MULTI CONTRAST MAGNETIC RESONANCE IMAGING BASED DISCRIMINANTS FOR NEURO BEHCET SYNDROME

by

Yıldız Tüzün

B.S., in Medical Engineering, Acibadem University, 2022

Submitted to the Institute of Biomedical Engineering
in partial fulfillment of the requirements
for the degree of
Master of Science
in
Biomedical Engineering

Boğaziçi University
2025

MULTI CONTRAST MAGNETIC RESONANCE IMAGING BASED DISCRIMINANTS FOR NEURO BEHCET SYNDROME

APPROVED BY:

Prof. Dr. Alpay Özcan ..... (Thesis Advisor)

Prof. Dr. Esin Öztürk Işık ..... (Thesis Co-advisor)

Assist. Prof. Dr. Pınar Özbay .....

Assist. Prof. Dr. Mustafa Kemal Ruhi .....

Assist. Prof. Dr. Engin Baysoy .....

DATE OF APPROVAL: 28 August 2025

ACKNOWLEDGMENTS

I would like to express my sincere gratitude to my thesis advisor, Prof. Dr. Alpay Özcan for being an excellent guide during my master of science program. His support has been instrumental in helping me find direction and shape my career, and it continues to have a lasting impact on my professional development. I also would like to thank Prof. Dr. Esin Öztürk Işık for giving me the opportunity to work with her. I would also like to thank Prof. Dr. Alp Dinçer for his support and valuable time for helping me during this study. I am also thankful to Dr. Abdullah Arcan for answering my questions and helping me. I also respectfully acknowledge the contribution of Prof. Dr. Uğur Uygunoğlu and Prof. Dr. Aksel Siva.

I am thankful to my laboratory friends, especially Hamit Alp Cömert, for his support and contributions during this study. I also like to thank my other colleagues in Systems Science and Mathematics laboratory; Adeola Aderemi,İrem Nur Çevikcan, and Merve Nur Bayer for their support and friendship.

Essentially, my deepest appreciation goes to my family for their endless support, patience, and love during my master of science. I would like to thank my mother Figen Tüzün and my father Cihat Tüzün for their infinite support and curiosity for this long journey. I could not have accomplished this without their support.

I also extend my deepest gratitude to my beloved partner, Öner Aydın, who stood by me through every setback and in moments when I felt I could not carry on. This thesis would not have been completed without his support.

This work was funded by the Council of Higher Education of Türkiye with YÖK–ADP Grant Number 50006 'Machine Learning Based Differentiation of Neuro-Behcet Syndrome with Multi-Contrast Magnetic Resonance Imaging' project.

ACADEMIC ETHICS AND INTEGRITY STATEMENT

I, Yıldız Tüzün, hereby certify that I am aware of the Academic Ethics and Integrity Policy issued by the Council of Higher Education YÖK and I fully acknowledge all the consequences due to its violation by plagiarism or any other way.

Name :

Signature:

Date:

ABSTRACT

MULTI CONTRAST MAGNETIC RESONANCE IMAGING BASED DISCRIMINANTS FOR NEURO BEHCET SYNDROME

Neuro–Behcet Syndrome (NBS) is extremely rare worldwide but has a higher prevalence in Türkiye, allowing meaningful cohort formation. The goal of this work is the characterization of white matter (WM) integrity by analyzing diffusion tensor imaging (DTI) metrics in NBS. Retrospective data from 14 patients and a control cohort of 38 healthy individuals were analyzed through WM segmentation in DTI data. The mean of DTI variables in WM of both groups were compared with Mann–Whitney U Test. λ_2 and apparent diffusion coefficient (ADC) showed the highest statistical significant differences between the NBS and control cohorts. These indicate potentially that in addition to damage along the axons, the WM integrity is more significantly compromised in the axons transversally. Accordingly, WM DTI characterization in this study might be considered as the first step for investigating differentiating quantitative MR biomarkers for this family of neurological pathologies. Machine Learning (ML) results indicate that combining multiple diffusion metrics yields superior classification accuracy in distinguishing NBS patients from healthy controls compared to using individual variables alone. In particular ADC, even when paired with less discriminative variables, exhibited outstanding classification performance. The highest classification performance, achieving 100% accuracy, was obtained with λ_3 and ADC. Notably, these results were achieved with a lower number of components, indicating that these DTI variables provide highly discriminative power with minimal feature complexity. This thesis establishes a foundation for future work aimed at enhancing classification accuracy through larger cohorts, inclusion of NBS subtypes, and personalized ML based on individual diffusion profiles.

