonable performance. As shown in Figure 2.33, the solid line is the observed error rate and the dashed line is the cross-validation estimate of the true error rate. After about 25 function evaluations for ICA, the estimated error rate did not match the observed error rate. However, it reached the minimum estimated error of 0.282. On the contrary, the estimate of the true rate exactly matched the observed error rate employing SVD data reduction scheme and reached the minimum value of 0.282.

Figure 2.34 present the violin plot illustration of the optimization process of the ONB classifier during 51-times run for different data reduction methods. As previously shown in Figure 2.33c, the ONB along with the SVD reached the minimum 10-folds cross-validation error.

To compare ICA and SVD algorithms, we can say that both of the algorithms reached the minimum estimated error. However, the overall prediction accuracy of the SVD algorithm was 13% better than ICA. It is true that the ICA reached an estimated error better than PCA, but the ONB classifier along with the PCA algorithm even predicted the subjects with better accuracy. This could be due to the same nature of the PCA and SVD as previously discussed in data reduction



(a) ICA

(b) PCA

(c) SVD

Figure 2.33: Convergence of the optimized Naive-Bayes classifier for ICA, PCA, and SVD data reduction methods.

section. The ONB along with the ICA algorithm predicted all the subjects in the relapse class correctly. However, it failed at the prediction of 61.5% of the subjects in the non-relapse class.

In conclusion, ICA could not extract enough structural information to be employed in prediction of non-relapse subjects for this study. It should be noted that even ICA, along with the ONB classifier, showed reasonable results in predicting subjects in the relapse class. However, it seemed ICA could not extract salient features to detect the subjects in the non-relapse class. This might be the reason that the FN values were high. This is so obvious by considering ICA along with the CART classifier, where the line represents a classifier that did not do better than random prediction.

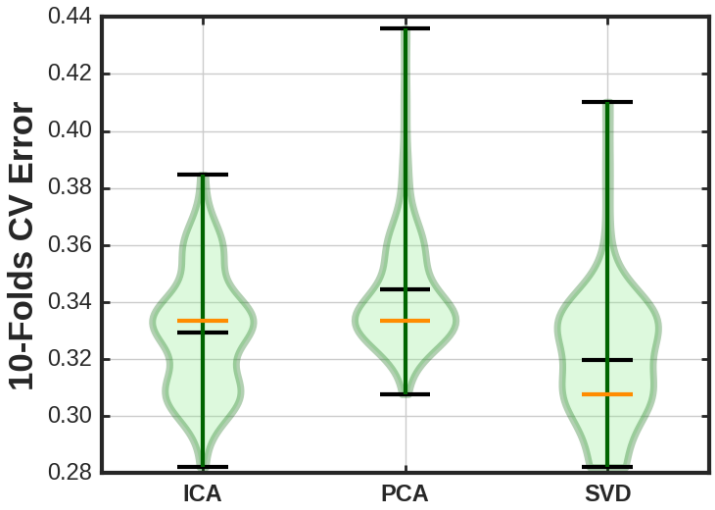
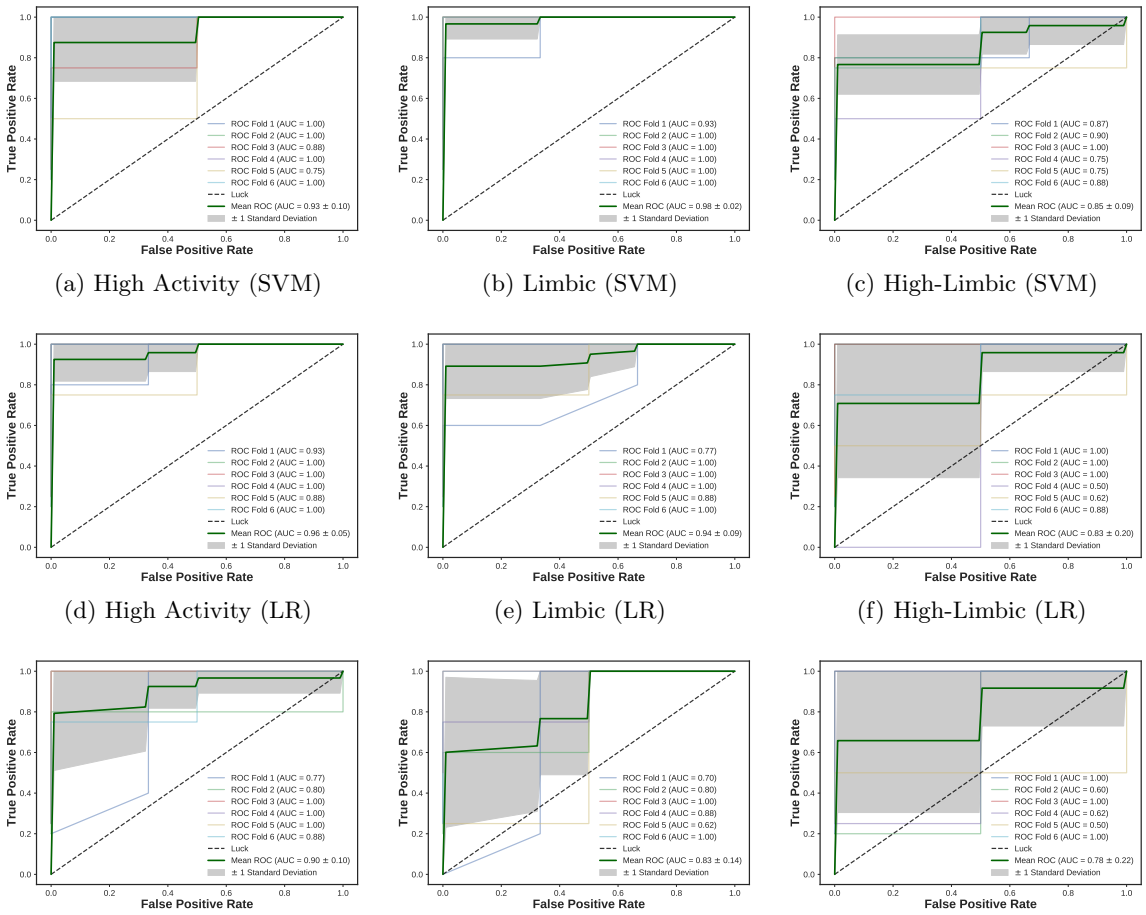


Figure 2.34: 10-folds cross-validation error for the ONB classifier with ICA, PCA, and SVD data reduction methods. The top, bottom, and middle dark lines present maximum, minimum, and mean values of 51 times run, respectively. The orange line presents the median value.

Next, the classification results employing random forests based on L_1 regularization and tree-based feature extraction are shown in Figure 2.35. In this regard, L_1 regularization and tree-based feature extraction applied on three different region of interests. As previously defined, three voxel selection schemes including high activity areas of the brain, limbic system, and the high activity parts of the limbic system were chosen as the ROI masks. For L_1 regularization linear support vector machine and logistic regression, and extra trees classifier for tree-based feature extraction were chosen as the base classifiers. The hyper-parameters of each classifier were optimized using a Bayesian optimization method [101, 89]. Figure 2.35 shows the ROC curves employing 6-folds cross-validation as the first row presents the L_1 regularization using linear SVM, the second row illustrates the L_1 regularization using logistic regression, and the third row shows the tree-based feature extraction using extra trees classifier. It turned out that the L_1 regularizations along with limbic system resulted in the best results in prediction of relapse in heavy smokers. LR with a mean AUC value of 0.94 ± 0.09, and SVM with a mean AUC value of 0.98 ± 0.02 showed the best results. The gray shaded area around the ROC curves presents the confidence interval of the classification. It can be used as a performance metric to see how valid are the presented results. As



(a) High Activity (SVM) (b) Limbic (SVM) (c) High-Limbic (SVM)

(d) High Activity (LR) (e) Limbic (LR) (f) High-Limbic (LR)

(g) High Activity (DT) (h) Limbic (DT) (i) High-Limbic (DT)

Figure 2.35: ROC curves of classification using random forests based on L_1 regularization using linear support vector machine (SVM), logistic regression (LR) , and tree-based feature selection extracted from the features produced by three voxel selection schemes from high activity areas of brain, limbic system, and the combination of both masks employing 6-folds cross-validation.

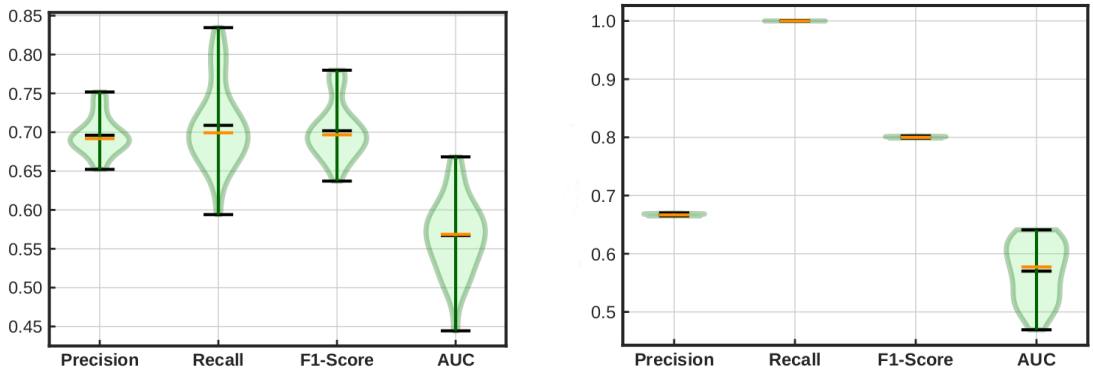
shown the results related to the high activity areas in the limbic system (high-limbic) as presented in the third column of Figure 2.35, the area covered in gray are the most in comparison to the other two voxel selection schemes. In better words, employing L_1 regularizations along with linear classifiers such as SVM and LR extracted salient features from the low activity parts of the brain which is interesting and can be used new biomarker in this research. However, the results presented using the tree-based feature extraction have less accuracy than L_1 regularization. However, the

result using features extracted from the high activity regions of brain using tree-based model is in agreement with the previously presented results [27]. In conclusion, L_1 regularizations can be a proper choice for region-based feature extraction. This can be proved through re-doing some examples in white-box researches such as speech and finger-tapping which we do know what parts in the brain are involved. Applying the ROI masks on the brain and selecting those voxels along with L_1 regularization would confirm the presented results [104, 105].

Finally, the salient features that were extracted from fMRI scans using convolutional layers developed as an autoencoder and similarity metrics were fed into various machine learning algorithms for the classification of the subjects into relapse and non-relapse classes employing leave-one-out cross-validation to overcome over-fitting. The results of seven classification algorithms including (1) decision tree (DT), (2) random forest (RF), (3) quadratic discriminant analysis (QDA), (4) k-th (k=3) nearest neighbors (kNN), (5) support vector machine (SVM) with a radial basis function (RBF) kernel, (6) adaptive boosting (AdaBoost), and (7) extreme gradient boosting (XGBoost) were presented in Figure 2.36.

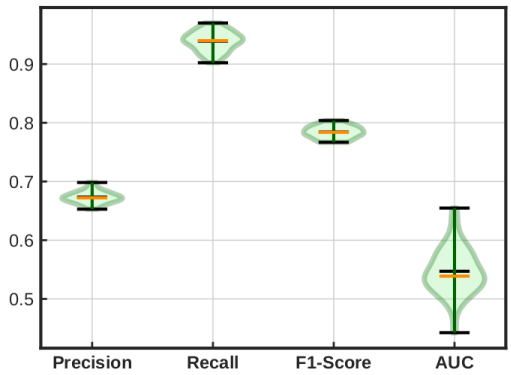
The violin plots of four different classification metrics including precision, recall, F1-score and AUC were presented based on the results of each fold of the leave-one-out cross-validation (total 39-times). As seen, the best recall (also named as true positive rate or sensitivity) was gained using SVM. However, as shown in Figure 2.36b, SVM did not show to be a very promising model with a mean AUC value of 0.57 ± 0.04. As shown, the challenge would be decreasing of false positives in the model prediction. Most of the models have shown reasonable results in prediction of true positives. However, the results of precision and AUC (the metrics are related to false positives) are around 60%-70% except for XGBoost model. As shown in Figure 2.36f, the XGBoost model has a mean precision value of 86%, mean recall value of 95%, mean F1-score of 90%, and mean AUC value of 92% which is reasonable enough to prove that the salient extracted features can be used to predict relapse. As seen, the XGBoost model outperformed the other machine learning algorithms. This is clearly shown in Figure 2.37f which is the ROC curves for leave-one-out cross-validation using XGBoost.

All the presented ROC curves of the 39-folds are within the shaded gray confidence interval. Each lighter curve shows the ROC curve for each fold of cross-validation (total 39-folds). As the ROC curve gets closer to top left corner, the AUC value will be higher and the model would

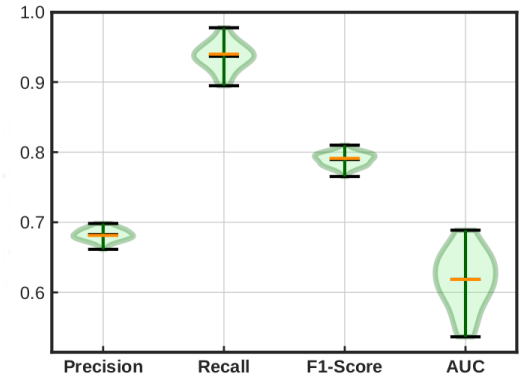


(a) DT

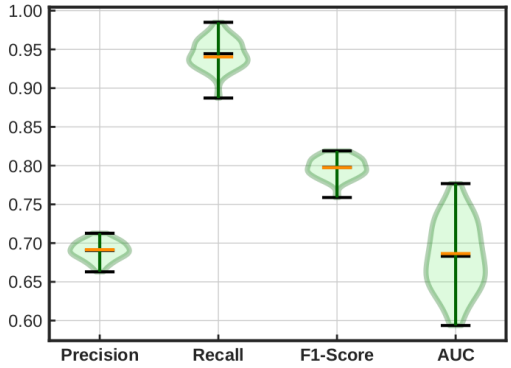
(b) SVM



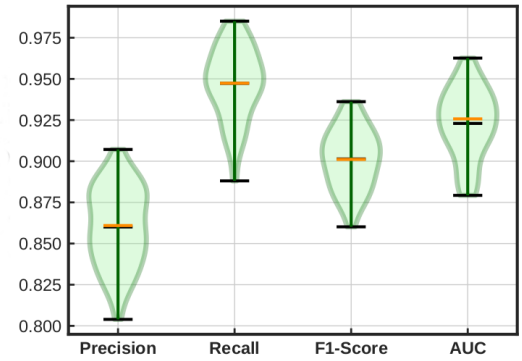
(c) QDA



(d) RF

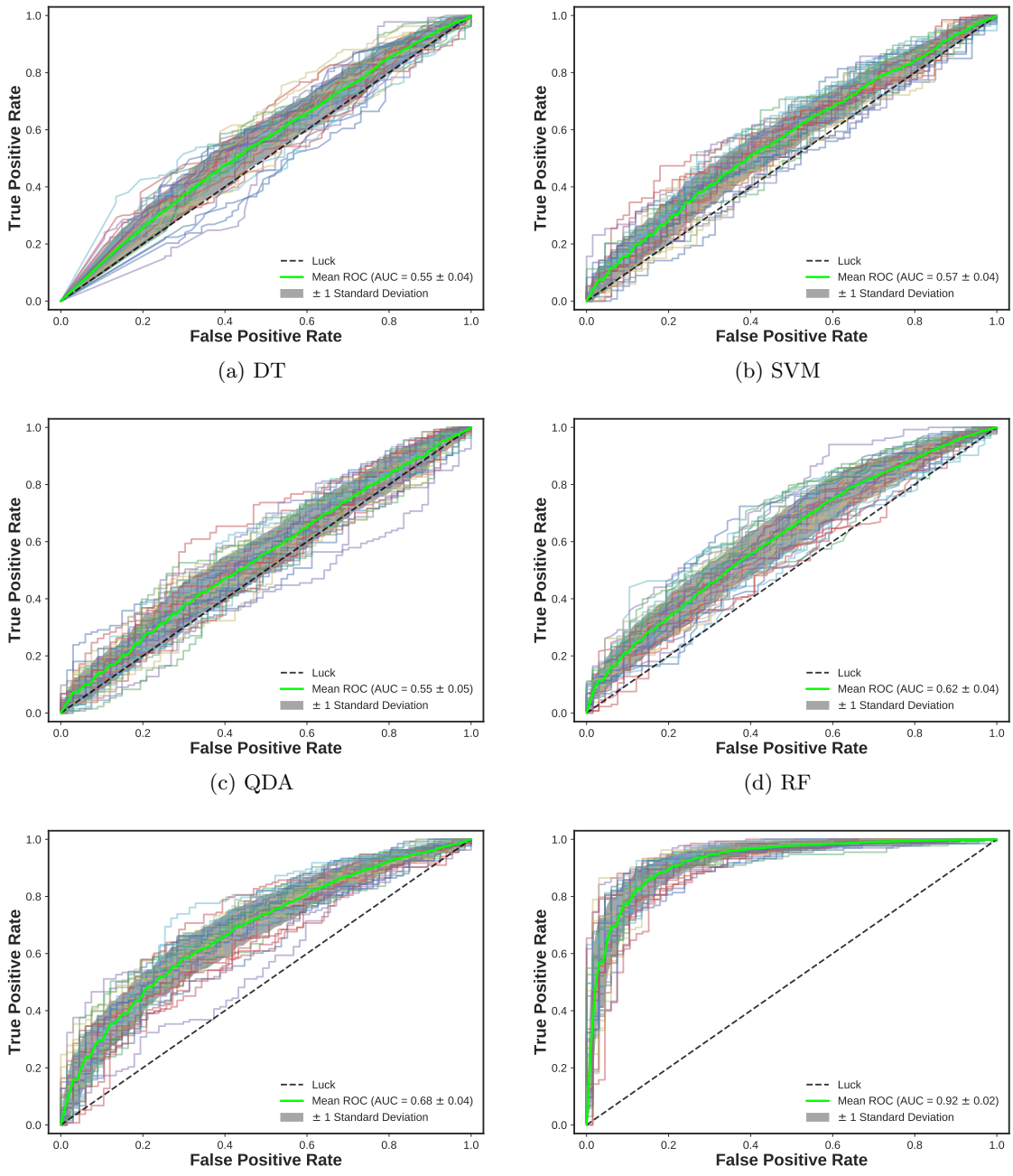


(e) AdaBoost



(f) XGBoost

Figure 2.36: Violin plots of leave-one-out cross-validation for four different classification metrics using several classification algorithms to predict relapse in heavy smokers.



(a) DT (b) SVM

(c) QDA (d) RF

(e) AdaBoost (f) XGBoost

Figure 2.37: ROC curves of various classification algorithms using leave-one-out cross-validation to predict relapse in heavy smokers.

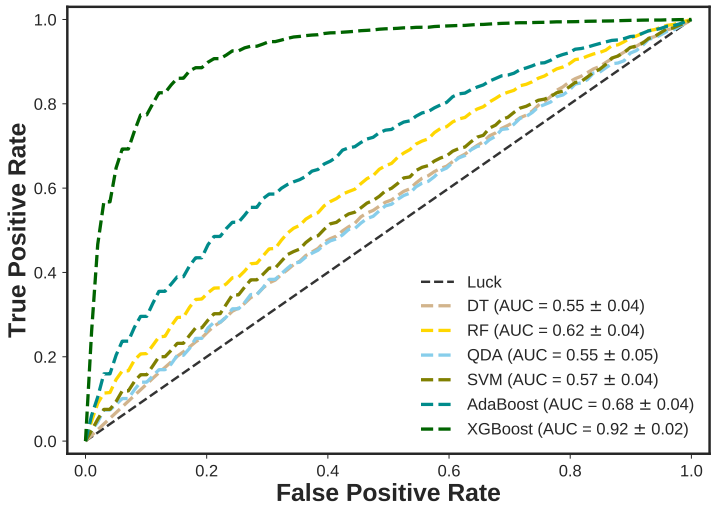


Figure 2.38: Mean ROC curves of leave-one-out cross-validation using several classification methods including decision tree (DT), support vector machine (SVM) with radial basis function (RBF) kernel, quadratic discriminant analysis (QDA), random forest (RF), AdaBoost, and XGBoost to predict relapse and non-relapse smokers based on features extracted from autoencoder.

show better accuracy. In contrast, as the ROC curve gets closer to the dashed black line (Luck), it indicates that the predictions are more stochastic and cannot be generalized. Furthermore, a comparison of the mean ROC curves of the employed machine learning algorithms is presented in Figure 2.38. This can be a clear illustration of the power of the XGBoost in prediction of complex features [106].

Additionally, the salient extracted features were backtracked exhaustively. This would clear out what part of the brain and on what snapshot was involved in the classification task. This would suggest to make new biomarkers according to the specific part of the brain to reduce relapse in heavy smokers. Figure 2.39 shows the extracted features from a subject from the non-relapse class were mapped on the subject's brain template. The highest intensity that is indicated in red was seen close to the mesolimbic system which is a collection of dopaminergic neurons that regulate incentive salience, motivation, reinforcement learning, and fear, among other cognitive processes.

It was previously shown that the dys-regulation of the mesolimbic pathway and its output neurons in the nucleus accumbens plays a significant role in the development and maintenance of

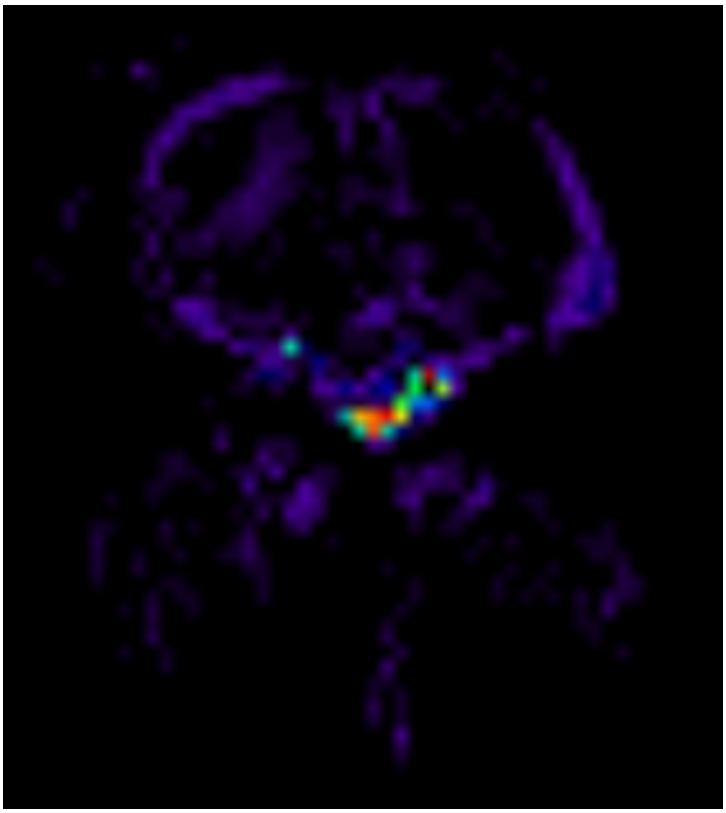


Figure 2.39: The mapped extracted features by the developed autoencoder from a subject from the non-relapse class.

an addiction. Drugs of abuse modulate gene expression, and produce their rewarding effects of euphoria or pleasure through an interaction with the mesolimbic dopaminergic system, leading to persistent alterations (neuroplastic, structural and functional) in the reward-related and memory-related brain centers [107, 108, 109, 110]. The proposed results are in agreement with the previously published results [27, 102, 103].

CHAPTER 3

BREAST: MULTI-PARAMETRIC MRI FOR NEO-ADJUVANT CHEMOTHERAPY

3.1 Background & Previous Works

In recent years, neo-adjuvant chemotherapy is widely used in patients with locally advanced breast cancer (LABC) offering several advantages such as a reduction in the tumor and enabling breast-conservation surgery instead of mastectomy as well as response-guided neo-adjuvant chemotherapy approaches [111, 112, 113, 114, 115, 116, 117]. In patients undergoing neo-adjuvant chemotherapy for breast cancer, the achievement of a pathological complete response (pCR) is associated with an increase in being significantly improved, disease-free, and overall survival [118, 119, 120, 121]. The most common definition of pCR can be the abscense of invasive disease in the breast and auxiliary lymph nodes [122, 121]. However, a pCR is achieved in only 30% of the patients after the completion of neo-adjuvant chemotherapy, and clinical studies have shown that the therapeutic outcome can be improved after the treatment modifications during the neo-adjuvant chemotherapy. Therefore, accurate means to predict treatment response as early as possible are desirable to identify women who do not benefit from cytotoxic therapy. Several studies have demonstrated that dynamic contrast-enhanced (DCE) MRI is the most sensitive method for the assessment and prediction of treatment response [113, 123, 124]. Additionally, it has been demonstrated that multi-parametric MRI (mpMRI) using morphological as well as additional functional parameters such as diffusion-weighted imaging (DWI) has potential for an improved prediction of treatment response.

In the past decade, the field of medical image analysis has grown exponentially, with an increased numbers of pattern recognition [71] tools and an increase in data set sizes. These advances have facilitated the development of processes for high-throughput extraction of quantitative features that result in the conversion of images into meaningful data, and the subsequent analysis of these data for decision support. This emerging field in medical research is termed radiomics [125, 126]. O'Flynn et al. [127] have recently shown that machine learning algorithms such as linear discriminant analysis along with statistical methods based on multi-parametric MRI features

such as enhancement fraction, tumor volume, initial area under the gadolinium curve, and also maco-kinetic parameters such as K_{trans} and K_{ep} can be employed to predict patients in terms of responders and non-responders to neo-adjuvant chemotherapy. However, it should be noted that they have employed only seven features along with one machine learning algorithm based on thirty-two patients. The summary of the recent literature review about the response after completion of neo-adjuvant chemotherapy is presented including medical statistics, sensitivity (true positive rate), specificity (true negative rate), positive predictive value (PPV), negative predictive value (NPV), and area under ROC curve (AUC) in Table 3.1. The details of the classification accuracy metrics is also presented in Figure 3.1.

Table 3.1: Assessing response after completion of neo-adjuvant chemotherapy with DCE-MRI.

Author	Number of Subjects	Sensitivity	Specificity	PPV	NPV	AUC
De Los Santos et al. 2013 [128]	746	83%	47%	47%	74%	74%
Hayashi et al. 2013 [129]	260	44%	90%	73%	73%	78%
Ko et al. 2013 [130]	166	96%	65%	-	-	89%
Fu et al. 2014 [131]	64	70%	89%	88%	71%	79%
Hylton et al. 2012 [132]	216	-	-	-	-	84%

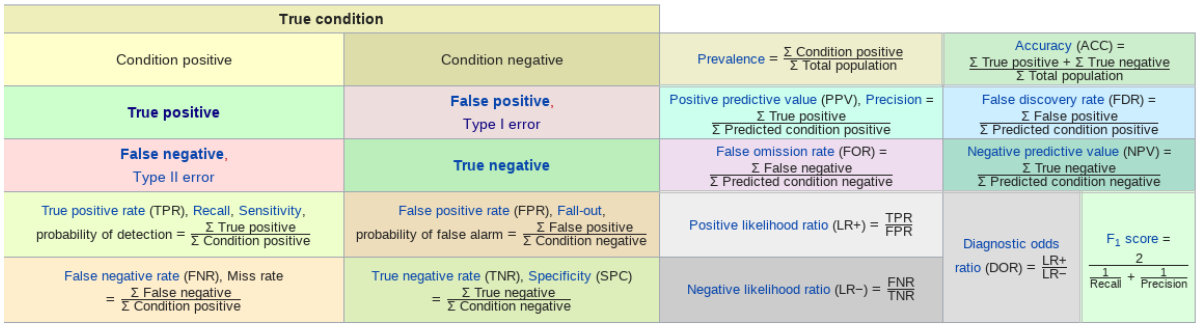


Figure 3.1: Details of all score metrics for classification problems.[5]

The aim of this study was to assess radiomics along with machine learning methods using multi-parametric MRI using, T_2-weighted, DCE with pharmaco-kinetic modeling MRI, apparent diffusion

[5]https://en.wikipedia.org/wiki/Receiver_operating_characteristic

coefficient (ADC) with DWI for the early prediction of pCR in breast cancer patients undergoing neo-adjuvant chemotherapy [120].

3.2 Data Acquisition

A prospectively populated study data base was searched for patients with newly diagnosed histopathologically proven breast cancer during the years 2009-2015, and who fulfilled the following inclusion criteria: treatment with neo-adjuvant cytotoxic systemic therapy, baseline multi-parametric magnetic resonance imaging (mpMRI) with T_2-weighted and dynamic contrast-enhanced (DCE) MRI fourteen days prior to initiation of therapy and early response assessment with mpMRI prior to and after two cycles of neo-adjuvant cytotoxic systemic therapy. Forty-one eligible patients (age range, 25-80 years; mean age, 51 years) were identified. Electronic medical records were reviewed and the following patient characteristics were recorded for each patient: age at therapy; type of therapy; start date of systemic therapy; histological type; tumor grade; receptor status; tumor proliferation rate (ki67), nodal status, date of progression (local recurrence, distant metastases); date of death; or date of last follow-up. If death was caused by breast cancer, this was also recorded.

All patients underwent mpMRI of the breast in a prone position using a 3.0 Tesla MRI unit (Trio Tim; Siemens Medical Solutions, Erlangen, Germany) with a dedicated four-channel breast coil (In Vivo, Orlando, FL, USA) and the following protocol before and during neo-adjuvant chemotherapy:

- A T_2-weighted turbo spin echo sequence with fat suppression: time of repetition (TR)/time of echo (TE) $4800/59msec$; field of view (FOV) $340mm$; 44 slices at $4mm$; flip angle 120 degree; matrix 384×512; and acquisition time (TA) $2:35min$.

- For DCE-MRI until 12/2011 a hybrid DCE-MRI protocol was used with the following sequences: T_1-weighted Volume-Interpolated-Breathhold-Examination sequences (TR/TE $3.62/1.4msec$; FOV $320mm$; 72 slices; $1.7mm$ isotropic; matrix 192×192; one average; TA $13.2sec$ per volume, 37 measurements) and T_1-weighted turbo fast-low-angle-shot-3D sequences with selective water-excitation (TR/TE $877/3.82msec$; FOV $320mm$; 96 slices; $1mm$ isotropic; matrix 320×134; one average; TA $2min$) with a total time of acquisition of $9:20min$ [133]. From 01/2012 onwards a transversal T_1-weighted time-resolved angiography with stochastic trajectories (TWIST) was acquired water excitation fat-saturation; TR/TE $6.23msec/2.95msec$; flip angle 15 degree, FOV $196 \times 330\ mm^2$; 144 slices; spatial resolution $0.9 \times 0.9 \times 1\ mm^3$; temporal interpolation factor 2; temporal resolution $14sec$; matrix 384×384; one average; center k-space region with a re-sampling rate of 23%; reacquisition density of peripheral k-space 20%; and TA $6:49min$.

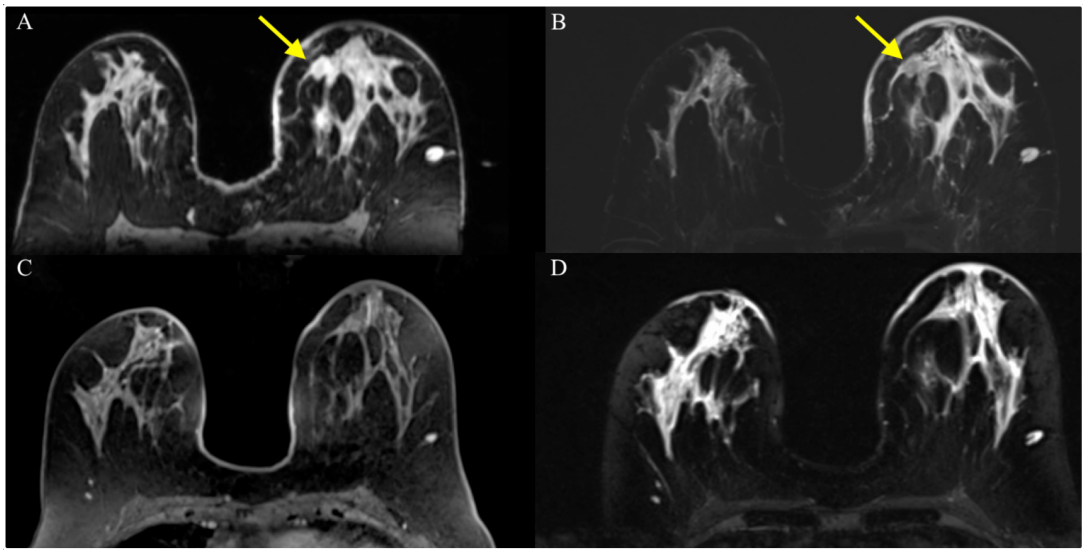


Figure 3.2: An illustration of complete imaging and pathological response after two cycles of neo-adjuvant chemotherapy.

A standard dose ($0.1 mmol/kg$ body-weight) of Gadotaremeglumine (Gd-DOTA; Dotarem, Guerbet, France) was injected intravenously as a bolus at $4 ml/sec$ with a saline flush after injection. Total MRI examination time was approximately $10-12$ minutes. Figure 3.2 illustrates a breast scan of a 65 years old patient with multi-centric breast cancer and index lesion left breast retroareolar medial (invasive ductal carcinoma grade 3, triple negative molecular subtype: estrogen receptor/ progesterone receptor/ human epidermal growth factor receptor 2 negative, ki67 90%): complete imaging and pathological response after two cycles of neo-adjuvant chemotherapy.

3.3 Feature Extraction

3.3.1 Initial Trial

MpMRI data was evaluated by an experienced breast radiologist (K. P.; 12 years of experience) and a resident in consensus. The below detailed image evaluation of multi-parametric MRI was repeated for the follow-up examination. For all lesions, size and location as well as the largest diameter on DCE-MRI was recorded. Signal intensity on T_2-weighted sequences (hypointense, isointense, and hyperintense) and the presence or absence of a peri-tumoral edema was noted. In

DCE-MRI tumors were classified as mass or non-mass enhancing (NME) lesions. According to the 5th edition of the American College of Radiology (ACR) and Breast Imaging Reporting and Data System (BI-RADS) (2) the following descriptors were assessed for masses: shape (round, oval, and irregular), margins (circumscribed, irregular, and spiculated), and internal enhancement characteristics (homogeneous, heterogeneous, rim enhancement, and dark internal septations). For NME the distribution (focal, linear, regional, segmental, multiple, and diffuse), internal enhancement pattern (homogeneous, heterogeneous, clumped, and clustered ring) and symmetry (symmetric and asymmetric) were evaluated. For pharmaco-kinetic (PK) assessment of DCE-MRI the mean plasma flow (PF), the volume distribution (VD), and the mean transit time (MTT) were assessed with parametric maps using a 3D-based region of interest (ROI) segmentation approach (UMM-perfusion tool, University of Heidelberg, OsiriX Imaging Software version 7.0).

Table 3.2: Features extracted from mpMRI using morphological, and functional imaging.

1. AP Diameter (mm)	8. Mass Internal EH
2. RL Diameter (mm)	9. MTT (Sec)
3. CC Diameter (mm)	10. DCE Plasma Flow (ml/min)
4. T_2 Signal Intensity	11. Distribution Volume (ml/100ml)
5. T_2 Peritumoral Edema	12. ADC (Min)
6. Mass Shape	13. ADC (Max)
7. Mass Margins	14. ADC (Mean)

For each lesion a total number of 14 features were extracted ranging from morphological, quantitative kinetic, and ADC parameters. Table 3.2 presents the list of the features extracted in this study from the mpMRI including: T_2-weighted: hyper-, hypo-, and isointense, presence of peritumoral edema, size: diameters in all three planes in millimeters, mass shape in terms of round, oval, and irregular, mass margins: circumscribed, irregular, and spiculated: mass internal enhancement characteristics: homogeneous, heterogeneous, rim enhancement, dark internal spatations: quantitative enhancement kinetics: mean transit time (MTT) in seconds, plasma flow (ml/min), distribution volume (ml/100 mls), and ADC min, max, and mean. Figure 3.3 depicts the correlation matrix of the extracted features with color bar which highlights the probability of the correlation of each

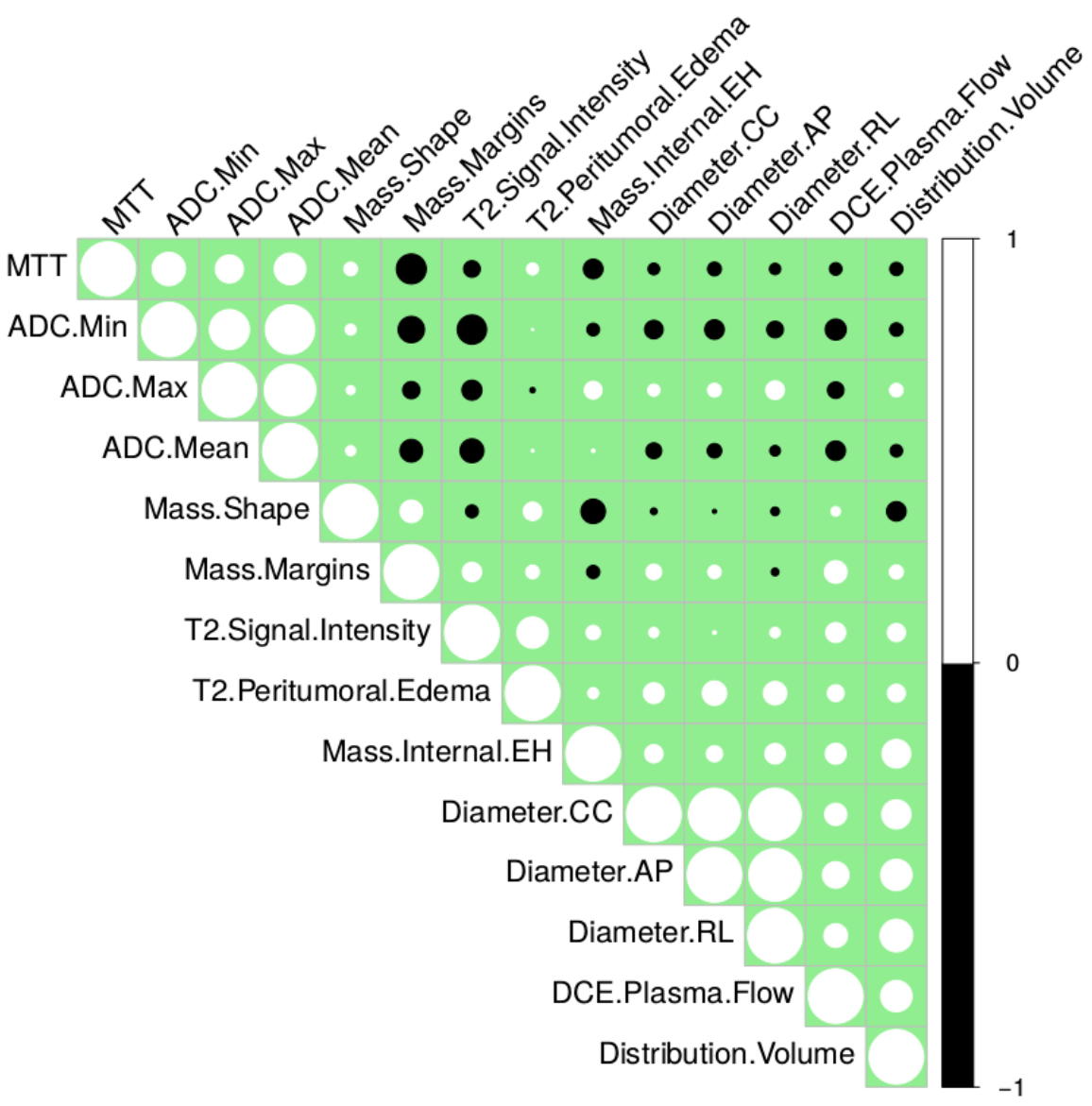


Figure 3.3: Correlation matrix plot of the extracted features with hierarchical clustering.

of the extracted features with each other. It is clear that the diagonal has the probability of one (white color). In this plot, to illustrate the correlation matrix the hierarchal clustering method was employed. Positive correlations are displayed in white and negative correlations in black color. The

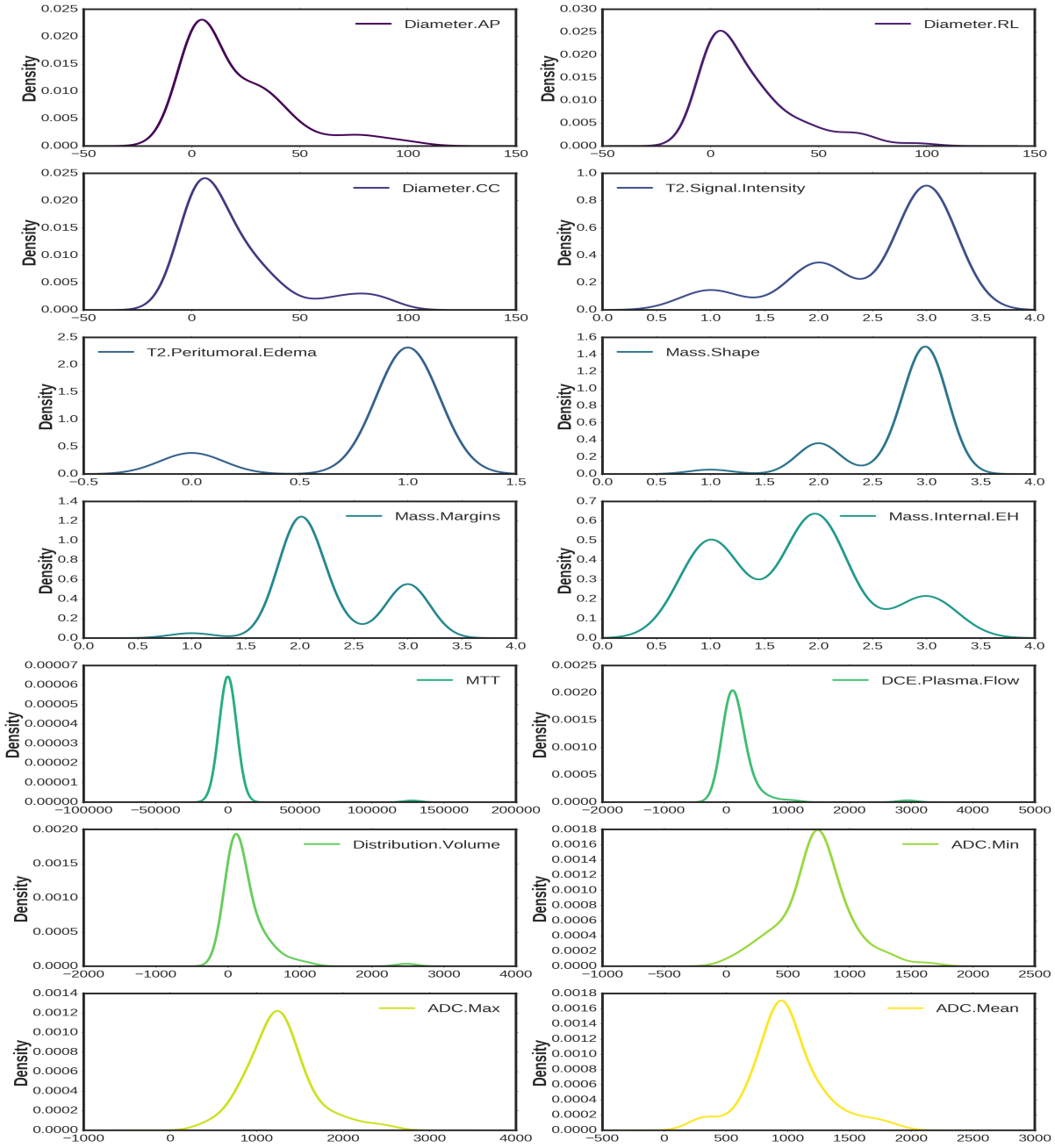


Figure 3.4: The kernel density function for each of the extracted features.

size of the circles are proportional to the correlation coefficients. Thus, as the circle gets progressively larger this indicated the features are more correlated which in turn can be both positive or

negative (black or white). For example, the mean value of ADC is correlated with the minimum and maximum values of ADC which was expected. The other point which was expected is the correlation of the distribution volume with the diameters AP, RL, and CC. In addition to this, Figure 3.4 illustrates the kernel density function of each of the extracted features. It should be noted that, for the initial feature extraction trial, pre-treatment and post-treatment scans of each subject were determined as two different subject to increase the number of subjects in classification. However, the change in condition of the patients can be overlooked. Thus, to see the improvement of the tumor for each patient, the change for each feature was considered as the final trial.