Keywords: Neuro–Behcet Syndrome, White Matter Integrity, Diffusion Tensor Imaging, Biomarker, Machine Learning, Classification.

ÖZET

NÖRO–BEHÇET SENDROMU İÇİN YÜKSEK KONTRASTLI MANYETİK REZONANS GÖRÜNTÜLEME TABANLI AYRIŞTIRICILAR

Nöro–Behçet Sendromu (NBS) dünya genelinde son derece nadir görülmesine rağmen, Türkiye'de daha yüksek bir yaygınlığa sahiptir; bu durum anlamlı bir kohort oluşturulmasına olanak tanımaktadır. Bu çalışmanın amacı, beyaz madde (BM) bütünlüğünün difüzyon tensör görüntüleme (DTG) aracılığıyla karakterize edilmesidir. Geriye dönük elde edilmiş 14 NBS hastası ve 38 sağlıklıdan oluşan kontrol kohortuna ait DTG verileri BM segmentasyonu ile analiz edilmiştir. Her iki grubun BM içindeki DTG değişkenlerinin ortalamaları Mann–Whitney U testi ile karşılaştırılmıştır. Elde edilen bulgulara göre, λ_2 ve görünür difüzyon katsayısı (GDK) NBS ve kontrol grupları arasında istatistiksel olarak anlamlı farklılıklar göstermiştir. Bu durum, akson boyunca meydana gelen hasara ek olarak, aksonların transversal yönünde BM bütünlüğünün daha belirgin şekilde bozulduğuna işaret etmektedir. Dolayısıyla, bu çalışmada gerçekleştirilen BM DTG karakterizasyonu, bu nörolojik hastalık ailesi için ayırt edici kantitatif manyetik rezonans (MR) biyobelirteçlerinin araştırılmasında ilk adım olarak değerlendirilebilir. Makine öğrenimi (MÖ) sonuçları, birden fazla difüzyon metriğinin bir arada kullanılmasının, yalnızca bazı tekil değişkenlerin kullanımına kıyasla NBS hastaları ile sağlıklı kontrolleri ayırt etmede yüksek sınıflandırma doğruluğu sağladığını göstermiştir. Bireysel değişkenler kombinasyon olarak da denenmiştir. En yüksek sınıflandırma başarımı, λ_3 ve GDK ile elde edilmiştir. Gelecekte, daha geniş hasta kohortları kullanılarak sınıflandırma başarımının artırılmasına ve bireysel difüzyon profillerine dayalı kişiselleştirilmiş MÖ algoritmalarının uygulanmasına odaklanacaktır.

Anahtar Sözcükler: Nöro–Behçet Sendromu, Ak Madde Bütünlüğü, Difüzyon Tensör Görüntüleme, Biyobelirteç, Makine Öğrenimi, Sınıflandırma.