3.3.2 Final Trial

As discussed, to consider the change of the tumors due to the neo-adjuvant chemotherapy, the final feature extraction trial was carried out. In this regard, out of 41 patients only 38 patients who had complete pCR were considered and the total 27 features including sub-molecular types (Her2+, TN, Luminal A, and Luminal B), DCE plasma flow, DCE distribution volume, MRI MTT in seconds, MRI ADC (Min, Max, and Mean), MRI diameter size (AP, RL, and CC), pattern of shrinkage (concentric, fragmentation, and complete), mass internal enhancement (homogeneous, heterogeneous, and rim), mass margins (irregular, and spiculated), mass shape (oval, and irregular), MRI edema, and MRI STIR (hypo, hyper, and iso) were extracted. For the continuous features the difference between pre-treatment and post-treatment scans were considered and for the categorical features one-hot-encoder was employed. Figure 3.5 shows the correlation matrix plot of the extracted features with hierarchical clustering.

3.4 Machine Learning

In addition to support vector machines as discussed in machine learning section in chapter 2, we have employed logistic regression, linear discriminant analysis, stochastic gradient descent, and random forests.

3.4.1 Linear Discriminant Analysis (LDA)

LDA [134] is a method based on generalization of the Fisher's linear discriminant [135] to find a linear combination of attributes that separates two or more classes by determining a subspace of lower dimension of the original data and classification as well. Statistical measures, such as

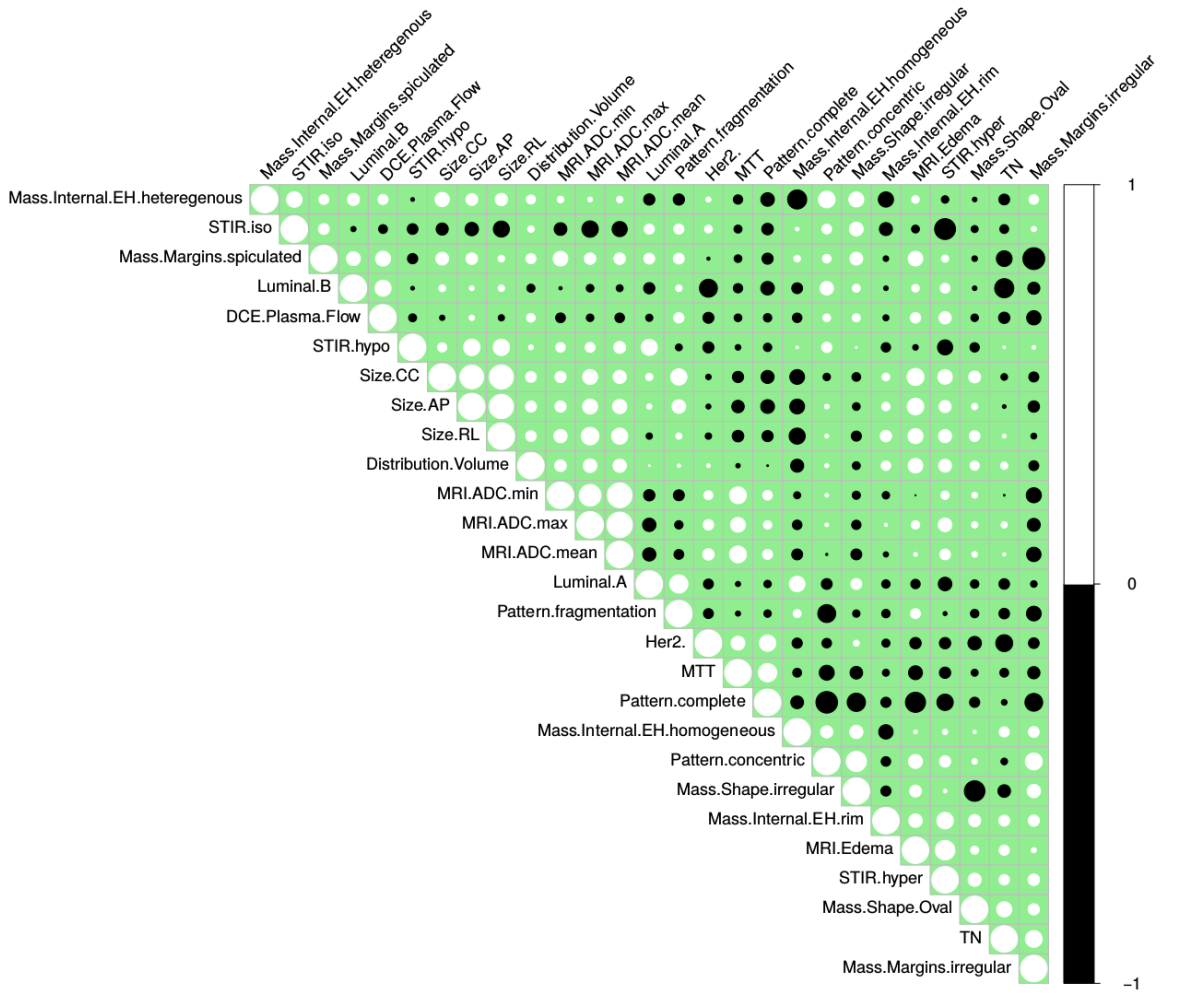


Figure 3.5: Correlation matrix plot of the extracted features with hierarchical clustering.

variance and mean, are used to determine separability. Figure 3.6 illustrates the LDA model for a binary class problem. As shown in Figure 3.6 , LDA finds the best Gaussian distribution with their mean and covariance parameters (μ, Σ) for each class with maximum margin between each class. LDA can also be derived from Bayesian rule by assigning a pattern with the maximal probability by comparing the posterior probability of all classes. In other words, it is desired to maximize the projected class means with minimization of the classes variance in that direction by fitting a Gaussian density to each class with the assumption that all classes share the same covariance matrix

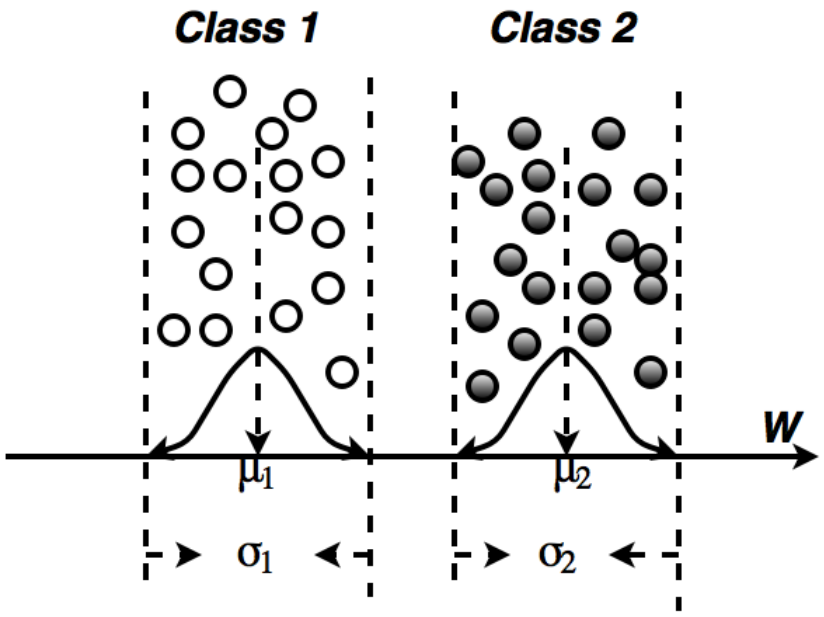


Figure 3.6: Schematic illustration of linear discriminant analysis for a two-class problem.

[134]. For probabilistic modeling for class, maximize the conditional probability as following:

$$p(y=k|x) = \frac{p(x|y=k)p(y=k)}{p(x)} = \frac{p(x|y=k)p(y=k)}{\sum_l p(x|y=l)p(y=l)} \tag{3.1}$$

For the classification task, we predict $p(y=k)$from the training data for class k with means μ_k and the covariance matrices Σ_k. Assuming the same covariance matrix for Gaussian, the log-probability ratios of the class k and class l ($log(\frac{p(y=k|x)}{p(y=l|x)}) = 0$) is:

$$(\mu_k - \mu_l)\Sigma^{-1}x = \frac{1}{2}(\mu_k^T\Sigma^{-1}\mu_k - \mu_l^T\Sigma^{-1}\mu_l) \tag{3.2}$$

where (μ, Σ) are the mean and covariance parameters. For a binary class problem, LDA deals with $p(x|y=0)$ and $p(x|y=1)$ with normal distribution parameters (μ_0, Σ_0) and (μ_1, Σ_1) [136].

LDA as a supervised model is robust against noise, but prone to over-fitting ("memorizing the training cases") in the classification task which might reduce classification accuracy on previously unseen cases. Another problem is under-fitting which usually happens where the number of records in the training dataset is small compared to the number of features. In this regard, shrinkage has been used in LDA model as a tool to improve the estimation of the covariance matrix [137, 101].

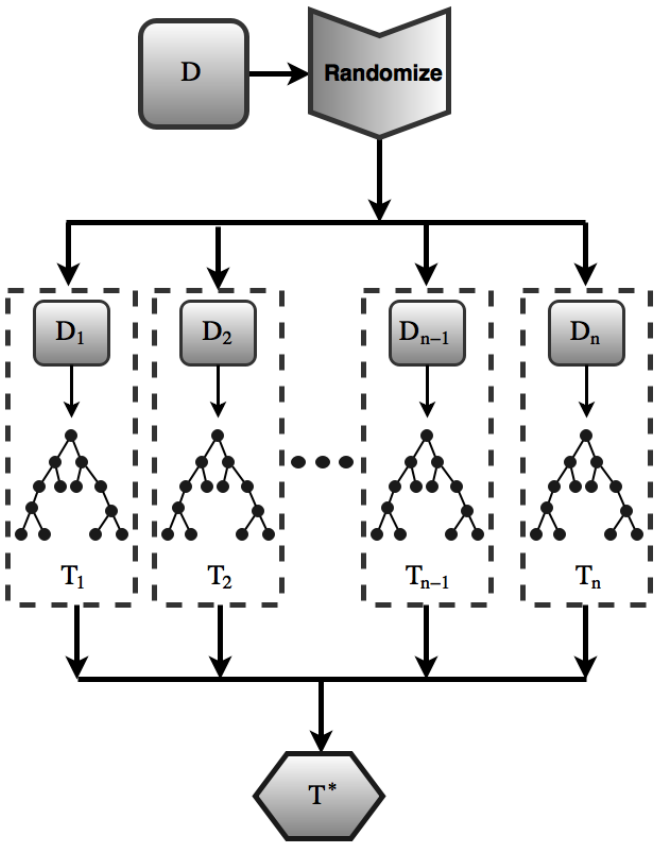


Figure 3.7: Schematic illustration of random forests.

Over-fitting is less of an issue in logistic regression model due to the low complexity. For big data problems with more variables, data reduction methods such as principal component analysis (PCA) or independent component analysis (ICA), and feature selection schemes would be useful to overcome over-fitting. Employing several methods such as forward selection, backward selection, and stepwise selection would help to test the statistical significance of the coefficients in the logistic model [138]. In general, the p-value threshold is set to 0.05 for statistical testing, however it could be modified based on the problem.

3.4.2 Logistic Regression (LR)

A logistic function, is an S shape function which combines two characteristic kinds of exponential growth, (1) the familiar pattern of increase at an increasing rate, and (2) bounded exponential growth which means as the decaying exponential dies out, the difference rises up to the bound. For a binary class, the odds based on the values of the independent variables can be modeled as the conditional probability using the logistic function:

$$p(y = \pm 1|x) = \frac{1}{1 + exp(-yw^T x)} \tag{3.3}$$

where x is the input data, y is the class label, and w is the weight vector. The logistic regression minimizes the negative log-likelihood of conditional probability via optimization algorithms. It is designed to find cumulative logistic distribution by measuring the relationship between one dependent and one or more independent variable(s). A major challenge to implementing the logistic regression method is to determine the values for the weights (w) [139]. For binary classes with small training data sets, LIBLINEAR [140] solver along with L_1 as a norm in penalization would be a great choice to determine the weights. On the other hand, stochastic average gradient (SAG) [141] optimization algorithm along with L_2 norm could handle large data sets. For multi-class cases, optimization algorithms such as NEWTON-CG, and LBFGS [142] should be used along with L_2 norm.

3.4.3 Stochastic Gradient Descent (SGD)

Gradient descent has been proposed by Rumelhart [143, 144] as a method for unconstrained optimization problems. For a given binary labeled $y_i \in \} - 1, 1\}$ training set $x_i \in R^n$, SGD tries to learn a linear scoring function such as $f(x) = w^T x + b$, where $w \in R^m$ is weight parameters and $b \in R$ is the intercept. The parameters can be found by minimizing a loss function. In machine learning, loss functions for classification represent the price paid for inaccuracy of prediction. Using SGD gives us more degrees of freedom to choose different loss functions [145]. SGD supports both binary and multi-class classification problems. It also should be noted that it is recommended to use SGD in large-scale and sparse data sets. SGD has been successfully applied to big data problem with more than 10^5 training instances containing more than 10^5 attributes. It is highly efficient due to the opportunities of code tuning in parameters and choice of loss functions [101]. There are

Algorithm 3: Recursive feature elimination incorporating k-Folds cross-validation

Input: Data D, Number of folds k

Output: Optimum Number of Features ONF , Variable Ranking R, Cross-Validation Accuracy $\bar{Acc}$

Partition the data D into k subsets $\{D_1, \ldots, D_k\}$;
for $i \in \{1, \ldots, k\}$ **do**
    $S_i \leftarrow D - D_i$;
    Train the model with S_i data;
    Test the model on D_i data;
    Calculate the classification accuracy Acc_i for subset D_i;
    Calculate the variable importance R_i;
    Determine the optimum number of features ONF;
end
$\bar{Acc} \leftarrow \frac{\sum_{i=1}^{k} Acc_i}{k}$ Cross-Validation Accuracy;
Determine the optimum list of features incorporating their ranking;

some disadvantages regarding using SGD in machine learning such as, being sensitive to attribute scaling. It is also time consuming since the parameters need to be regularized during iterations.

3.4.4 Random Forests (RF)

In 2001, Leo Breiman [146] presented the new concept of decision trees called the random forests with employing hierarchical and sequential nodes to illustrate the patterns and structure of the input data. The patterns would finally represent the data in terms of traditional decision trees [147]. A prediction by random forests consists of the combination of the results from several decision trees. In this way, random forest method is robust to noisy data and can improve the classification accuracy in comparison to decision trees as an ensemble method. The ensemble structure can be also translated in terms of being robust to over-fitting. However, over-fitting is so common in traditional tree based methods such as CART [70]. Figure 3.7 demonstrates schematic illustration of random forests. The stochastic construction of the trees based on bootstrapped sample is the main basis of the random forest method. The growing process of the trees is based on the stochastic

feature selection, best split feature selection, and node splitting [147]. The results for each tree would be combined based on the majority votes for the classification task.

3.4.5 Recursive Feature Elimination (RFE)

In this study, recursive feature elimination method as presented in Algorithm 3 along with classifiers was employed to find the optimum ranking of the features extracted from mpMRI. By employing recursive feature elimination, we can select features by recursively considering smaller and smaller sets of features by training a classifier on the initial set of features, and weights [101]. Then, features whose absolute weights are the minimum are pruned from the current set features. By repeating this procedure the desired optimum number of features with the maximum accuracy would be found. It should be noted that, the number of individuals in each class was not equal. In better words, we were dealing with biased classification tasks. In this regard, 10-folds cross-validations was employed to decrease the possibility of over-fitting the models. For each classifier, the optimum number of features (ONF) is also reported. It should be noted that all the computing codes, and analysis were written in `R`, and `Python` programming languages [101].

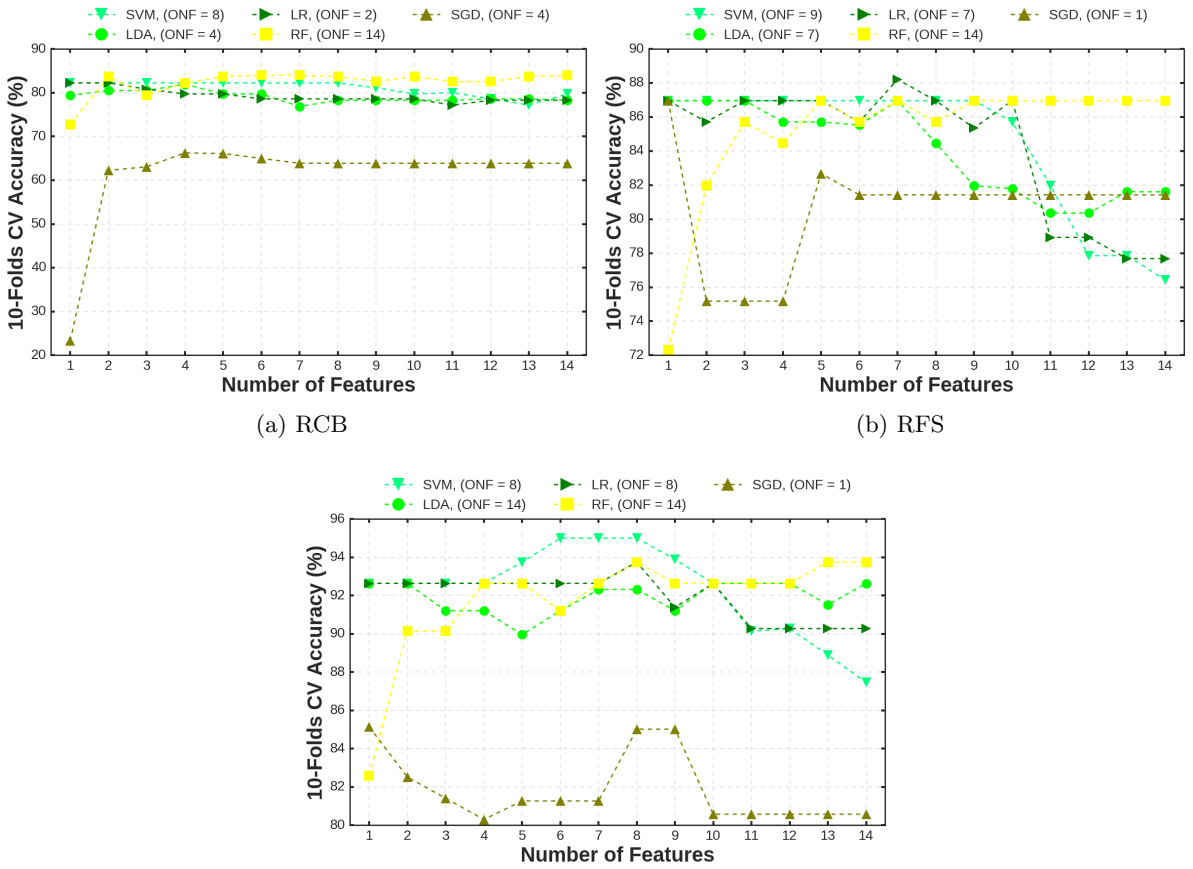
3.5 Results & Discussion

3.5.1 Initial Feature Extraction

As discussed in Algorithm 3, we have employed recursive feature elimination along with 10-folds cross-validation to rank the extracted features based on their importance for each of the machine learning methods. Figure 3.8 shows the 10-folds cross-validation accuracy for each of the machine learning algorithms including RF, LDA, linear SVM, LR, and SGD to predict RCB score, RFS, and DSS. For each of the classifier, the optimum number of features (ONF) were also presented.

The trend of improvement in accuracy is reasonable for RCB prediction as shown in Figure 3.8a. All the classifiers except SGD showed good performance with an accuracy around 80%. However, this trend is not usual for RFS prediction as shown in Figure 3.8b. It is obvious that the trend until 6 features can be reasonable. However, the classification accuracy for some methods dropped after 6 features which proved that the classifier could not predict the test cases after certain numbers of features.

This suggests that the presence of over-fitting in the models. Similar trend was seen in the prediction of DSS as shown in Figure 3.8c. The results of the features ranking based on recursive



(a) RCB

(b) RFS

(c) DSS

Figure 3.8: Recursive feature elimination along with 10-folds cross-validation incorporating different classifiers including linear support vector machine (SVM), logistic regression (LR), random forest (RF), stochastic gradient descent (SGD), and linear discriminant analysis (LDA) to predict residual cancer burden (RCB) score, recurrence free survival (RFS), and disease-specific death (DSS).

feature elimination algorithm in prediction of RCB, RFS, and DSS are presented in the Tables 3.3, 3.4, and 3.5. In this regard, we have tested different combinations of the best five features to predict RCB score, RFS, and DSS class labels. Thus, we could come up with a feature set including the five optimum features along with a classifier to predict all the aforementioned class labels. After testing the possible combinations, the best accuracy was gained using the features set including: 1) mass internal EH, 2) mass shape, 3) mass margins, 4) T_2 peritumoral edema, and 5) T_2 signal intensity.

Table 3.3: Features ranking for linear support vector machine (SVM), logistic regression (LR), random forest (RF), stochastic gradient descent (SGD), and linear discriminant analysis (LDA) to predict residual cancer burden (RCB) score.

	SVM	LDA	LR	RF	SGD
AP Diameter	8	7	5	2	7
RL Diameter	7	6	6	7	9
CC Diameter	6	8	7	6	8
T_2 Signal Intensity	5	3	3	14	12
T_2 Peritumoral Edema	4	1	2	12	13
Mass Shape	2	5	4	13	10
Mass Margins	3	4	8	11	14
Mass Internal EH	1	2	1	10	11
MTT	14	14	14	4	1
DCE Plasma Flow	11	13	13	9	4
Distribution Volume	12	11	12	1	5
ADC (Min)	13	12	11	5	3
ADC (Max)	10	10	10	8	6
ADC (Mean)	9	9	9	3	2

The box plot presentations of the classification accuracy along with 4-folds cross-validation for all the class predictions were presented in Figure 3.9.

As seen, among all the classifiers, random forests showed a reasonable performance in prediction of RCB score, RFS, and DSS. To explore deeper the details of the hyper-parameters involved in random forests, several multi-metric evaluations incorporating 4-folds cross-validation on the chosen features were applied. This would help to find the optimized structure and hyper-parameters for random forest of each of the classes in prediction. The first hyper-parameter is the number of trees in random forest. As shown in Figure 3.10, two different metrics, classification accuracy and area under ROC curve for both training and testing set were employed to find the best number of trees in each forest to predict RCB score, RFS, and DSS classes.