TABLE OF CONTENTS

ACKNOWLEDGMENTS . . . iii
ACADEMIC ETHICS AND INTEGRITY STATEMENT . . . iv
ABSTRACT . . . v
ÖZET . . . vi
LIST OF FIGURES . . . ix
LIST OF TABLES . . . xii
LIST OF SYMBOLS . . . xiii
LIST OF ABBREVIATIONS . . . xiv
1. INTRODUCTION . . . 1
1.1 Behcet Syndrome . . . 1
1.2 Neuro Behcet Syndrome . . . 1
1.3 Central nervous system symptoms . . . 2
1.3.1 Parenchymal NBS . . . 2
1.3.2 Neuro–psychiatric BD . . . 2
1.3.3 Myelitis . . . 3
1.3.4 Headaches . . . 3
1.3.5 Non–parenchymal NBS . . . 3
1.3.6 Peripheral nervous system symptoms . . . 4
1.4 Diagnosis of Neuro–Behcet Syndrome . . . 4
1.5 Neuroimaging in Neuro–Behcet Syndrome . . . 4
1.6 Diffusion Tensor Imaging . . . 6
1.7 Machine Learning in Neuro–Behcet Syndrome . . . 9
1.8 Literature Review . . . 10
1.9 Research Gap . . . 12
2. MATERIALS AND METHODS . . . 14
2.1 Medical Data . . . 14
2.2 Imaging Data . . . 15
2.3 White Matter Segmentation . . . 15
2.4 Coregistration . . . 17

2.5 DTI Variables . . . 18
2.6 Statistical Analysis . . . 19
2.7 Classification . . . 19
2.8 Histogram Based Feature Extraction . . . 37
2.9 Histogram Normalization . . . 39
2.10 Mean Centering of Histogram Features . . . 39
2.11 DTI Variable Combination . . . 40
2.12 Classification Models . . . 41
2.13 Classifier Performance . . . 41
3. RESULTS . . . 43
3.1 Statistical Results . . . 43
3.2 Machine Learning Results . . . 46
4. DISCUSSION . . . 73
5. CONCLUSION . . . 77
5.1 List of publications produced from the thesis . . . 79
APPENDIX A. MACHINE LEARNING MODELS . . . 80
A.1 Support Vector Machines . . . 80
A.2 Decision Trees . . . 81
A.3 K–Nearest Neighbors . . . 82
A.4 Naive Bayes . . . 82
A.5 Ensemble Methods . . . 83
A.6 Kernel Methods . . . 83
A.7 Neural Network . . . 84
A.8 Logistic Regression . . . 85
REFERENCES . . . 86

LIST OF FIGURES

Figure 1.1	The depiction of water molecules diffusing equally in all directions in an isotropic environment (A), and more along one direction in an anisotropic environment (B).	6
Figure 2.1	T2W image and WM mask derived from T2W: (A) T2W image; (B) WM mask obtained from T2W; (C) WM mask overlaid on the T2W anatomical image.	16
Figure 2.2	B_0 and coregistered masks: B_0 image (D); T2 mask overlaid on B_0 (E); coregistered T2 mask on T2WI (F).	18
Figure 2.3	30–bin histogram distribution of λ_1 in NBS and Healthy WM.	20
Figure 2.4	60–bin histogram distribution of λ_1 in NBS and Healthy WM.	20
Figure 2.5	90–bin histogram distribution of λ_1 in NBS and Healthy WM	21
Figure 2.6	120–bin histogram distribution of λ_1 in NBS and Healthy WM.	21
Figure 2.7	30–bin histogram distribution of λ_2 in NBS and Healthy WM.	22
Figure 2.8	60–bin histogram distribution of λ_2 in NBS and Healthy WM.	22
Figure 2.9	90–bin histogram distribution of λ_2 in NBS and Healthy WM.	23
Figure 2.10	120–bin histogram distribution of λ_2 in NBS and Healthy WM.	23
Figure 2.11	30–bin histogram distribution of λ_3 in NBS and Healthy WM.	24
Figure 2.12	60–bin histogram distribution of λ_3 in NBS and Healthy WM.	24
Figure 2.13	90–bin histogram distribution of λ_3 in NBS and Healthy WM.	25
Figure 2.14	120–bin histogram distribution of λ_3 in NBS and Healthy WM.	25
Figure 2.15	30–bin histogram distribution of ADC in NBS and Healthy WM.	26
Figure 2.16	60–bin histogram distribution of ADC in NBS and Healthy WM.	26
Figure 2.17	90–bin histogram distribution of ADC in NBS and Healthy WM.	27
Figure 2.18	120–bin histogram distribution of ADC in NBS and Healthy WM.	27
Figure 2.19	30–bin histogram distribution of RD in NBS and Healthy WM.	28
Figure 2.20	60–bin histogram distribution of RD in NBS and Healthy WM.	28
Figure 2.21	90–bin histogram distribution of RD in NBS and Healthy WM.	29
Figure 2.22	120–bin histogram distribution of RD in NBS and Healthy WM.	29
Figure 2.23	30–bin histogram distribution of FA in NBS and Healthy WM.	30