Random forest achieves a lower test error solely by variance reduction. Thus, increasing the number of trees in the ensemble only decreased the variance of the forests and did not have any effect on the bias of the model [148]. In addition to this, the best score based on each metric was

Table 3.4: Features ranking for linear support vector machine (SVM), logistic regression (LR), random forest (RF), stochastic gradient descent (SGD), and linear discriminant analysis (LDA) to predict recurrence free survival (RFS).

	SVM	LDA	LR	RF	SGD
AP Diameter	5	6	7	2	7
RL Diameter	7	8	8	9	8
CC Diameter	4	7	6	6	9
T_2 Signal Intensity	2	3	1	13	14
T_2 Peritumoral Edema	12	1	2	14	12
Mass Shape	6	5	3	12	13
Mass Margins	1	2	5	11	10
Mass Internal EH	3	4	4	10	11
MTT	14	14	14	1	1
DCE Plasma Flow	11	11	9	4	4
Distribution Volume	13	12	13	8	3
ADC (Min)	8	9	11	3	2
ADC (Max)	10	13	12	7	5
ADC (Mean)	9	10	10	5	6

pointed on the curve. This would help to choose the best number of trees for each class. The shadow area for the testing sets demonstrates the standard deviation based on 4-folds cross-validation and the lines illustrates the mean value of the 4-folds for each set. Similarly, this procedure was repeated to find the minimum number of samples required to be at a leaf node (Figure 3.11). Additionally, Figure 3.12 presents the minimum number of samples required to spit an internal node. By taking all the optimized hyper-parameters into account, random forest was built to predict each of the classes individually incorporating 4-folds cross-validation.

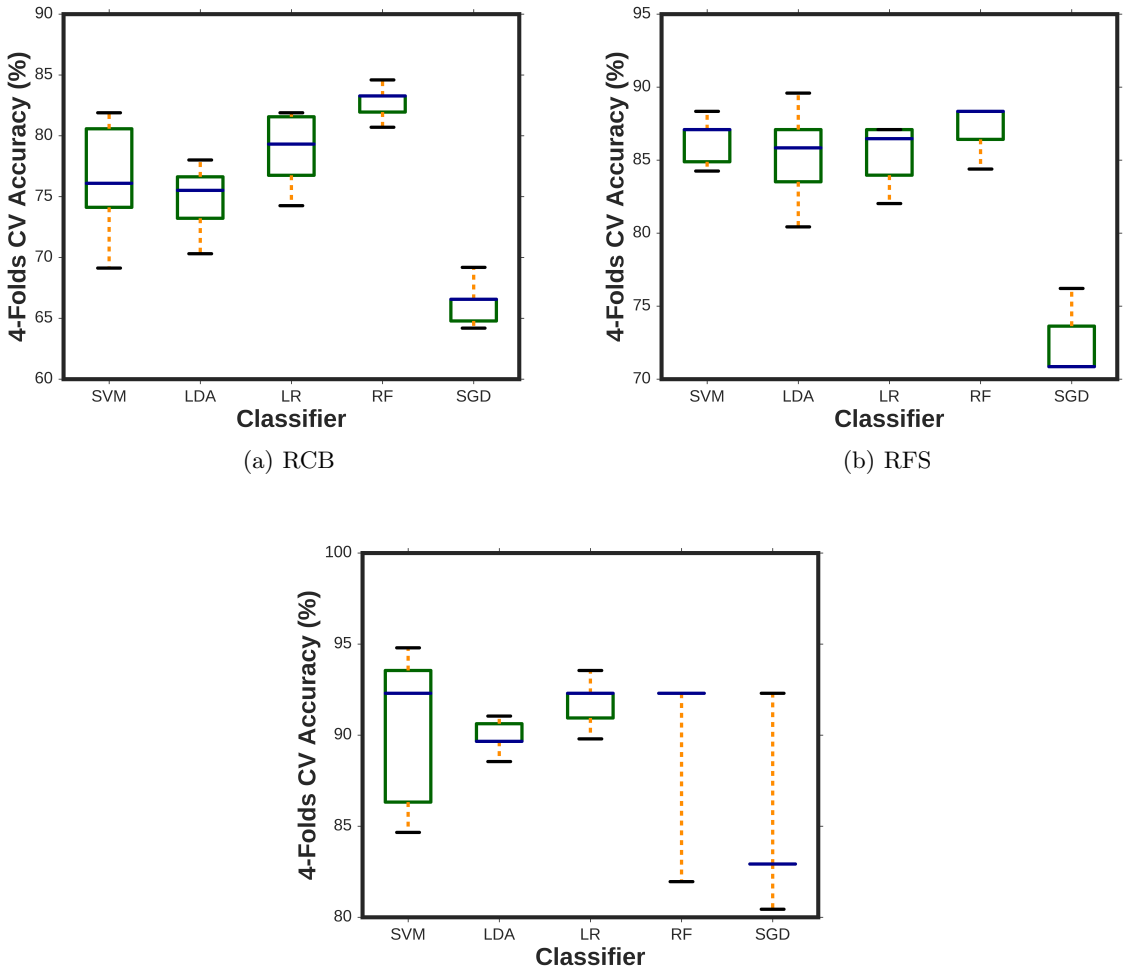
In addition to find the most accurate model using the optimized hyper-parameters, identifying which of the input variables (features) are the most import ones to make the predictions [148]. In this regard, Figure 3.13 illustrates the relative importance of the features in prediction of RCB score, RFS, and DSS. In this regard, the importance of each of the variables for predicting of the classes was calculated by adding up the weighted impurity decreases based on Gini index for all nodes where the specific variable was used, and averaged over all trees [146, 81, 148]. As shown, the

Table 3.5: Features ranking for linear support vector machine (SVM), logistic regression (LR), random forest (RF), stochastic gradient descent (SGD), and linear discriminant analysis (LDA) to predict disease-specific death (DSS).

	SVM	LDA	LR	RF	SGD
AP Diameter	5	6	8	4	7
RL Diameter	7	8	7	1	8
CC Diameter	4	7	6	7	9
T_2 Signal Intensity	2	3	3	12	13
T_2 Peritumoral Edema	12	2	4	14	14
Mass Shape	6	5	5	10	11
Mass Margins	1	4	1	13	12
Mass Internal EH	3	1	2	11	10
MTT	10	14	14	2	1
DCE Plasma Flow	13	11	13	5	4
Distribution Volume	14	10	12	9	3
ADC (Min)	8	9	9	3	2
ADC (Max)	11	12	11	8	5
ADC (Mean)	9	13	10	6	6

mass internal enhancement with four different categories including homogeneous, heterogeneous, rim enhancement, and dark internal septation had the most importance in prediction of all three classes. This would suggest a new path to radiologists to make accurate decisions before any surgery.

To explore deeper the performance of the random forests, the receiver operating characteristic (ROC) curves with area under curve (AUC) for each one of the folds along with the mean value and its standard deviation were presented in Figure 3.14. An ROC curve presents false positive rate versus true positive rate under different classification thresholds. The true positive rate is the proportion of positive cases that are correctly classified. The false positive rate is the proportion of negative cases that are incorrectly classified as positive. he performance can be evaluated through how well a method separates the true positive rate from the false positive rate. The area under the ROC curve provides a straightforward measure. An AUC of 1.0 represents a perfect test and an AUC of 0.5 represents a worthless test. The closer the AUC to 1.0, the better the test [69].



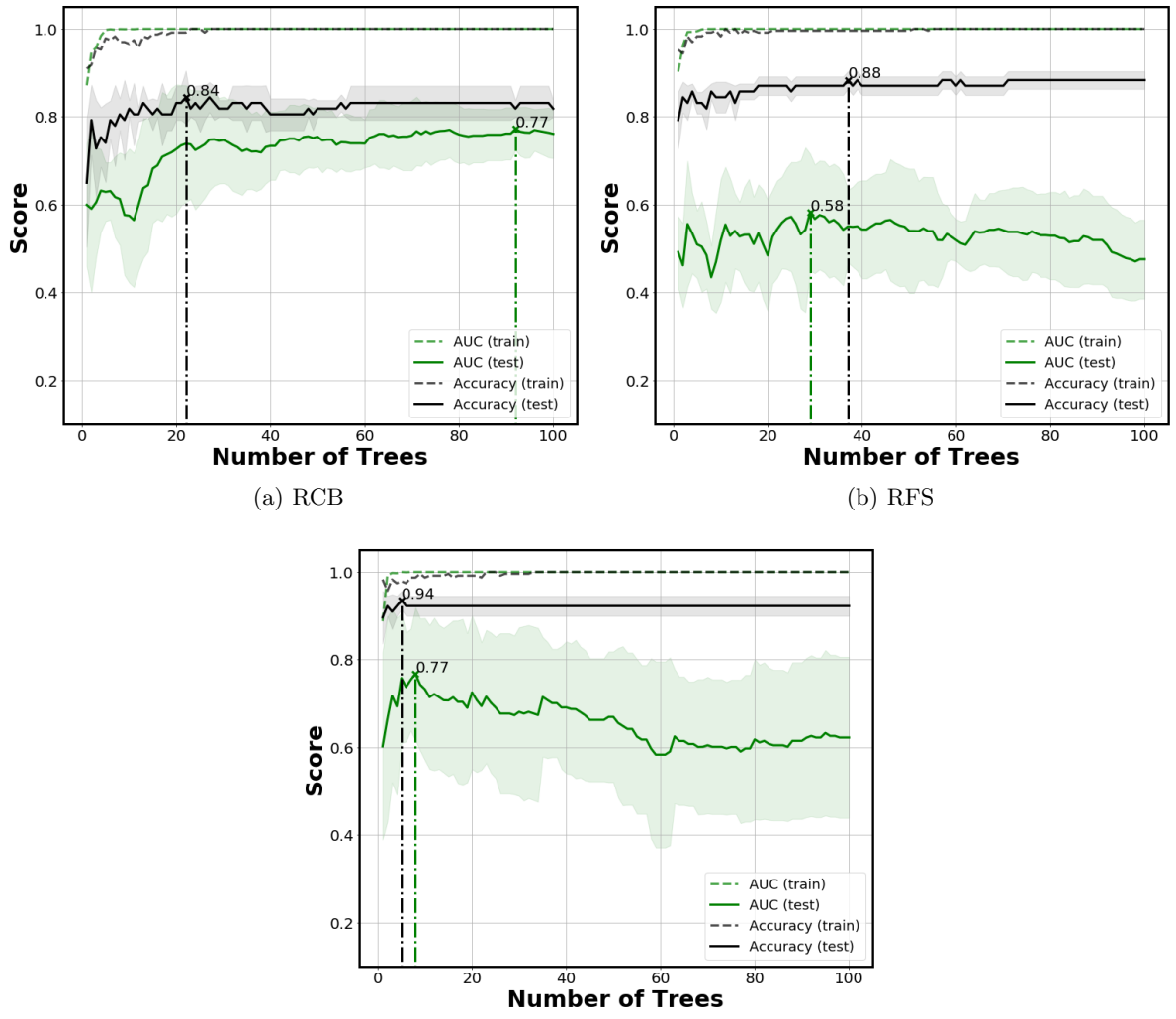
(a) RCB

(b) RFS

(c) DSS

Figure 3.9: Box plot presentations of 4-folds cross-validation accuracy incorporating different classifiers including linear support vector machine (SVM), logistic regression (LR), random forest (RF), stochastic gradient descent (SGD), and linear discriminant analysis (LDA) to predict residual cancer burden (RCB) score, recurrence free survival (RFS), and disease-specific death (DSS).

The gray area presented in the plot can be used as the confidence interval for the predictions. AS seen, clearly most of the classifications happened in this interval which prove that the classifications are valid. Based on the ROC results, random forests predicted the RCB score class with a mean AUC of (0.88 ± 0.09), the RFS class with a mean AUC of (0.85 ± 0.07), and the DSS class with



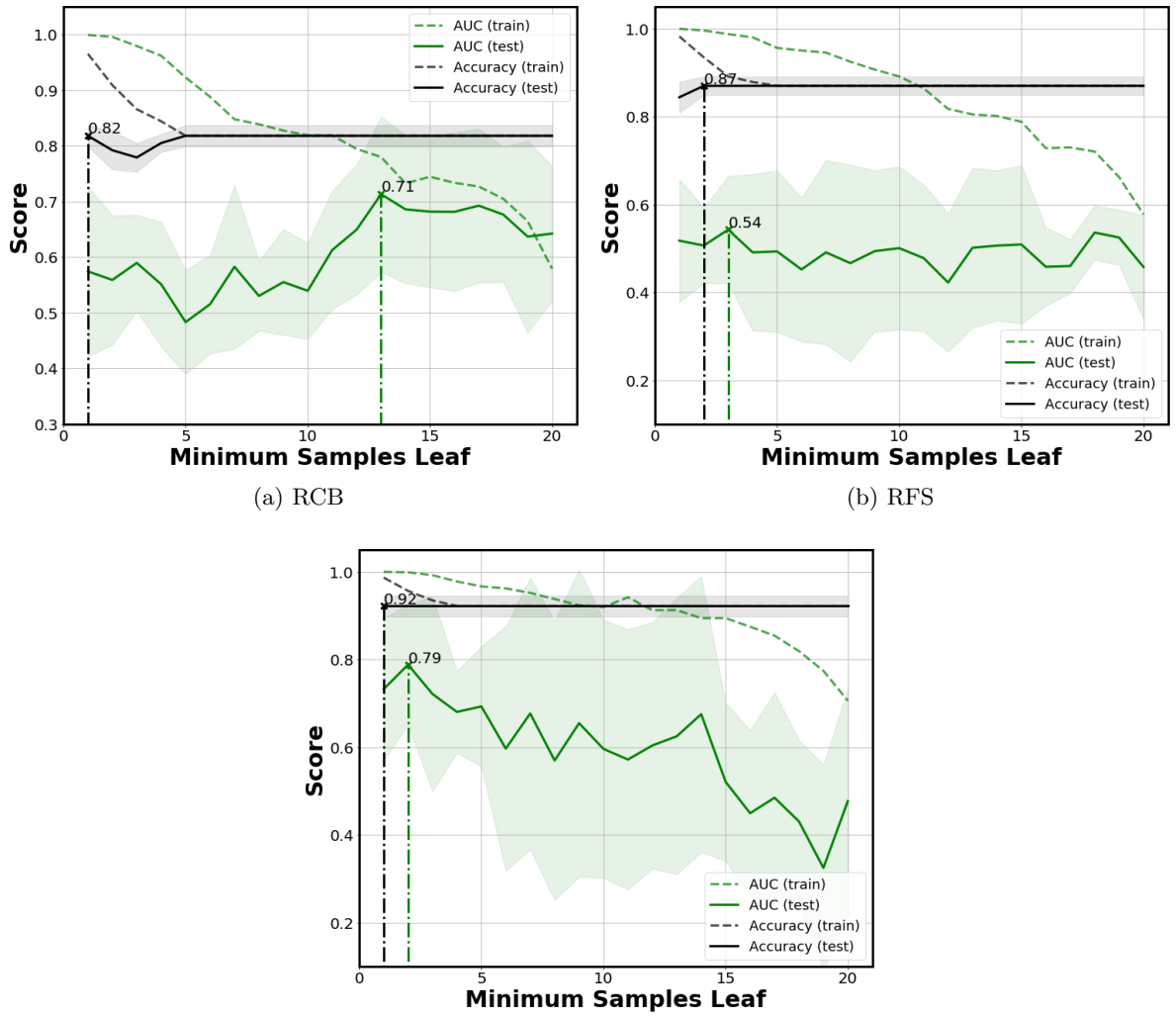
(a) RCB

(b) RFS

(c) DSS

Figure 3.10: Multi-metric evaluation of 4-folds cross-validation incorporating RF classification method based on the numbers of trees to predict RCB score, RFS, and DSS.

a mean AUC of (0.84 ± 0.02). This numbers suggest us that random forest can be a valid and robust method to be employed for multi-parametric classification of breast cancers. Finally, Figure 3.15 illustrates the decision boundaries of the random forest classifier in prediction of RCB score (Figure 3.15a), RFS (Figure 3.15b), and DSS (Figure 3.15c).



(a) RCB

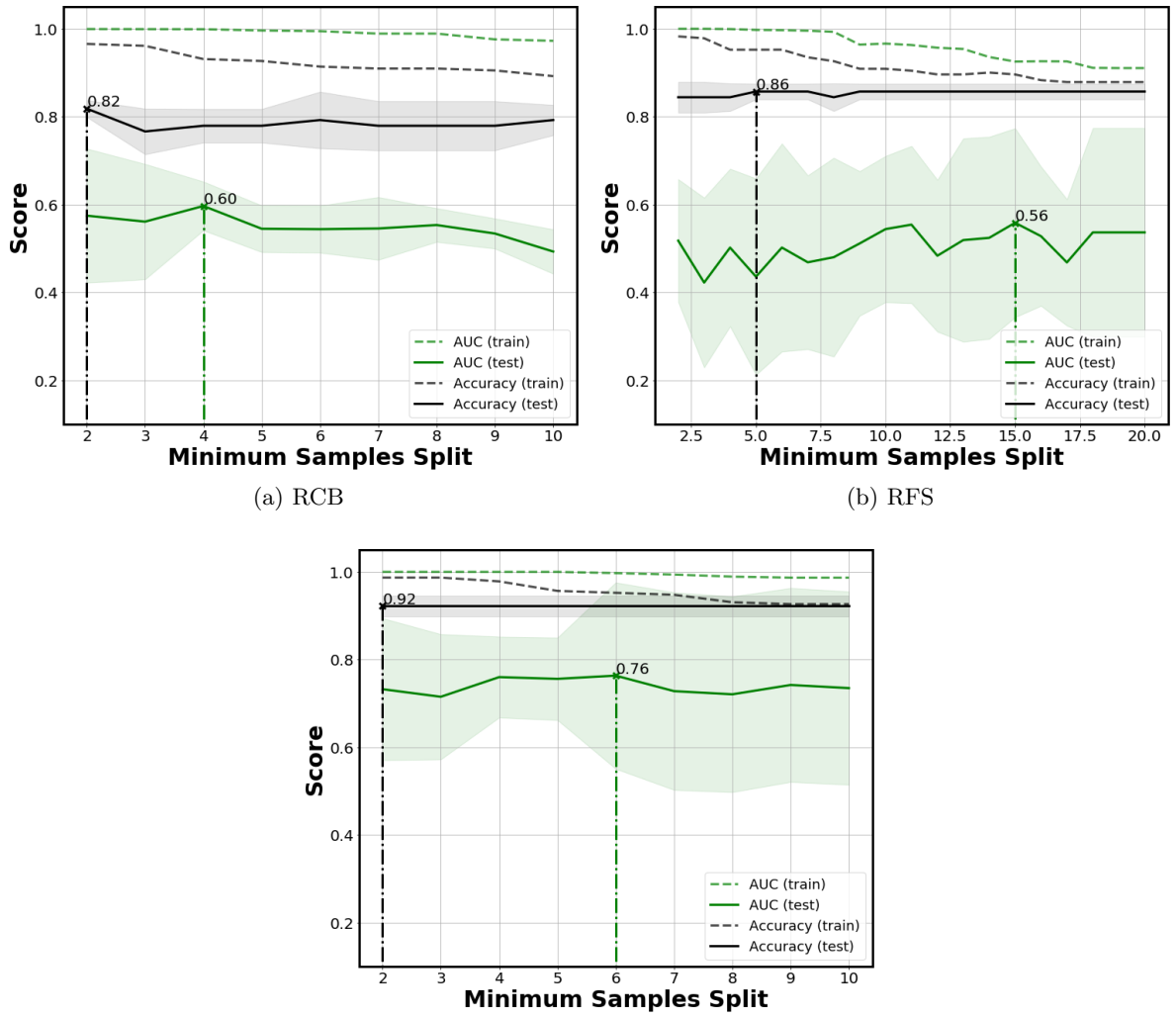
(b) RFS

(c) DSS

Figure 3.11: Multi-metric evaluation of 4-folds cross-validation incorporating RF classification method based on the minimum number of samples required to be at a leaf node to predict RCB score, RFS, and DSS.

3.5.2 Final Feature Extraction

As discussed, 27 radiomics features were extracted and fed into recursive feature elimination algorithm employing 4-folds cross-validation to over-come over-fitting. AUC was considered as the classification accuracy metric. Figure 3.16 shows the box plot presentations of the RFE accuracy



(a) RCB

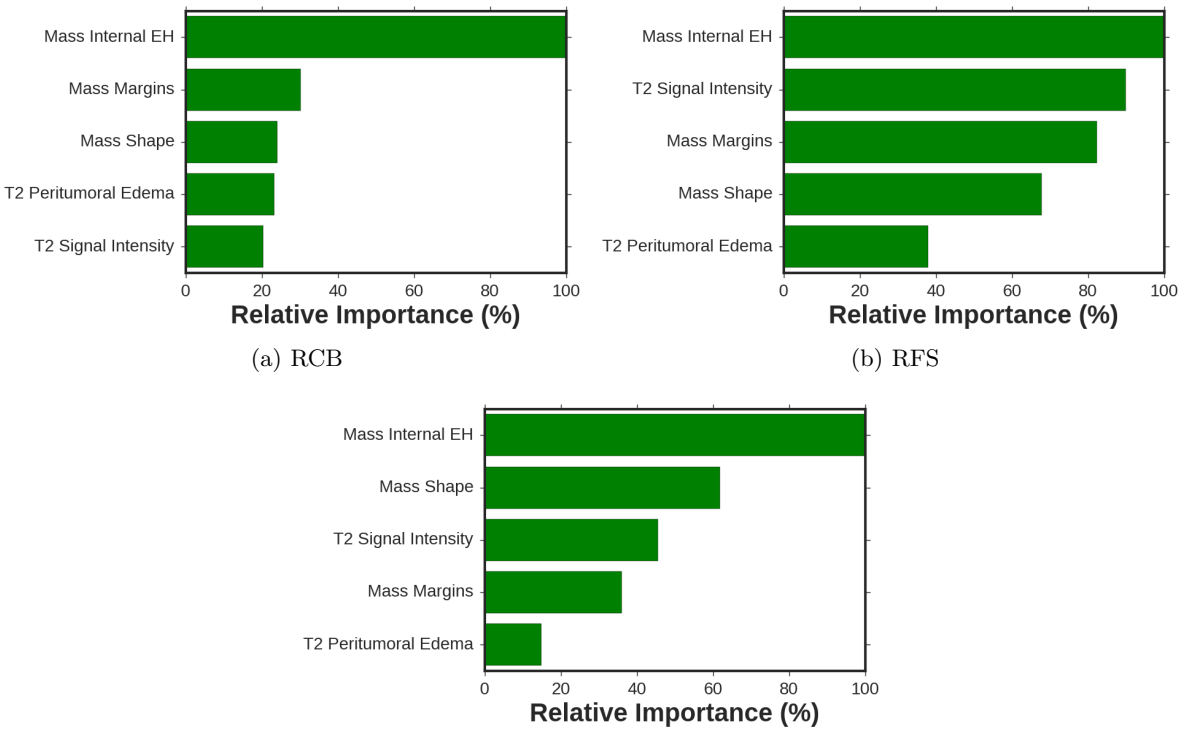
(b) RFS

(c) DSS

Figure 3.12: Multi-metric evaluation of 4-folds cross-validation incorporating RF classification method based on the minimum number of samples required to split an internal node to predict RCB score, RFS, and DSS.

for RCB (Figure 3.16a), RFS (Figure 3.16b), and DSS (Figure 3.16c). Additionally, Figure 3.17 presents the relative feature importance of the radiomics features of the multi-parametric model in prediction of the RCB score, RFS, and DSS using recursive feature elimination algorithm along with XGBoost classifier.

To go deeper in analyzing the importance of the extracted features, they were divided into five



(a) RCB

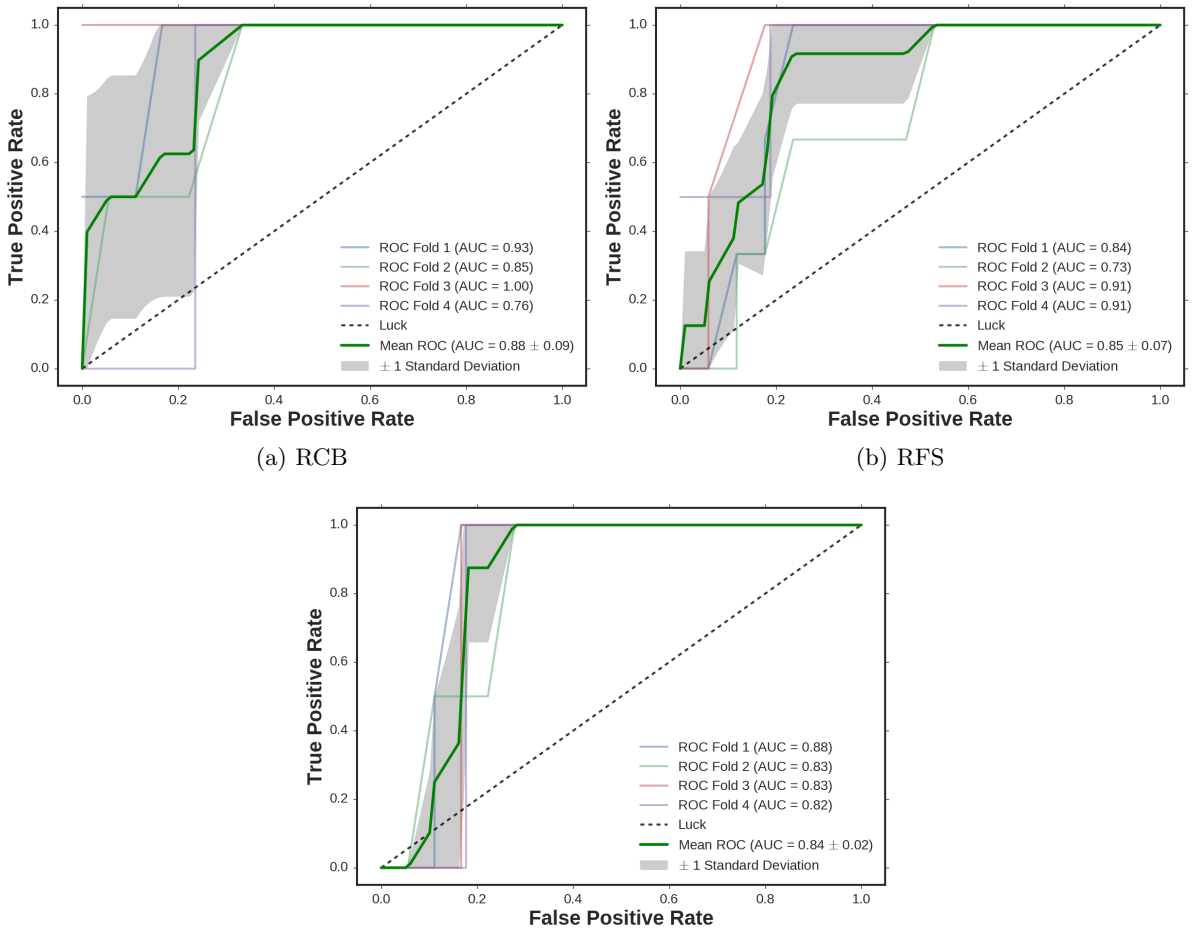
(b) RFS

(c) DSS

Figure 3.13: Relative features importance for RF classification method to predict RCB score, RFS, and DSS.

main categories including (1) kinetic, (2) functional, (3) molecular, (4) morphological, and (5) multi-parametric. As shown in Figure 3.16, XGBoost has shown a stable performance in prediction of all the three classes. Thus, XGBoost was chosen as the main classifier to investigate the importance of the defined main categories. In this regard, the ROC curves of each folds along with the mean ROC and ± 1 standard deviation of them as the confidence interval with an AUC value for each fold based on each main category were presented. 4-folds cross-validation was employed in prediction of RCB score, and RFS and 3-folds cross-validation was employed for predicting the DSS class since only 3 cases as death were reported in the data.

Figure 3.18 presents the ROC curves of prediction of the classes based on only functional features using XGBoost classifier. As shown, the highest accuracy was gained for the RFS class with an AUC of 0.75 ± 0.15. In addition to this, the prediction of RCB score using the functional features



(a) RCB

(b) RFS

(c) DSS

Figure 3.14: ROC curves of 4-folds cross-validation incorporating RF classification method to predict RCB score, RFS, and DSS.

was reasonable as well with an AUC value of 0.71 ± 0.13 as shown in Figure 3.18a.

Figure 3.19 shows the ROC curves of classification of RCB score, RFS, and DSS classes based on only extracted kinetic features employing k-folds cross-validation using XGBoost classifier. As seen, RFS class was predicted with a reasonable AUC value of 0.75 ± 0.15 with a maximum AUC value of 0.88 for two folds. However, the prediction of RCB score and DSS class was not stable. The prediction of RCB score using the kinetic features had a maximum AUC value of 0.86 in one folds and a minimum AUC value of 0.18 in another fold. That is why the mean AUC value of

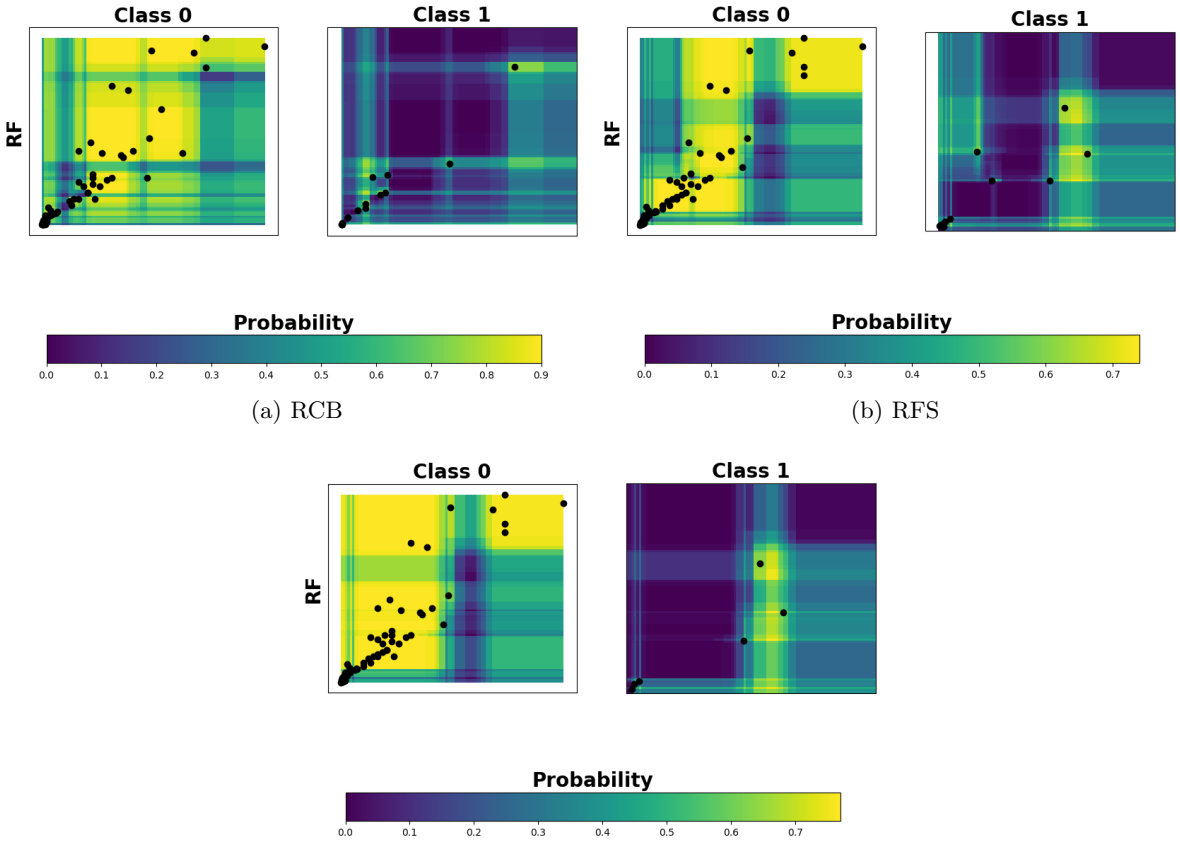
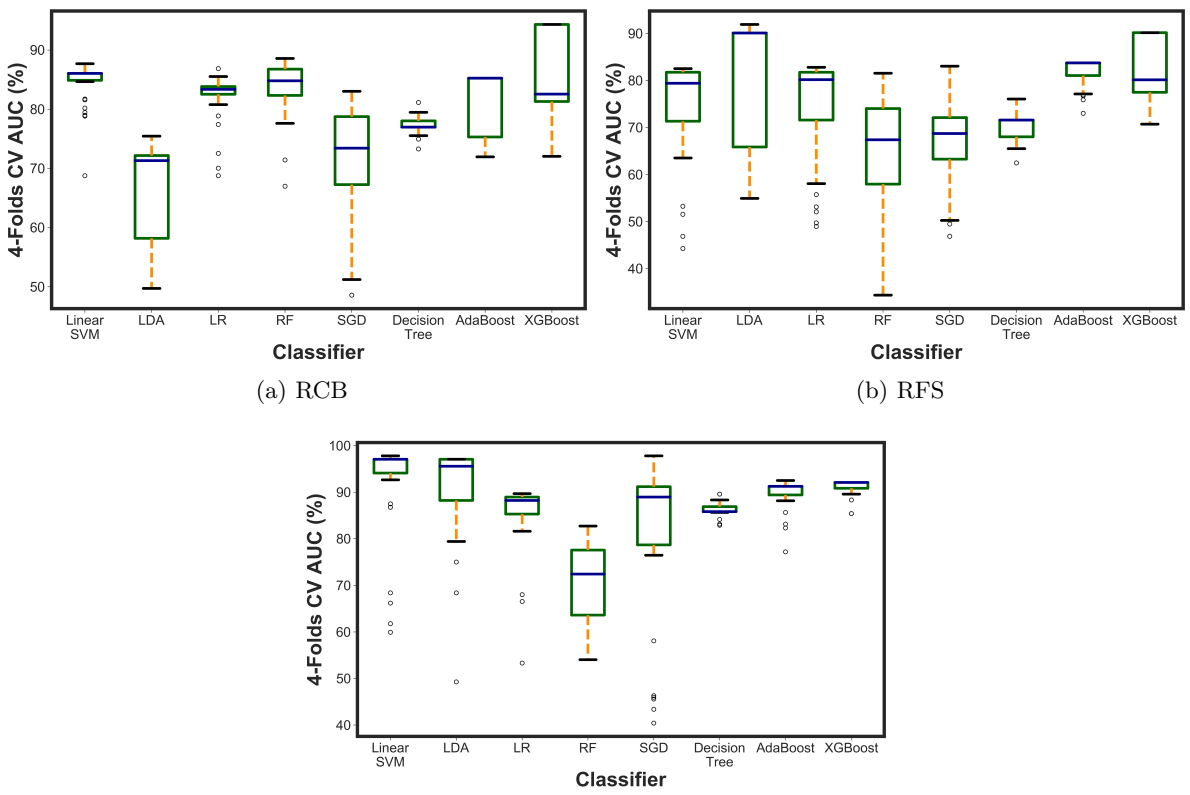


Figure 3.15: The decision boundaries of random forest in prediction of RCB score, RFS, and DSS. The x-axis was set to RL diameter and the y-axis was set to CC diameter.

0.62 ± 0.27 was reported which shows roughly 30% error in prediction. This might be due to the low number of patients in the data.

The classification results of the XGBoost based on morphological features in prediction of the RCB score, RFS, and DSS classes are presented in Figure 3.20. As shown the most stable result was presented in Figure 3.20a in prediction of RCB score with an AUC value of 0.78 ± 0.15. Although, the functional features showed a mean AUC value of 0.80 ± 0.25 in prediction of DSS class, the chance of random classification is still high due to a poor AUC value of 0.46 for one of the folds. It was pretty well depicted in Figure 3.20c with gray confidence interval which the green line is totally out of the shaded area. Furthermore, RFS was also predicted with a reasonable AUC value



(a) RCB

(b) RFS

(c) DSS

Figure 3.16: Box plot presentations of 4-folds cross-validation of AUC score based on recursive feature elimination using eight classifiers including linear SVM, LDA, RF, LR, SGD, decision tree, AdaBoost, and XGBoost in prediction of RCB score, RFS, and DSS.

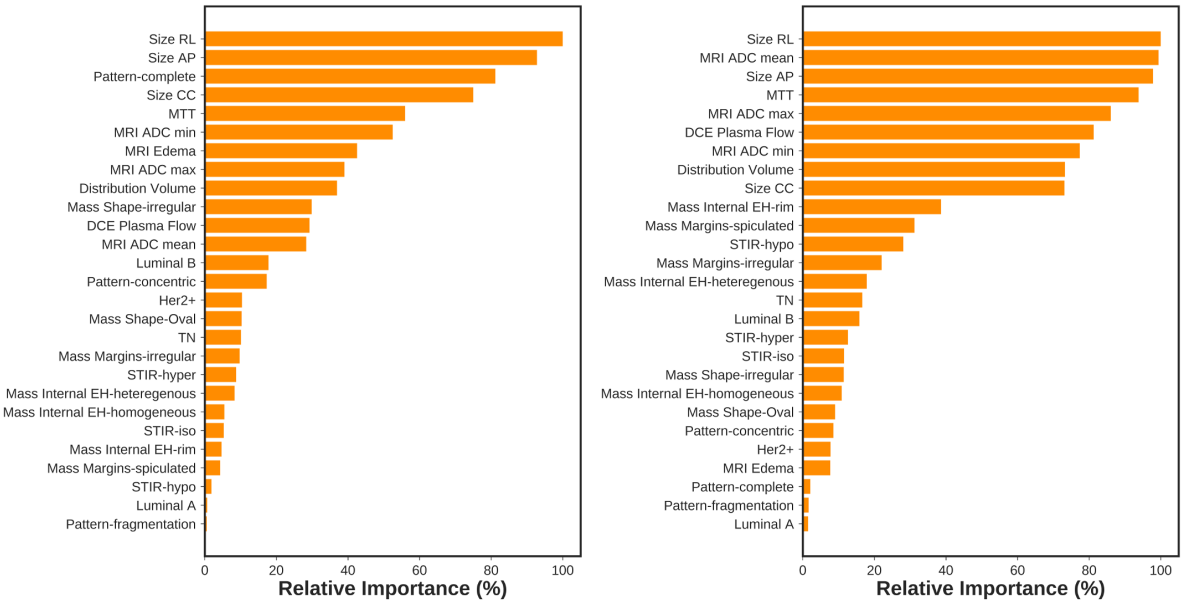
of 0.74 ± 0.15 with a maximum AUC value of 0.88.

Figure 3.21 shows the performance of the immuno-histo-chemical features and intrinsic molecular subtypes (luminal A, luminal B, Her2+ and triple-negative (TN) tumors) in prediction of RCB score via XGBoost classifier employing 4-folds cross-validation. As shown, luminal B has shown the best performance in prediction of RCB score with an AUC value of 0.68 ± 0.08. Based on its ROC curve, prediction using luminal B suffers from high false positive rate. However, by taking a look at the prediction of the RCB score using Her2+ and TN, it is obvious that it contains lower false positive rate. It is the trade-off (higher accuracy with high false positives, and lower accuracy with low false positives) that the radiologist should take into account. In addition to this, Figure

3.22 illustrates the performance of molecular subtypes in prediction of RFS. Similar to the results shown in Figure 3.21, classification using luminal A contains high numbers of false positives. On the contrary, the prediction of RFS using Her2+ and TN also contains high values of false positive. Moreover, the prediction of RFS using luminal B is totally random and it would be smarter to be neglected in prediction of RFS. Finally, Figure 3.23 illustrates the classification performance of DSS class using Her2+, TN, luminal A, and luminal B. Similar to RFS prediction using Her2+, high values of false positive can be seen. The best performance was reported using TN with an AUC value of 0.69 ± 0.25. It should be taken into account that 25% error might make the model stochastic for some instances. Thus, prediction using Her2+ with an AUC value of 0.64 ± 0.05 with only 5% error in 3-folds cross-validation would give us a better confidence interval to trust the resulted 64% accuracy.

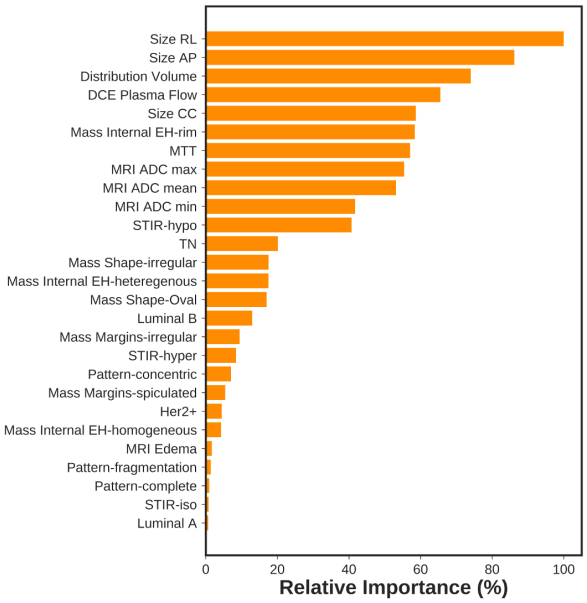
Lastly, Figure 3.24 shows the ROC curves of prediction of RCB score, RFS, and DSS classes employing multi-parametric features using XGBoost classifier. As shown, RCB score, RFS, and DSS had mean ACU values of 0.95 ± 0.05, 0.90 ± 0.13, and 0.88 ± 0.06, respectively.

To recapitulate and compare the importance of the each defined category with the performance of the multi-parametric results, Figure 3.25 was presented. As shown, the results using multi-parametric radiomics outperformed the results using the defined main categories. Other than the multi-parametric analysis, in prediction of the RCB score, morphological features with an AUC value of 0.78 ± 0.15 and functional features with an AUC value of 0.71 ± 0.13 showed the best performances. Similarly, in prediction of the RFS, kinetic and functional features with an AUC value of 0.75 ± 0.15 and morphological features with an AUC value of 0.74 ± 0.15 showed the best performances. Finally, in prediction of DSS, morphological features with an AUC value of 0.80 ± 0.25 and TN with an AUC value of 0.69 ± 0.25 showed the best performances.



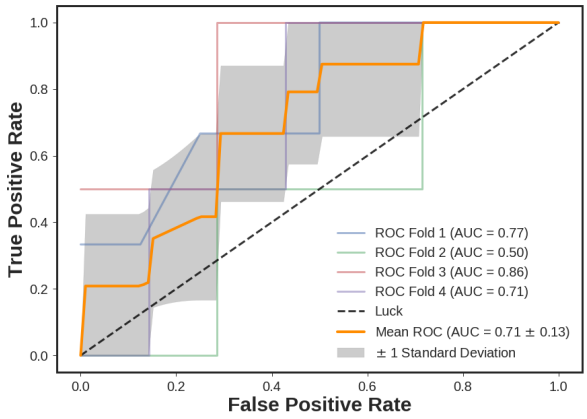
(a) RCB

(b) RFS

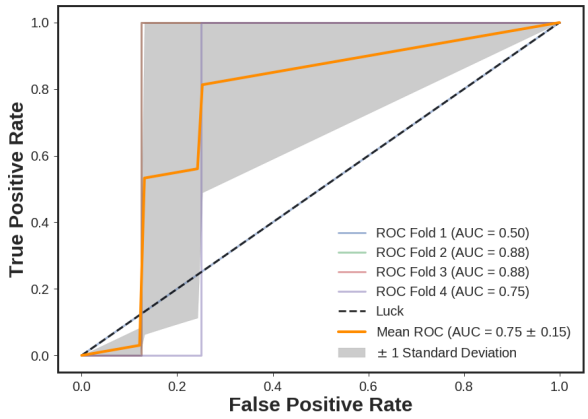


(c) DSS

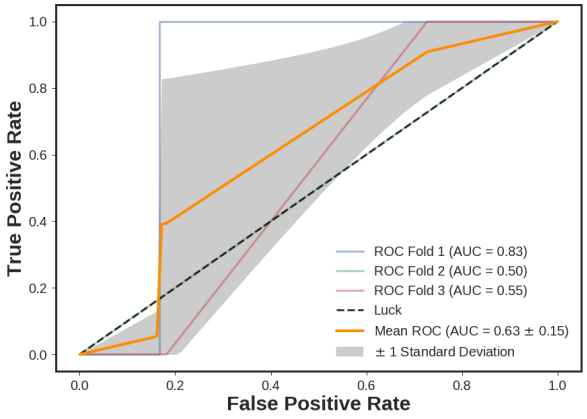
Figure 3.17: Relative feature importance of the radiomics features of the multi-parametric model in prediction of the RCB score, RFS, and DSS using recursive feature elimination algorithm along with XGBoost classifier.