Figure 2.24	60–bin histogram distribution of FA in NBS and Healthy WM.	30
Figure 2.25	90–bin histogram distribution of FA in NBS and Healthy WM.	31
Figure 2.26	120–bin histogram distribution of FA in NBS and Healthy WM.	31
Figure 2.27	30–bin histogram distribution of RA in NBS and Healthy WM.	32
Figure 2.28	60–bin histogram distribution of RA in NBS and Healthy WM.	32
Figure 2.29	90–bin histogram distribution of RA in NBS and Healthy WM.	33
Figure 2.30	120–bin histogram distribution of RA in NBS and Healthy WM	33
Figure 2.31	1000–bin histogram distribution of λ_1 in NBS and Healthy WM.	34
Figure 2.32	1000–bin histogram distribution of λ_2 in NBS and Healthy WM.	34
Figure 2.33	1000–bin histogram distribution of λ_3 in NBS and Healthy WM.	35
Figure 2.34	1000–bin histogram distribution of ADC in NBS and Healthy WM.	35
Figure 2.35	1000–bin histogram distribution of RD in NBS and Healthy WM.	36
Figure 2.36	1000–bin histogram distribution of FA in NBS and Healthy WM.	36
Figure 2.37	1000–bin histogram distribution of RA in NBS and Healthy WM.	37
Figure 3.1	Box plots comparing the mean values of DTI metrics between the NBS and control cohorts, extracted from white matter regions. Panels (a)–(c) show the eigenvalues $\lambda_1, \lambda_2, \lambda_3$; (d) ADC; (e) RD; and (f), (g) the anisotropy indices FA and RA, respectively. Reported p-values indicate the statistical significance of the observed group differences.	45
Figure 3.2	Confusion matrix for λ_1 (Linear SVM).	59
Figure 3.3	ROC curve for λ_1 (Linear SVM).	60
Figure 3.4	Confusion matrix for λ_2 (Linear SVM).	60
Figure 3.5	ROC curve for λ_2 (Linear SVM).	61
Figure 3.6	Confusion matrix for λ_3 (Narrow Neural Network).	61
Figure 3.7	ROC curve for λ_3 (Narrow Neural Network).	62
Figure 3.8	Confusion matrix for ADC (Linear SVM).	62
Figure 3.9	ROC curve for ADC (Linear SVM).	63
Figure 3.10	Confusion matrix for RD (RUSBoosted Trees).	63
Figure 3.11	ROC curve for RD (RUSBoosted Trees).	64
Figure 3.12	Confusion matrix for FA (Quadratic SVM).	64
Figure 3.13	ROC curve for FA (Quadratic SVM).	65

Figure 3.14	Confusion matrix for RA (Coarse Tree).	65
Figure 3.15	ROC curve for RA (Coarse Tree).	66
Figure 3.16	Confusion matrix for $\lambda_1,\ \lambda_2$ (Linear SVM).	66
Figure 3.17	ROC curve for $\lambda_1,\ \lambda_2$ (Linear SVM).	67
Figure 3.18	Confusion matrix for $\lambda_1,\ \lambda_3$ (Linear SVM).	67
Figure 3.19	ROC curve for $\lambda_1,\ \lambda_3$ (Linear SVM).	68
Figure 3.20	Confusion matrix for $\lambda_2,\ \lambda_3$ (Linear SVM).	68
Figure 3.21	ROC curve for $\lambda_2,\ \lambda_3$ (Linear SVM).	69
Figure 3.22	Confusion matrix for $\lambda_1,\ \lambda_2,\ \lambda_3$ (Linear SVM).	69
Figure 3.23	ROC curve for $\lambda_1,\ \lambda_2,\ \lambda_3$ (Linear SVM).	70