(a) RCB

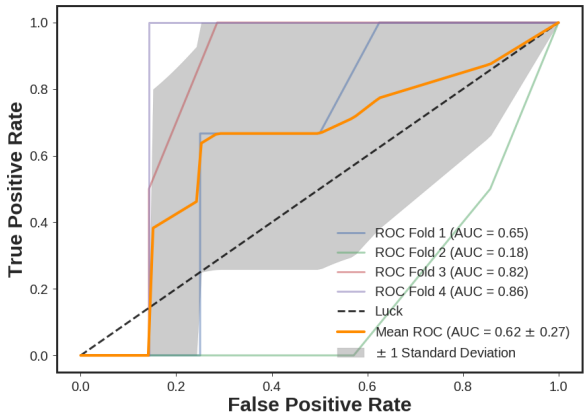


(b) RFS

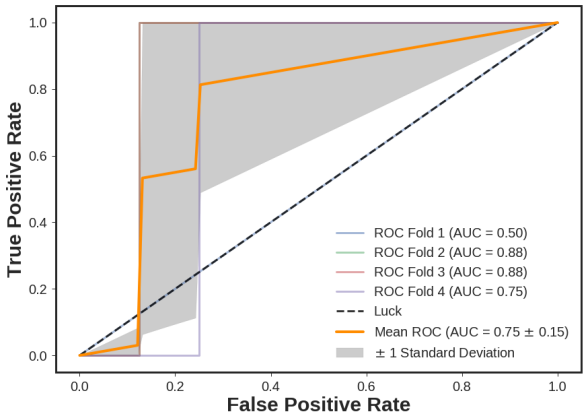


(c) DSS

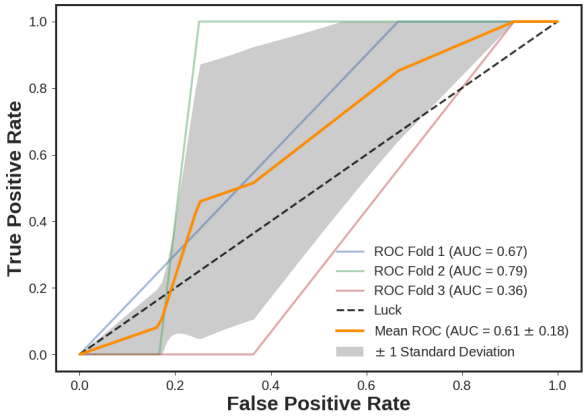
Figure 3.18: ROC curves of prediction of the RCB score, RFS, and DSS using XGBoost based on functional features.



(a) RCB

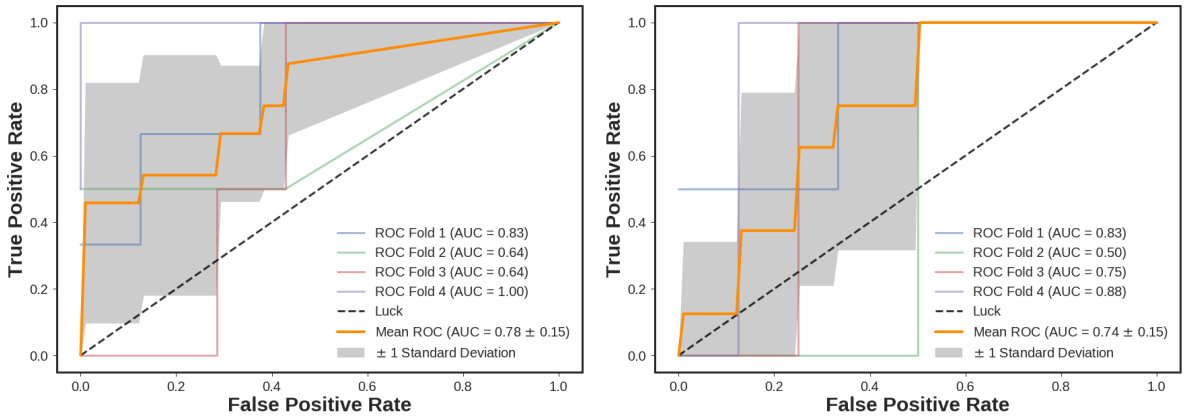


(b) RFS



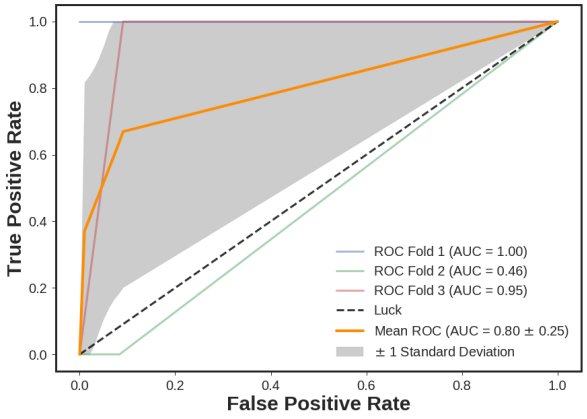
(c) DSS

Figure 3.19: ROC curves of prediction of the RCB score, RFS, and DSS using XGBoost based on kinetic features.



(a) RCB

(b) RFS



(c) DSS

Figure 3.20: ROC curves of prediction of the RCB score, RFS, and DSS using XGBoost based on morphological features.

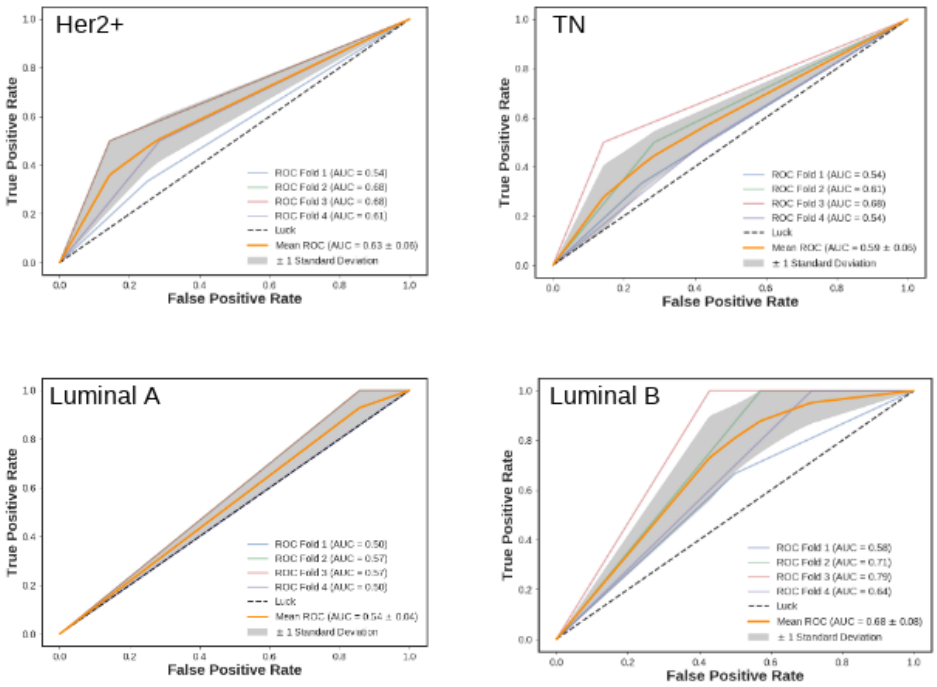


Figure 3.21: ROC curves of prediction of the RCB score using XGBoost based on Her2+, TN, Luminal A, and Luminal B.

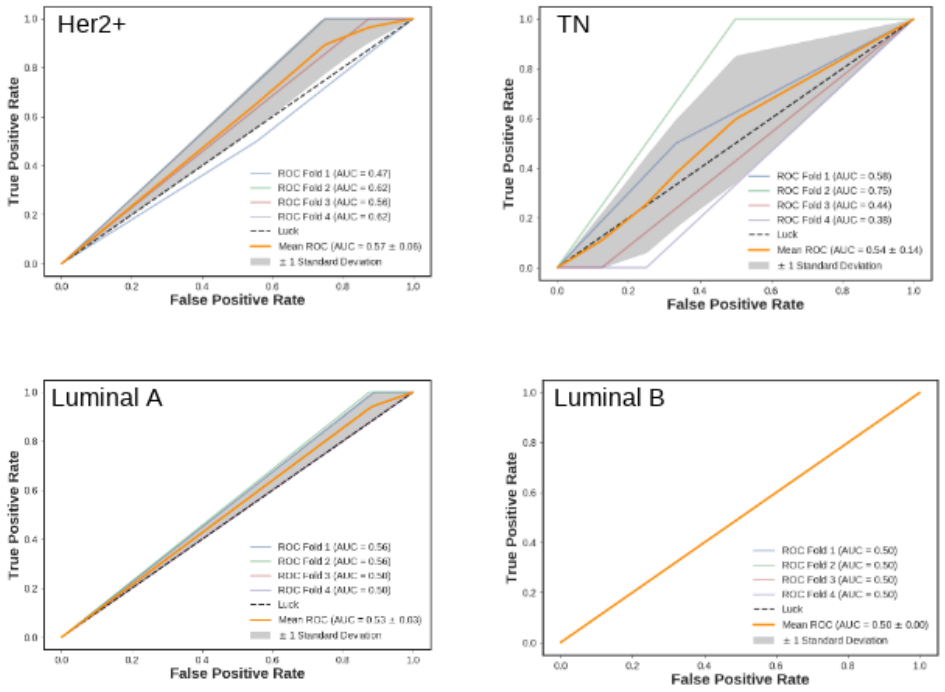


Figure 3.22: ROC curves of prediction of the RFS using XGBoost based on Her2+, TN, Luminal A, and Luminal B.

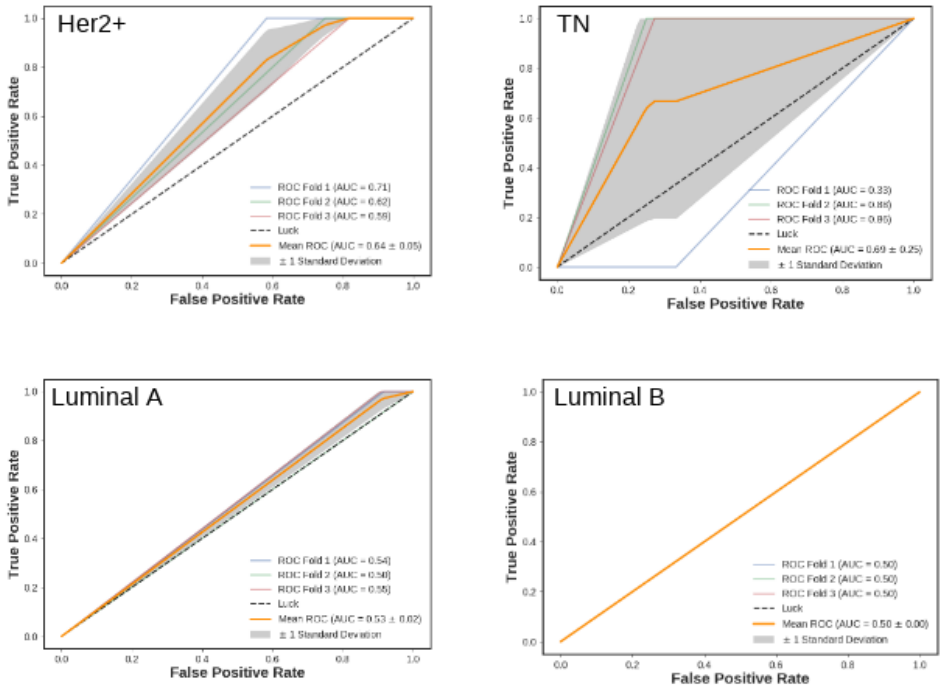
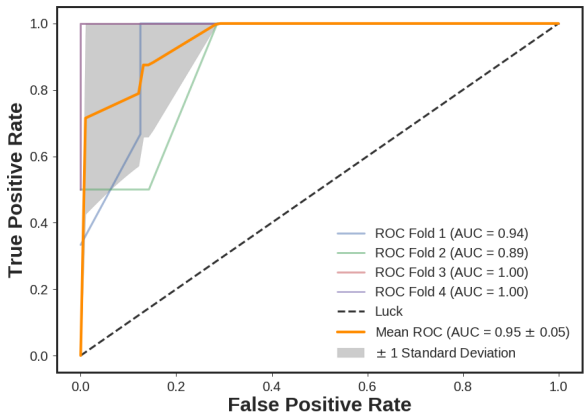
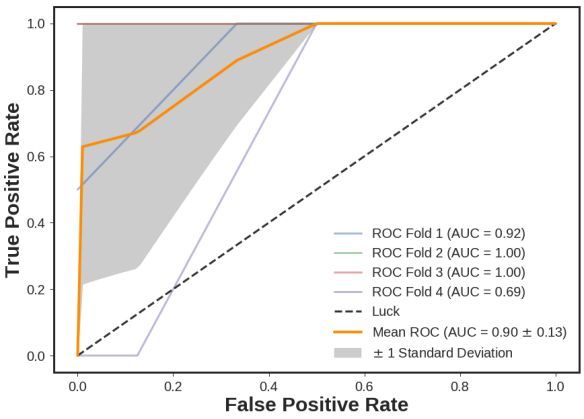


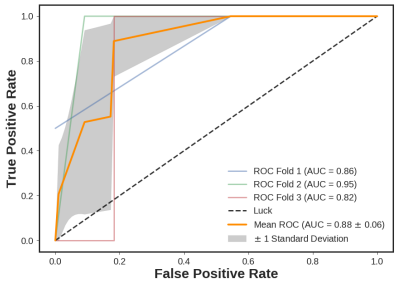
Figure 3.23: ROC curves of prediction of the DSS using XGBoost based on Her2+, TN, Luminal A, and Luminal B.



(a) RCB

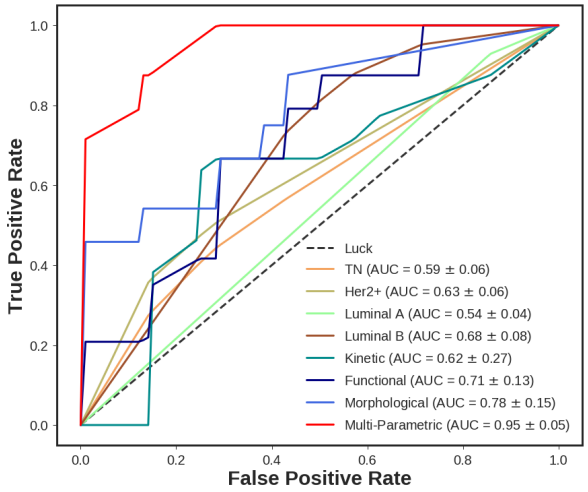


(b) RFS

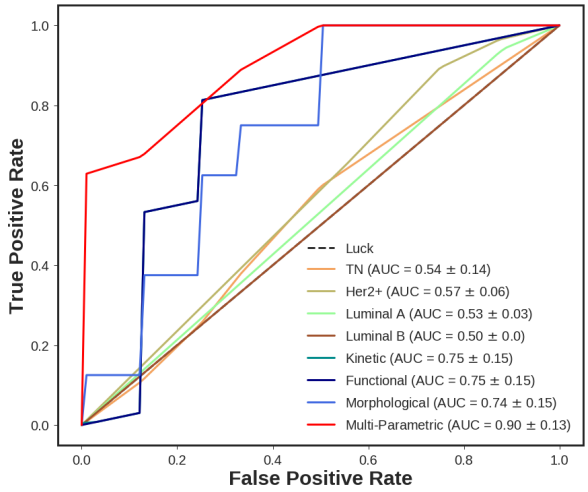


(c) DSS

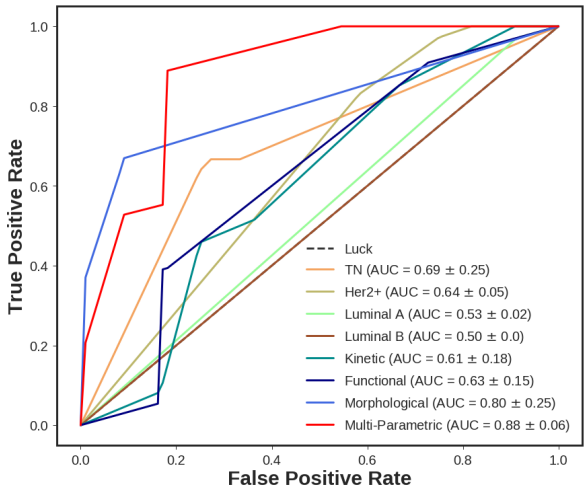
Figure 3.24: ROC curves of prediction of the RCB score, RFS, and DSS using XGBoost based on multi-parametric features.



(a) RCB



(b) RFS



(c) DSS

Figure 3.25: An exhaustive comparison of the ROC curves of prediction of the RCB score, RFS, and DSS based on all of the proposed models.

CHAPTER 4

SUMMARY

Resting-state functional magnetic resonance imaging (fMRI) images allow the level of activity in a patient's brain to be observed. The fMRI of patients before and after they underwent a double-blind smoking cessation treatment were considered. For the first time, this study aims at developing new theory-driven biomarkers by implementing and evaluating novel techniques from resting-state scans that can be used in relapse prediction in nicotine-dependent patients and future treatment efficacy. In this regard two classes of patients were studied, one took the drug N-acetylcysteine and the other took a placebo. The goal was to classify the patients as treatment or non-treatment, based on their fMRI scans and predict relapse in the future. The image slices of brain are used as the variable, and the results consisted of a big data problem with about 95,000,000 inputs per subject. To handle this problem, the data had to be reduced and the first process in doing that was to create three masks to apply to all images. The masks were created by averaging the before images for all patients and selecting the top 40% of voxels from that average, selecting voxels just in the limbic system, and a combination of both. These masks were then applied to all fMRI images for all patients. The average of the difference in the before treatment and after fMRI images for each patient were found and these were flattened to one dimension. A matrix was made by stacking these 1D arrays on top of each other and a data reduction algorithm was applied on it. Lastly, the matrix was fed into some machine learning algorithms and leave-one-run-out cross-validation was used to test out the accuracy. Various classifiers were compared including genetic programming, support vector machines with different kernels, decision trees, and Naive-Bayes along with independent component analysis, principal component analysis, and singular value decomposition. Also compared the classifiers' accuracy at high activity parts of the brain, limbic system, and a combination of both. The results suggest that there is a big difference in the resting-state fMRI images of a smoker that undergoes this smoking cessation treatment compared to a smoker that receives a placebo. Additionally, an under-complete autoencoder employing different convolutional layers was employed to extract features from the 4D fMRI images to feed into several robust classifiers including boosting

algorithms. The developed pipeline has the ability to backtrack the features to map on the brain template to visualize the involved areas of the brain for each subject. The extracted features along with XGBoost classifier confirm the areas around meso-limbic system can be used to predict relapse in heavy smokers subjects.

Neo-adjuvant chemotherapy is the treatment of choice in patients with locally advanced breast cancer to reduce tumor burden, and potentially enable breast conservation. Response to treatment is assessed by histopathology from surgical specimen, a pathological complete response (pCR), or a minimal residual disease are associated with an improved disease-free, and overall survival. Early identification of non-responders is crucial as these patients might require different, or more aggressive treatment. Multi-parametric magnetic resonance imaging (mpMRI) using different morphological and functional MRI parameters such as T_2-weighted, dynamic contrast-enhanced (DCE) MRI, and diffusion weighted imaging (DWI) has emerged as the method of choice for the early response assessments to neo-adjuvant chemotherapy. Although, mpMRI is superior to conventional mammography for predicting treatment response, and evaluating residual disease, yet there is still room for improvement. In the past decade, the field of medical imaging analysis has grown exponentially, with an increased numbers of pattern recognition tools, and an increase in data sizes. These advances have heralded the field of radiomics. Radiomics allows the high-throughput extraction of the quantitative features that result in the conversion of images into mineable data, and the subsequent analysis of the data for an improved decision support with response monitoring during neo-adjuvant chemotherapy being no exception. In this study, we determine the importance and ranking of the extracted parameters from mpMRI using T_2-weighted, DCE, and DWI for prediction of pCR and patient outcomes with respect to metastases and disease-specific death.

REFERENCES

[1] Anke Meyer-Baese. *Pattern recognition for medical imaging*. Academic Press, 2004.

[2] Anke Meyer-Baese and Volker J Schmid. *Pattern recognition and signal analysis in medical imaging*. Elsevier, 2014.

[3] Ann E Kelley. Memory and addiction: shared neural circuitry and molecular mechanisms. *Neuron*, 44(1):161–179, 2004.

[4] Amirhessam Tahmassebi, Katja Pinker-Domenig, Georg Wengert, Marc Lobbes, Andreas Stadlbauer, Francisco J Romero, Diego P Morales, Encarnacion Castillo, Antonio Garcia, Guillermo Botella, et al. Dynamical graph theory networks techniques for the analysis of sparse connectivity networks in dementia. In *Smart Biomedical and Physiological Sensor Technology XIV*, volume 10216, page 1021609. International Society for Optics and Photonics, 2017.

[5] Amirhessam Tahmassebi, Amir H Gandomi, Ian McCann, Mieke HJ Schulte, Lianne Schmaal, Anna E Goudriaan, and Anke Meyer-Baese. fmri smoking cessation classification. *IEEE Transactions on Cybernetics*, 2017.

[6] Amirhessam Tahmassebi, Katja Pinker-Domenig, Georg Wengert, Marc Lobbes, Andreas Stadlbauer, Norelle C Wildburger, Francisco J Romero, Diego P Morales, Encarnacion Castillo, Antonio Garcia, et al. The driving regulators of the connectivity protein network of brain malignancies. In *Smart Biomedical and Physiological Sensor Technology XIV*, volume 10216, page 1021605. International Society for Optics and Photonics, 2017.

[7] Amirhessam Tahmassebi, Ali Moradi Amani, Katja Pinker-Domenig, and Anke Meyer-Baese. Determining disease evolution driver nodes in dementia networks. In *Medical Imaging 2018: Biomedical Applications in Molecular, Structural, and Functional Imaging*, volume 10578, page 1057829. International Society for Optics and Photonics, 2018.

[8] Martin A Lindquist. Functional causal mediation analysis with an application to brain connectivity. *Journal of the American Statistical Association*, 107(500):1297–1309, 2012.

[9] Hua Xie, Vince Calhoun, Javier Gonzalez-Castillo, Eswar Damaraju, Robyn Miller, Peter Bandettini, and Sunanda Mitra. Whole-brain connectivity dynamics reflect both task-specific and individual-specific modulation: a multitask study. *NeuroImage*, 2017.

[10] Anil K Seth, Paul Chorley, and Lionel C Barnett. Granger causality analysis of fmri bold signals is invariant to hemodynamic convolution but not downsampling. *Neuroimage*, 65:540–555, 2013.

[11] Tingting Zhang, Fan Li, Lane Beckes, Casey Brown, and James A Coan. Nonparametric inference of the hemodynamic response using multi-subject fmri data. *NeuroImage*, 63(3):1754–1765, 2012.

[12] Amirhessam Tahmassebi. *Fluid Flow Through Carbon Nanotubes And Graphene Based Nanostructures.* PhD thesis, University of Akron, 2015.

[13] Amirhessam Tahmassebi and Alper Buldum. Fluid flow calculations of graphene composites. In *APS March Meeting Abstracts*, 2015.

[14] Aria Smitha, Anahid Ehtemami, Daniel Frattea, Anke Meyer-Baesea, Olmo Zavala-Romeroa, Anna E Goudriaanb, Lianne Schmaalc, and Mieke HJ Schulteb. Functional connectivity analysis of resting-state fmri networks in nicotine dependent patients. In *SPIE Medical Imaging*, pages 978827–978827. International Society for Optics and Photonics, 2016.

[15] Yi-Jen Lin and Alan P Koretsky. Manganese ion enhances t1-weighted mri during brain activation: An approach to direct imaging of brain function. *Magnetic resonance in medicine*, 38(3):378–388, 1997.

[16] Karen Kinkel, Thomas H Helbich, Laura J Esserman, John Barclay, Ellen H Schwerin, Edward A Sickles, and Nola M Hylton. Dynamic high-spatial-resolution mr imaging of suspicious breast lesions: diagnostic criteria and interobserver variability. *American journal of Roentgenology*, 175(1):35–43, 2000.

[17] Christiane K Kuhl, Hans H Schild, and Nuschin Morakkabati. Dynamic bilateral contrast-enhanced mr imaging of the breast: trade-off between spatial and temporal resolution. *Radiology*, 236(3):789–800, 2005.

[18] Reiko Woodhams, Keiji Matsunaga, Keiichi Iwabuchi, Shinichi Kan, Hirofumi Hata, Masaru Kuranami, Masahiko Watanabe, and Kazushige Hayakawa. Diffusion-weighted imaging of malignant breast tumors: the usefulness of apparent diffusion coefficient (adc) value and adc map for the detection of malignant breast tumors and evaluation of cancer extension. *Journal of computer assisted tomography*, 29(5):644–649, 2005.

[19] Patric Hagmann, Lisa Jonasson, Philippe Maeder, Jean-Philippe Thiran, Van J Wedeen, and Reto Meuli. Understanding diffusion mr imaging techniques: from scalar diffusion-weighted imaging to diffusion tensor imaging and beyond. *Radiographics*, 26(suppl_1):S205–S223, 2006.

[20] Guillaume Lemaitre. *Computer-Aided Diagnosis for Prostate Cancer using Multi-Parametric Magnetic Resonance Imaging.* PhD thesis, Ph. D. dissertation, Universitat de Girona and Université de Bourgogne, 2016.

[21] Ian McCann, Amirhessam Tahmassebi, Simon Y Foo, Gordon Erlebacher, and Anke Meyer-Baese. Newsanalyticaltoolkit: an online natural language processing platform to analyze news. In *Next-Generation Analyst VI*, volume 10653, page 106530P. International Society for Optics and Photonics, 2018.

[22] Amirhessam Tahmassebi, Amir H Gandomi, and Anke Meyer-Baese. Stock risk assessment via multi-objective genetic programming. *Journal of Postdoctoral Research*, 6(3), 2018.

[23] Behshad Mohebali, Amirhessam Tahmassebi, Amir H Gandomi, Anke Meyer-Baese, and Simon Y Foo. A scalable communication abstraction framework for internet of things applications using raspberry pi. In *Disruptive Technologies in Information Sciences*, volume 10652, page 1065205. International Society for Optics and Photonics, 2018.

[24] Francisco J Romero, Diego P Morales, Encarnación Castillo, Antonio García, Amirhessam Tahmassebi, and Anke Meyer-Baese. Reconfigurable wearable to monitor physiological variables and movement. In *Smart Biomedical and Physiological Sensor Technology XIV*, volume 10216, page 1021608. International Society for Optics and Photonics, 2017.

[25] Victor Toral-Lopez, Salvador Criado, Francisco J Romero, Diego P Morales, Encarnación Castillo, Antonio García, Amirhessam Tahmassebi, and Anke Meyer-Baese. Wearable biosignal acquisition system for decision aid. In *Smart Biomedical and Physiological Sensor Technology XV*, volume 10662, page 106620F. International Society for Optics and Photonics, 2018.

[26] Francisco J Romero-Maldonado, Santiago Juarez, Inmaculada Ortiz-Gomez, Diego P Morales, Alfonso Salinas-Castillo, Encarnacion Castillo, Antonio García, Amirhessam Tahmassebi, and Anke Meyer-Bäse. Reconfigurable instrument for measuring variations of capacitor's dielectric: an application to olive oil quality monitoring. In *Sensing for Agriculture and Food Quality and Safety X*, volume 10665, page 106650F. International Society for Optics and Photonics, 2018.

[27] Amirhessam Tahmassebi, Amir H Gandomi, Ian McCann, Mieke HJ Schulte, Lianne Schmaal, Anna E Goudriaan, and Anke Meyer-Baese. An evolutionary approach for fmri big data classification. In *Evolutionary Computation (CEC), 2017 IEEE Congress on*, pages 1029–1036. IEEE, 2017.

[28] Lianne Schmaal, Dick J Veltman, Aart Nederveen, Wim Van Den Brink, and Anna E Goudriaan. N-acetylcysteine normalizes glutamate levels in cocaine-dependent patients: a randomized crossover magnetic resonance spectroscopy study. *Neuropsychopharmacology*, 37(9):2143–2152, 2012.

[29] Steven D LaRowe, Pascale Mardikian, Robert Malcolm, Hugh Myrick, Peter Kalivas, Krista McFarland, Michael Saladin, Aimee McRae, and Kathleen Brady. Safety and tolerability of n-acetylcysteine in cocaine-dependent individuals. *The American Journal on Addictions*, 15(1):105–110, 2006.

[30] David A Baker, Krista McFarland, Russell W Lake, Hui Shen, Xing-Chun Tang, Shigenobu Toda, and Peter W Kalivas. Neuroadaptations in cystine-glutamate exchange underlie cocaine relapse. *Nature neuroscience*, 6(7):743–749, 2003.

[31] Florian Schubert, Jürgen Gallinat, Frank Seifert, and Herbert Rinneberg. Glutamate concentrations in human brain using single voxel proton magnetic resonance spectroscopy at 3 tesla. *Neuroimage*, 21(4):1762–1771, 2004.

[32] Jon E Grant, Suck Won Kim, and Brian L Odlaug. N-acetyl cysteine, a glutamate-modulating agent, in the treatment of pathological gambling: a pilot study. *Biological psychiatry*, 62(6):652–657, 2007.

[33] Lori A Knackstedt, Steven LaRowe, Pascale Mardikian, Robert Malcolm, Himanshu Upadhyaya, Sarra Hedden, Athina Markou, and Peter W Kalivas. The role of cystine-glutamate exchange in nicotine dependence in rats and humans. *Biological psychiatry*, 65(10):841–845, 2009.

[34] Kevin M Gray, Noreen L Watson, Matthew J Carpenter, and Steven D LaRowe. N-acetylcysteine (nac) in young marijuana users: an open-label pilot study. *The American journal on addictions/American Academy of Psychiatrists in Alcoholism and Addictions*, 19(2):187, 2010.

[35] Mengyu Dai, Zhengwu Zhang, and Anuj Srivastava. Testing stationarity of brain functional connectivity using change-point detection in fmri data. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 19–27, 2016.

[36] Derrek P Hibar, Jason L Stein, Miguel E Renteria, Alejandro Arias-Vasquez, Sylvane Desrivières, Neda Jahanshad, Roberto Toro, Katharina Wittfeld, Lucija Abramovic, Micael Andersson, et al. Common genetic variants influence human subcortical brain structures. *Nature*, 520(7546):224–229, 2015.

[37] Junwei Han, Changyuan Chen, Ling Shao, Xintao Hu, Jungong Han, and Tianming Liu. Learning computational models of video memorability from fmri brain imaging. *IEEE transactions on cybernetics*, 45(8):1692–1703, 2015.

[38] Farid Yaghouby, Christopher J Schildt, Kevin D Donohue, Bruce F O'Hara, and Sridhar Sunderam. Validation of a closed-loop sensory stimulation technique for selective sleep restriction in mice. In *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE*, pages 3771–3774. IEEE, 2014.

[39] Tom M Mitchell, Svetlana V Shinkareva, Andrew Carlson, Kai-Min Chang, Vicente L Malave, Robert A Mason, and Marcel Adam Just. Predicting human brain activity associated with the meanings of nouns. *science*, 320(5880):1191–1195, 2008.

[40] David D Cox and Robert L Savoy. Functional magnetic resonance imaging (fmri)brain reading: detecting and classifying distributed patterns of fmri activity in human visual cortex. *Neuroimage*, 19(2):261–270, 2003.

[41] Kenneth A Norman, Sean M Polyn, Greg J Detre, and James V Haxby. Beyond mind-reading: multi-voxel pattern analysis of fmri data. *Trends in cognitive sciences*, 10(9):424–430, 2006.

[42] Thomas A Carlson, Paul Schrater, and Sheng He. Patterns of activity in the categorical representations of objects. *Journal of cognitive neuroscience*, 15(5):704–717, 2003.

[43] Francisco Pereira, Tom Mitchell, and Matthew Botvinick. Machine learning classifiers and fmri: a tutorial overview. *Neuroimage*, 45(1):S199–S209, 2009.

[44] Tom M Mitchell, Rebecca Hutchinson, Radu S Niculescu, Francisco Pereira, Xuerui Wang, Marcel Just, and Sharlene Newman. Learning to decode cognitive states from brain images. *Machine learning*, 57(1-2):145–175, 2004.

[45] Janaina Mourão-Miranda, Arun LW Bokde, Christine Born, Harald Hampel, and Martin Stetter. Classifying brain states and determining the discriminating activation patterns: support vector machine on functional mri data. *NeuroImage*, 28(4):980–995, 2005.

[46] Anahid Ehtemami, Aria Smith, Daniel Fratte, Anke Meyer-Baese, Anna E Goudriaan, Lianne Schmaal, Mieke HJ Schulte, and Olmo Zavala-Romero. Functional connectivity analysis of resting-state fmri networks in nicotine dependent patients. In *SPIE Medical Imaging*, pages 978827–978827. International Society for Optics and Photonics, 2016.

[47] Lianne Schmaal, Lotte Berk, Kai P Hulstijn, Janna Cousijn, Reinout W Wiers, and Wim van den Brink. Efficacy of n-acetylcysteine in the treatment of nicotine dependence: a double-blind placebo-controlled pilot study. *European addiction research*, 17(4):211–216, 2011.

[48] Hui Shen, Lubin Wang, Yadong Liu, and Dewen Hu. Discriminative analysis of resting-state functional connectivity patterns of schizophrenia using low dimensional embedding of fmri. *Neuroimage*, 49(4):3110–3121, 2010.

[49] B Froeliger, PA McConnell, N Stankeviciute, EA McClure, PW Kalivas, and KM Gray. The effects of n-acetylcysteine on frontostriatal resting-state functional connectivity, withdrawal symptoms and smoking abstinence: a double-blind, placebo-controlled fmri pilot study. *Drug and alcohol dependence*, 156:234–242, 2015.

[50] Nathan W Churchill, Anita Oder, Herve Abdi, Fred Tam, Wayne Lee, Christopher Thomas, Jon E Ween, Simon J Graham, and Stephen C Strother. Optimizing preprocessing and analysis pipelines for single-subject fmri. i. standard temporal motion and physiological noise correction methods. *Human brain mapping*, 33(3):609–627, 2012.

[51] Jeremy D Schmahmann, Julien Doyon, David McDonald, Colin Holmes, Karyne Lavoie, Amy S Hurwitz, Noor Kabani, Arthur Toga, Alan Evans, and Michael Petrides. Three-dimensional mri atlas of the human cerebellum in proportional stereotaxic space. *Neuroimage*, 10(3):233–260, 1999.

[52] Nathalie Tzourio-Mazoyer, Brigitte Landeau, Dimitri Papathanassiou, Fabrice Crivello, Olivier Etard, Nicolas Delcroix, Bernard Mazoyer, and Marc Joliot. Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain. *Neuroimage*, 15(1):273–289, 2002.

[53] Xiaoping Hu, Tuong Huu Le, Todd Parrish, and Peter Erhard. Retrospective estimation and correction of physiological fluctuation in functional mri. *Magnetic resonance in medicine*, 34(2):201–212, 1995.

[54] Karl J Friston, Steven Williams, Robert Howard, Richard SJ Frackowiak, and Robert Turner. Movement-related effects in fmri time-series. *Magnetic resonance in medicine*, 35(3):346–355, 1996.

[55] Karl J Friston, Andrew P Holmes, JB Poline, PJ Grasby, SCR Williams, Richard SJ Frackowiak, and Robert Turner. Analysis of fmri time-series revisited. *Neuroimage*, 2(1):45–53, 1995.

[56] Alan V Oppenheim and Jae S Lim. The importance of phase in signals. *Proceedings of the IEEE*, 69(5):529–541, 1981.

[57] A. Gholipour, N. Kehtarnavaz, R. Briggs, M. Devous, and K. Gopinath. Brain functional localization: A survey of image registration techniques. *IEEE Transactions on Medical Imaging*, 26(4):427–451, April 2007.

[58] JL Lancaster, LH Rainey, JL Summerlin, CS Freitas, PT Fox, AC Evans, AW Toga, and JC Mazziotta. Automated labeling of the human brain: a preliminary report on the development and evaluation of a forward-transform method. *Human brain mapping*, 5(4):238, 1997.

[59] Jack L Lancaster, Marty G Woldorff, Lawrence M Parsons, Mario Liotti, Catarina S Freitas, Lacy Rainey, Peter V Kochunov, Dan Nickerson, Shawn A Mikiten, and Peter T Fox. Automated talairach atlas labels for functional brain mapping. *Human brain mapping*, 10(3):120–131, 2000.

[60] Joseph A Maldjian, Paul J Laurienti, Robert A Kraft, and Jonathan H Burdette. An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fmri data sets. *Neuroimage*, 19(3):1233–1239, 2003.

[61] Russell A Poldrack. Region of interest analysis for fmri. *Social cognitive and affective neuroscience*, 2(1/4):67, 2007.

[62] Hans C Breiter, Randy L Gollub, Robert M Weisskoff, David N Kennedy, Nikos Makris, Joshua D Berke, Julie M Goodman, Howard L Kantor, David R Gastfriend, Jonn P Riorden, et al. Acute effects of cocaine on human brain activity and emotion. *Neuron*, 19(3):591–611, 1997.

[63] Rita Z Goldstein and Nora D Volkow. Drug addiction and its underlying neurobiological basis: neuroimaging evidence for the involvement of the frontal cortex. *American Journal of Psychiatry*, 159(10):1642–1652, 2002.

[64] Anke Meyer-Baese, Axel Wismueller, and Oliver Lange. Comparison of two exploratory data analysis methods for fmri: unsupervised clustering versus independent component analysis. *IEEE Transactions on Information Technology in Biomedicine*, 8(3):387–398, 2004.

[65] Vince D Calhoun, Jingyu Liu, and Tülay Adalı. A review of group ica for fmri data and ica for joint inference of imaging, genetic, and erp data. *Neuroimage*, 45(1):S163–S172, 2009.

[66] Martin J McKeown, Scott Makeig, Greg G Brown, Tzyy-Ping Jung, Sandra S Kindermann, Anthony J Bell, and Terrence J Sejnowski. Analysis of fmri data by blind separation into independent spatial components. Technical report, DTIC Document, 1997.

[67] Andrew Y Ng. Feature selection, l 1 vs. l 2 regularization, and rotational invariance. In *Proceedings of the twenty-first international conference on Machine learning*, page 78. ACM, 2004.

[68] Behshad Mohebali. *Characterization of the common mode features of a 3-phase full-bridge inverter using frequency domain approaches*. PhD thesis, The Florida State University, 2016.

[69] Christopher M Bishop. *Pattern recognition and machine learning*. springer, 2006.

[70] Pang-Ning Tan et al. *Introduction to data mining*. Pearson Education India, 2006.

[71] Richard O Duda, Peter E Hart, and David G Stork. *Pattern classification*. John Wiley & Sons, 2012.

[72] Joset A Etzel, Valeria Gazzola, and Christian Keysers. An introduction to anatomical roi-based fmri classification analysis. *Brain research*, 1282:114–125, 2009.

[73] Amirhessam Tahmassebi and Amir H Gandomi. Building energy consumption forecast using multi-objective genetic programming. *Measurement*, 118:164–171, 2018.

[74] Amirhessam Tahmassebi, Amir H Gandomi, Mieke HJ Schulte, Anna E Goudriaan, Simon Y Foo, and Anke Meyer-Baese. Optimized naive-bayes and decision tree approaches for fmri smoking cessation classification. *Complexity*, 2018, 2018.

[75] Tor D Wager and Thomas E Nichols. Optimization of experimental design in fmri: a general framework using a genetic algorithm. *Neuroimage*, 18(2):293–309, 2003.

[76] Amirhessam Tahmassebi and Amir H Gandomi. Genetic programming based on error decomposition: A big data approach. In *Genetic Programming Theory and Practice XV*, pages 135–147. Springer, 2018.

[77] Markus Brameier and Wolfgang Banzhaf. A comparison of linear genetic programming and neural networks in medical data mining. *IEEE Transactions on Evolutionary Computation*, 5(1):17–26, 2001.

[78] Thomas Loveard and Victor Ciesielski. Representing classification problems in genetic programming. In *Evolutionary Computation, 2001. Proceedings of the 2001 Congress on*, volume 2, pages 1070–1077. IEEE, 2001.

[79] Pedro G Espejo, Sebastián Ventura, and Francisco Herrera. A survey on the application of genetic programming to classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 40(2):121–144, 2010.

[80] J Krishna Kishore, Lalit M. Patnaik, V Mani, and VK Agrawal. Application of genetic programming for multicategory pattern classification. *IEEE transactions on evolutionary computation*, 4(3):242–258, 2000.

[81] Leo Breiman, Jerome Friedman, Charles J Stone, and Richard A Olshen. *Classification and regression trees*. CRC press, 1984.

[82] J Ross Quinlan. C4. 5: Programming for machine learning. *Morgan Kauffmann*, 38, 1993.

[83] J. Ross Quinlan. Induction of decision trees. *Machine learning*, 1(1):81–106, 1986.

[84] J. Ross Quinlan. Simplifying decision trees. *International journal of man-machine studies*, 27(3):221–234, 1987.

[85] Manish Mehta, Rakesh Agrawal, and Jorma Rissanen. Sliq: A fast scalable classifier for data mining. *Advances in Database TechnologyEDBT'96*, pages 18–32, 1996.

[86] John Shafer, Rakesh Agrawal, and Manish Mehta. Sprint: A scalable parallel classi er for data mining. In *Proc. 1996 Int. Conf. Very Large Data Bases*, pages 544–555. Citeseer, 1996.

[87] Georg Langs, Bjoern H Menze, Danial Lashkari, and Polina Golland. Detecting stable distributed patterns of brain activation using gini contrast. *NeuroImage*, 56(2):497–507, 2011.

[88] Earl B Hunt, Janet Marin, and Philip J Stone. Experiments in induction. 1966.

[89] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[90] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. *The elements of statistical learning*, volume 1. Springer series in statistics New York, 2001.

[91] MATLAB. *version 8.5.0.197613 (R2015a*. The MathWorks Inc., Natick, Massachusetts, 2015.

[92] Robert E Schapire. The strength of weak learnability. *Machine learning*, 5(2):197–227, 1990.

[93] Larry Wasserman. *All of statistics: a concise course in statistical inference*. Springer Science & Business Media, 2013.

[94] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794. ACM, 2016.

[95] Jerome Friedman, Trevor Hastie, Robert Tibshirani, et al. Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). *The annals of statistics*, 28(2):337–407, 2000.

[96] Ian Goodfellow, Yoshua Bengio, Aaron Courville, and Yoshua Bengio. *Deep learning*, volume 1. MIT press Cambridge, 2016.

[97] Amirhessam Tahmassebi. ideeple: Deep learning in a flash. In *Disruptive Technologies in Information Sciences*, volume 10652. International Society for Optics and Photonics, 2018.

[98] Alexandre Abraham, Fabian Pedregosa, Michael Eickenberg, Philippe Gervais, Andreas Mueller, Jean Kossaifi, Alexandre Gramfort, Bertrand Thirion, and Gaël Varoquaux. Machine learning for neuroimaging with scikit-learn. *Frontiers in neuroinformatics*, 8:14, 2014.

[99] François Chollet et al. Keras, 2015.

[100] Krzysztof Gorgolewski, Christopher D Burns, Cindee Madison, Dav Clark, Yaroslav O Halchenko, Michael L Waskom, and Satrajit S Ghosh. Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in python. *Front Neuroinform*, 5, 08 2011.

[101] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *Journal of Machine Learning Research*, 12(Oct):2825–2830, 2011.

[102] Amirhessam Tahmassebi, Amir H. Gandomi, Ian McCann, Mieke HJ Schulte, Lianne Schmaal, Anna E. Goudriaan, and Anke Meyer-Baese. fmri smoking cessation classification using genetic programming. *Workshop on Data Science meets Optimization*, 2017.

[103] Amirhessam Tahmassebi, Amir H Gandomi, and Anke Meyer-Bäse. High performance gp-based approach for fmri big data classification. In *Proceedings of the Practice and Experience in Advanced Research Computing 2017 on Sustainability, Success and Impact*, page 57. ACM, 2017.

[104] Axel Riecker, Dirk Wildgruber, Klaus Mathiak, Wolfgang Grodd, and Hermann Ackermann. Parametric analysis of rate-dependent hemodynamic response functions of cortical and subcortical brain structures during auditorily cued finger tapping: a fmri study. *Neuroimage*, 18(3):731–739, 2003.

[105] Penelope A Lewis, AM Wing, PA Pope, P Praamstra, and RC Miall. Brain activity correlates differentially with increasing temporal complexity of rhythms during initialisation, synchronisation, and continuation phases of paced finger tapping. *Neuropsychologia*, 42(10):1301–1312, 2004.

[106] Amirhessam Tahmassebi, Amir H. Gandomi, Ian McCann, Mieke H. J. Schulte, Anna E. Goudriaan, and Anke Meyer-Baese. Deep learning in medical imaging: Fmri big data analysis via convolutional neural networks. In *Proceedings of the Practice and Experience on Advanced Research Computing*, PEARC '18, pages 85:1–85:4, New York, NY, USA, 2018. ACM.

[107] Jean-Luc Dreyer. New insights into the roles of microrna s in drug addiction and neuroplasticity. *Genome medicine*, 2(12):92, 2010.

[108] Alfred J Robison and Eric J Nestler. Transcriptional and epigenetic mechanisms of addiction. *Nature reviews neuroscience*, 12(11):623, 2011.

[109] Kenneth Blum, Tonia Werner, Stefanie Carnes, Patrick Carnes, Abdalla Bowirrat, John Giordano, Marlene-Oscar-Berman, and Mark Gold. Sex, drugs, and rock nroll: hypothesizing common mesolimbic activation as a function of reward gene polymorphisms. *Journal of psychoactive drugs*, 44(1):38–55, 2012.

[110] Christopher M Olsen. Natural rewards, neuroplasticity, and non-drug addictions. *Neuropharmacology*, 61(7):1109–1122, 2011.

[111] M Kaufmann, G Von Minckwitz, HD Bear, A Buzdar, P McGale, H Bonnefoi, M Colleoni, C Denkert, W Eiermann, R Jackesz, et al. Recommendations from an international expert panel on the use of neoadjuvant (primary) systemic treatment of operable breast cancer: new perspectives 2006. *Annals of Oncology*, 18(12):1927–1934, 2007.

[112] Richard G Abramson, Xia Li, Tamarya Lea Hoyt, Pei-Fang Su, Lori R Arlinghaus, Kevin J Wilson, Vandana G Abramson, A Bapsi Chakravarthy, and Thomas E Yankeelov. Early assessment of breast cancer response to neoadjuvant chemotherapy by semi-quantitative analysis of high-temporal resolution dce-mri: preliminary results. *Magnetic resonance imaging*, 31(9):1457–1464, 2013.

[113] Lori R Arlinghaus, Xia Li, Mia Levy, David Smith, E Brian Welch, John C Gore, and Thomas E Yankeelov. Current and future trends in magnetic resonance imaging assessments of the response of breast tumors to neoadjuvant chemotherapy. *Journal of oncology*, 2010, 2010.

[114] Ignacio Alvarez Illan, Amirhessam Tahmassebi, Javier Ramirez, Juan M Gorriz, Simon Y Foo, Katja Pinker-Domenig, and Anke Mayer-Baese. Machine learning for accurate differentiation of benign and malignant breast tumors presenting as non-mass enhancement. In *Computational Imaging III*, volume 10669, page 106690W. International Society for Optics and Photonics, 2018.

[115] A Tahmassebi, K Pinker-Domenig, G Wengert, T Helbich, Z Bago-Horvath, and A Meyer-Baese. Determining the importance of parameters extracted from multi-parametric mri in the early prediction of the response to neo-adjuvant chemotherapy in breast cancer. *Medical Imaging*, 2018.

[116] Katja Pinker-Domenig, Amirhessam Tahmassebi, Georg Wengert, Thomas H Helbich, Zsuzsanna Bago-Horvath, Elizabeth A Morris, and Anke Meyer-Baese. Magnetic resonance imaging of the breast and radiomics analysis for an improved early prediction of the response to neoadjuvant chemotherapy in breast cancer patients, 2018.

[117] Amirhessam Tahmassebi, Dat Ngo, Antonio Garcia, Encarnacin Castillo, Diego P Morales, Katja Pinker-Domenig, Mark Lobbes, and Anke Meyer-Bäse. Multi-level analysis of spatio-temporal features in non-mass enhancing breast tumors. In *Smart Biomedical and Physiological Sensor Technology XV*, volume 10662, page 106620H. International Society for Optics and Photonics, 2018.

[118] Melanie A Lindenberg, Anna Miquel-Cases, Valesca P Retèl, Gabe S Sonke, Jelle Wesseling, Marcel PM Stokkel, and Wim H van Harten. Imaging performance in guiding response to neoadjuvant therapy according to breast cancer subtypes: A systematic literature review. *Critical Reviews in Oncology/Hematology*, 112:198–207, 2017.

[119] Qiufang Liu, Chen Wang, Panli Li, Jianjun Liu, Gang Huang, and Shaoli Song. The role of 18f-fdg pet/ct and mri in assessing pathological complete response to neoadjuvant chemotherapy in patients with breast cancer: a systematic review and meta-analysis. *BioMed research international*, 2016, 2016.

[120] Frank G Zöllner, Gerald Weisser, Marcel Reich, Sven Kaiser, Stefan O Schoenberg, Steven P Sourbron, and Lothar R Schad. Ummperfusion: an open source software tool towards quantitative mri perfusion analysis in clinical routine. *Journal of digital imaging*, 26(2):344–352, 2013.

[121] Gaiane M Rauch, Beatriz Elena Adrada, Henry Mark Kuerer, Raquel FD van la Parra, Jessica WT Leung, and Wei Tse Yang. Multimodality imaging for evaluating response to neoadjuvant chemotherapy in breast cancer. *American Journal of Roentgenology*, 208(2):290–299, 2017.

[122] Leticia De Mattos-Arruda, Ronglai Shen, Jorge S Reis-Filho, and Javier Cortés. Translating neoadjuvant therapy into survival benefits: one size does not fit all. *Nature Reviews Clinical Oncology*, 13(9):566–579, 2016.

[123] Li-An Wu, Ruey-Feng Chang, Chiun-Sheng Huang, Yen-Shen Lu, Hong-Hao Chen, Jo-Yu Chen, and Yeun-Chung Chang. Evaluation of the treatment response to neoadjuvant chemotherapy in locally advanced breast cancer using combined magnetic resonance vascular maps and apparent diffusion coefficient. *Journal of Magnetic Resonance Imaging*, 42(5):1407–1420, 2015.

[124] Lenka Minarikova, Wolfgang Bogner, Katja Pinker, Ladislav Valkovič, Olgica Zaric, Zsuzsanna Bago-Horvath, Rupert Bartsch, Thomas H Helbich, Siegfried Trattnig, and Stephan Gruber. Investigating the prediction value of multiparametric magnetic resonance imaging at 3 t in response to neoadjuvant chemotherapy in breast cancer. *European radiology*, 27(5):1901–1911, 2017.

[125] Robert J Gillies, Paul E Kinahan, and Hedvig Hricak. Radiomics: images are more than pictures, they are data. *Radiology*, 278(2):563–577, 2015.

[126] Kavya Ravichandran, Nathaniel Braman, Andrew Janowczyk, and Anant Madabhushi. A deep learning classifier for prediction of pathological complete response to neoadjuvant chemotherapy from baseline breast dce-mri. In *Medical Imaging 2018: Computer-Aided Diagnosis*, volume 10575, page 105750C. International Society for Optics and Photonics, 2018.

[127] Elizabeth AM OFlynn, David Collins, James DArcy, Maria Schmidt, and Nandita M de Souza. Multi-parametric mri in the early prediction of response to neo-adjuvant chemotherapy in breast cancer: Value of non-modelled parameters. *European journal of radiology*, 85(4):837–842, 2016.

[128] Jennifer F De Los Santos, Alan Cantor, Keith D Amos, Andres Forero, Mehra Golshan, Janet K Horton, Clifford A Hudis, Nola M Hylton, Kandace McGuire, Funda Meric-Bernstam, et al. Magnetic resonance imaging as a predictor of pathologic response in patients treated with neoadjuvant systemic treatment for operable breast cancer. *Cancer*, 119(10):1776–1783, 2013.

[129] Yuji Hayashi, Hiroyuki Takei, Satoshi Nozu, Yoshihiro Tochigi, Akihiro Ichikawa, Naoki Kobayashi, Masafumi Kurosumi, Kenichi Inoue, Takashi Yoshida, Shigenori E Nagai, et al. Analysis of complete response by mri following neoadjuvant chemotherapy predicts pathological tumor responses differently for molecular subtypes of breast cancer. corrigendum in/ol/5/4/1433. *Oncology letters*, 5(1):83–89, 2013.

[130] Eun Sook Ko, Boo-Kyung Han, Rock Bum Kim, Eun Young Ko, Jung Hee Shin, Soo Yeon Hahn, Seok Jin Nam, Jeong Eon Lee, Se Kyung Lee, Young-Hyuck Im, et al. Analysis of factors that influence the accuracy of magnetic resonance imaging for predicting response after neoadjuvant chemotherapy in locally advanced breast cancer. *Annals of surgical oncology*, 20(8):2562–2568, 2013.

[131] Jian-Fei Fu, Hai-Long Chen, Jiao Yang, Cheng-Hao Yi, and Shu Zheng. Feasibility and accuracy of sentinel lymph node biopsy in clinically node-positive breast cancer after neoadjuvant chemotherapy: a meta-analysis. *PloS one*, 9(9):e105316, 2014.

[132] Nola M Hylton, Jeffrey D Blume, Wanda K Bernreuter, Etta D Pisano, Mark A Rosen, Elizabeth A Morris, Paul T Weatherall, Constance D Lehman, Gillian M Newstead, Sandra Polin, et al. Locally advanced breast cancer: Mr imaging for prediction of response to neoadjuvant chemotherapyresults from acrin 6657/i-spy trial. *Radiology*, 263(3):663–672, 2012.

[133] Katja Pinker, Gunther Grabner, Wolfgang Bogner, Stephan Gruber, Pavol Szomolanyi, Siegfried Trattnig, Gertraud Heinz-Peer, Michael Weber, Florian Fitzal, Ursula Pluschnig, et al. A combined high temporal and high spatial resolution 3 tesla mr imaging protocol for the assessment of breast lesions: initial results. *Investigative radiology*, 44(9):553–558, 2009.

[134] Sebastian Mika, Gunnar Ratsch, Jason Weston, Bernhard Scholkopf, and Klaus-Robert Mullers. Fisher discriminant analysis with kernels. In *Neural Networks for Signal Processing IX, 1999. Proceedings of the 1999 IEEE Signal Processing Society Workshop.*, pages 41–48. IEEE, 1999.

[135] Ronald A Fisher. The use of multiple measurements in taxonomic problems. *Annals of human genetics*, 7(2):179–188, 1936.

[136] Geoffrey McLachlan. *Discriminant analysis and statistical pattern recognition*, volume 544. John Wiley & Sons, 2004.

[137] Lars Buitinck, Gilles Louppe, Mathieu Blondel, Fabian Pedregosa, Andreas Mueller, Olivier Grisel, Vlad Niculae, Peter Prettenhofer, Alexandre Gramfort, Jaques Grobler, Robert Layton, Jake VanderPlas, Arnaud Joly, Brian Holt, and Gaël Varoquaux. API design for machine learning software: experiences from the scikit-learn project. In *ECML PKDD Workshop: Languages for Data Mining and Machine Learning*, pages 108–122, 2013.

[138] Stephan Dreiseitl and Lucila Ohno-Machado. Logistic regression and artificial neural network classification models: a methodology review. *Journal of biomedical informatics*, 35(5):352–359, 2002.

[139] Hsiang-Fu Yu, Fang-Lan Huang, and Chih-Jen Lin. Dual coordinate descent methods for logistic regression and maximum entropy models. *Machine Learning*, 85(1):41–75, 2011.

[140] Rong-En Fan, Kai-Wei Chang, Cho-Jui Hsieh, Xiang-Rui Wang, and Chih-Jen Lin. Liblinear: A library for large linear classification. *Journal of machine learning research*, 9(Aug):1871–1874, 2008.

[141] Mark Schmidt, Nicolas Le Roux, and Francis Bach. Minimizing finite sums with the stochastic average gradient. *Mathematical Programming*, 162(1-2):83–112, 2017.

[142] Ciyou Zhu, Richard H Byrd, Peihuang Lu, and Jorge Nocedal. Algorithm 778: L-bfgs-b: Fortran subroutines for large-scale bound-constrained optimization. *ACM Transactions on Mathematical Software (TOMS)*, 23(4):550–560, 1997.

[143] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning representations by back-propagating errors. *nature*, 323(6088):533–536, 1986.

[144] David E Rumelhart, Geoffrey E Hinton, Ronald J Williams, et al. Learning representations by back-propagating errors. *Cognitive modeling*, 5(3):1, 1988.

[145] Léon Bottou. Large-scale machine learning with stochastic gradient descent. In *Proceedings of COMPSTAT'2010*, pages 177–186. Springer, 2010.

[146] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[147] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. Overview of supervised learning. In *The elements of statistical learning*, pages 9–41. Springer, 2009.

[148] Gilles Louppe. Understanding random forests: From theory to practice. *arXiv preprint arXiv:1407.7502*, 2014.

BIOGRAPHICAL SKETCH

Amirhessam Tahmassebi was born on April 13, 1988 in Tehran, Iran. He received his B.Sc. and M.Sc. degrees in Physics from The University of Tehran, Iran, and The University of Akron, Ohio, in 2010, and 2015, respectively. In August of 2015, he has started his PhD study in the Department of Scientific Computing at Florida State University, Tallahassee, Florida. In Spring of 2017, he has passed the PhD preliminary exam with a score of 95.1%. During his PhD study, he published 1 book chapter and more than 30 peer-reviewed papers in prestigious journals and conferences. In addition to receiving numerous awards and travel grants, in Spring of 2018, he was awarded the Florida State University Graduate Student Research and Creativity Award in the natural and physical sciences, including mathematics and engineering category. The Graduate Student Research and Creativity Award recognizes outstanding contributions to research and creative endeavors at the Celebration of Graduate Student Excellence held each spring. Awards are made to students in three categories: the natural and physical sciences, including mathematics and engineering, humanities and arts, and social and behavioral sciences. His research interests include data mining, machine learning, deep learning, medical imaging, scientific computing, and genetic programming.

Publications

1. **Tahmassebi, Amirhessam**, and Amir H. Gandomi. Genetic Programming Based on Error Decomposition: A Big Data Approach, In Book: Genetic Programming Theory and Practice XV, Springer, 2018.

2. **Tahmassebi, Amirhessam**, Amir H. Gandomi, Mieke HJ Schulte, Lianne Schmaal, Anna E. Goudriaan, and Anke Meyer-Baese. Optimized Naive-Bayes and Decision Tree Approaches for fMRI Smoking Cessation Classification, Journal of Complexity, Hindawi, 2018.

3. **Tahmassebi, Amirhessam**. iDeepLe: Deep Learning in a Flash." In SPIE Defense + Security, International Society for Optics and Photonics, 2018.

4. **Tahmassebi, Amirhessam**, Amir H. Gandomi, Ian McCann, Mieke H.J. Schulte, Lianne Schmaal, Anna E. Goudriaan, and Anke Meyer-Baese. "An evolutionary approach for fMRI big data classification." In Evolutionary Computation (CEC), 2017 IEEE Congress on, pp. 1029-1036. IEEE, 2017.

5. **Tahmassebi, Amirhessam**, Amir H. Gandomi, and Anke Meyer-Baese. "High Performance GP-Based Approach for fMRI Big Data Classification." In Proceedings of the Practice and Experience in Advanced Research Computing 2017 on Sustainability, Success and Impact, p. 57. ACM, 2017.

6. **Tahmassebi, Amirhessam**, Amir H. Gandomi, Ian McCann, Mieke H.J. Schulte, Anna E. Goudriaan, and Anke Meyer-Baese. "Deep Learning in Medical Imaging: fMRI Big Data Analysis via Convolutional Neural Networks." In Proceedings of the Practice and Experience in Advanced Research Computing 2018, ACM, 2018.

7. **Tahmassebi, Amirhessam**, Amir H. Gandomi, Ian McCann, Mieke HJ Schulte, Lianne Schmaal, Anna E. Goudriaan, and Anke Meyer-Baese. "fMRI Smoking Cessation Classification Using Genetic Programming." In Workshop on Data Science meets Optimization, 2017.

8. **Tahmassebi, Amirhessam**, Katja Pinker-Domenig, Georg Wengert, Marc Lobbes, Andreas Stadlbauer, Norelle C. Wildburger, Francisco J. Romeroa, and Meyer-Baese, Anke. "The driving regulators of the connectivity protein network of brain malignancies." In SPIE Commercial + Scientific Sensing and Imaging, pp. 1021605-1021605. International Society for Optics and Photonics, 2017.

9. **Tahmassebi, Amirhessam**, Katja Pinker-Domenig, Georg Wengert, Marc Lobbes, Andreas Stadlbauer, Francisco J. Romeroa, Diego P. Morales et al. "Dynamical graph theory networks

techniques for the analysis of sparse connectivity networks in dementia." In SPIE Commercial+ Scientific Sensing and Imaging, pp. 1021609-1021609. International Society for Optics and Photonics, 2017.

10. **Tahmassebi, Amirhessam**, Katja Pinker-Domenig, Georg Wengert, Anke Meyer-Baese. "Determining the importance of parameters extracted from multi-parametric MRI in the early prediction of the response to neo-adjuvant chemotherapy in breast cancer." In SPIE Medical Imaging, International Society for Optics and Photonics, 2018.

11. **Tahmassebi, Amirhessam**, Katja Pinker-Domenig, Georg Wengert, Anke Meyer-Baese. "Determining disease evolution driver nodes in dementia networks." In SPIE Medical Imaging, International Society for Optics and Photonics, 2018.

12. **Tahmassebi, Amirhessam**, Anke Meyer-Baese, Georg Wengert, Katja Pinker-Domenig. Radiomics with MRI for early prediction of the response to neo-adjuvant chemotherapy in breast cancer patients." In ECR, 2018.

13. **Tahmassebi, Amirhessam**, Alper Buldum, Michael Avon, and Graham Kelly. "Fluid flow through graphene based nanostructures. Journal of Applied Physics, 2018.

14. **Tahmassebi, Amirhessam**, Amir H. Gandomi, and Anke Meyer-Baese. "Stock Risk Assessment via Multi-Objective Genetic Programming." Journal of Postdoctoral Research 6.3, 2018.

15. **Tahmassebi, Amirhessam**, Amir H. Gandomi. Building Energy Consumption Forecast using Multi-Objective Genetic Programming, Journal of Measurements, Elsevier 2018.

16. **Tahmassebi, Amirhessam**, Amir H. Gandomi, Simon Fong, Anke Meyer-Baese, Simon Y. Foo. Multi-stage optimization of deep model: A case study on ground motion modeling, Plos one, 2018.

17. **Tahmassebi, Amirhessam**, Dat Ngo, Antonio Garcia, Encarnacin Castillo, Diego P. Morales, Katja Pinker-Domenig, Mark Lobbes, and Anke Meyer-Baese. "Multi-level analysis of spatio-temporal features in non-mass enhancing breast tumors." In SPIE Defense + Security, International Society for Optics and Photonics, 2018.

18. **Tahmassebi, Amirhessam**, Amir H. Gandomi, and Anke Meyer-Baese. "A Pareto Front Based Evolutionary Model for Airfoil Self-Noise Prediction." IEEE Congress on Evolutionary Computation, 2018.

19. **Tahmassebi, Amirhessam**, Anke Meyer-Baese, Georg J. Wengert, Thomas H. Helbich, and Katja Pinker-Domenig. "Radiomics with MRI for early prediction of the response to neo-adjuvant chemotherapy in breast cancer patients." Insights into Imaging, Springer, 2018.

20. **Tahmassebi, Amirhessam**, Amir H. Gandomi, and Anke Meyer-Baese. "An Evolutionary Online Framework for MOOC Performance using EEG Data." IEEE Congress on Evolutionary Computation, 2018.

21. **Tahmassebi, Amirhessam**, and Alper Buldum. "Fluid flow calculations of Graphene Composites." In APS March Meeting. 2015.

22. **Tahmassebi, Amirhessam**. "Fluid flow through carbon nanotubes and graphene based nanostructures." MS Thesis, The University of Akron, 2015.

23. **Tahmassebi, Amirhessam**, Amir H. Gandomi. Naturally Inspired Algorithms: GP & ANN, The National Aeronautics and Space Administration (NASA) Proposals, 2018.

24. Mohebali, Behshad, **Amirhessam Tahmassebi**, Amir H. Gandomi, Anke Meyer-Baese, and Simon Y. Foo. A Scalable Communication Abstraction Framework for Internet of Things Applications using Raspberry Pi." In SPIE Defense + Security, International Society for Optics and Photonics, 2018.

25. McCann, Ian, **Amirhessam Tahmassebi**, Simon Y. Foo, Gordon Erlebacher, and Anke Meyer-Baese. NewsAnalyticalToolkit: an online natural language processing platform to analyze news." In SPIE Defense + Security, International Society for Optics and Photonics, 2018.

26. Illan, Ignacio, **Tahmassebi, Amirhessam**, Javier Ramirez, Juan M. Gorriz, Simon Y. Foo, Katja Pinker-Domenig, Anke Meyer-Baese. "Machine learning for accurate differentiation of benign and malignant breast tumors presenting as non-mass enhancement." In SPIE Defense + Security, International Society for Optics and Photonics, 2018.

27. Pinker-Domenig, Katja, **Amirhessam Tahmassebi**, Georg J. Wengert, Thomas Helbich, Zsuzsanna Bago-Horvath, Elizabeth A. Morris, and Anke Meyer-Baese. "Magnetic resonance imaging of the breast and radiomics analysis for an improved early prediction of the response to neoadjuvant chemotherapy in breast cancer patients." In Proceedings of the 109th Annual Meeting of the American Association for Cancer Research, AACR.

28. Yazicioglu, Yasin, Katja Pinker-Domenig, **Amirhessam Tahmassebi**, and Anke Meyer-Baese. "Determining leader nodes in dementia networks." Insights into Imaging, Springer, 2018.

29. Romero, Fransisco, Santiago Juarez, Inmaculada Ortiz-Gomez, Diego P. Morales, Alfonso Salinas-Castillo, Encarnacion Castillo, Antonio Garcia, **Amirhessam Tahmassebi**, and Anke Meyer-Baese. "Reconfigurable instrument for measuring variations of capacitors dielectric: an application to olive oil quality monitoring."In SPIE Commercial+ Scientific Sensing and Imaging, International Society for Optics and Photonics, 2018.

30. Toral, Victor, Salvador Criado, Fransisco Romero, Diego P. Morales, Encarnacion Castillo, Antonio Garcia, **Amirhessam Tahmassebi**, and Anke Meyer-Baese. "Wearable biosignal acquisition system for decision aid." In SPIE Commercial+ Scientific Sensing and Imaging, International Society for Optics and Photonics, 2018.

31. Romero, Francisco J., Diego P. Morales, Encarnacion Castillo, Antonio Garcia, **Amirhessam Tahmassebi**, and Anke Meyer-Baese. "Reconfigurable wearable to monitor physiological variables and movement." In SPIE Commercial+ Scientific Sensing and Imaging, pp. 1021608-1021608. International Society for Optics and Photonics, 2017.