LIST OF TABLES

Table 2.1	Patient Demographics.	14
Table 2.2	Types of NBS Observed in the Cohort.	14
Table 3.1	Statistical significance of group differences for DTI metrics between NBS and control cohorts.	44
Table 3.2	Validation accuracy for λ_1.	46
Table 3.3	Validation accuracy for λ_2.	47
Table 3.4	Validation accuracy for λ_3.	47
Table 3.5	Validation accuracy for ADC.	49
Table 3.6	Validation accuracy for RD.	50
Table 3.7	Validation accuracy for FA.	51
Table 3.8	Validation accuracy for RA.	51
Table 3.9	Validation accuracy for $\lambda_1,\ \lambda_2$.	53
Table 3.10	Validation accuracy for $\lambda_1,\ \lambda_3$.	55
Table 3.11	Validation accuracy for $\lambda_2,\ \lambda_3$.	56
Table 3.12	Validation accuracy for $\lambda_1,\ \lambda_2,\ \lambda_3$.	58

LIST OF SYMBOLS

D	Diffusion tensor (3×3, symmetric)
D_{xx}, D_{yy}, D_{zz}	Diagonal elements of D
D_{xy}, D_{xz}, D_{yz}	Off–diagonal elements of D
B_0	Non–diffusion–weighted MR image
x	DTI scalar ($\lambda_1, \lambda_2, \lambda_3$, ADC, RD, FA, RA)
i	Subject index, $i \in \{1, \ldots, N\}$
N	Total number of subjects
k	WM voxel index for subject i
K_i	Number of WM voxels for subject i
n_{bins}	Number of histogram bins
h_i	Histogram feature vector of subject i
h_{mean}	Mean histogram vector
h_{ci}	Centered histogram ($h_i - h_{\text{mean}}$)
Z	Feature–label matrix
y_i	Class label (NBS = 1, Control = 0)
t	Threshold value
$\bar{x}$	Sample mean
V	Sample Variance
ϕ	Elevation angle (spherical coords.)
θ	Azimuthal angle (spherical coords.)
$\lambda_1, \lambda_2, \lambda_3$	Eigenvalues of D ($\lambda_1 \geq \lambda_2 \geq \lambda_3$)

LIST OF ABBREVIATIONS

ADC	Apparent Diffusion Coefficient
AD	Axial Diffusivity
AI	Artificial Intelligence
ANTs	Advanced Normalization Tools
AUC	Area Under the ROC Curve
BD	Behcet Disease
B0	Non–diffusion–weighted MR Image
CNS	Central Nervous System
CSF	Cerebrospinal Fluid
CVT	Cerebral Venous Thrombosis
DICOM	Digital Imaging and Communications in Medicine
DTE	Diffusion Tensor Eigenvalues ($\lambda_1, \lambda_2, \lambda_3$)
DTI	Diffusion Tensor Imaging
DWI	Diffusion–Weighted Imaging
EPI	Echo Planar Imaging
FA	Fractional Anisotropy
FLAIR	Fluid–Attenuated Inversion Recovery
GM	Gray Matter
HLA–B51	Human Leukocyte Antigen B51
IL–6	Interleukin–6
IRB	Institutional Review Board
MD	Mean Diffusivity (see ADC)
ML	Machine Learning
MLAs	Machine Learning Algorithms
MIPAV	Medical Image Processing, Analysis and Visualization
MRA	Magnetic Resonance Angiography
MRI	Magnetic Resonance Imaging
MRV	Magnetic Resonance Venography

MRS	Magnetic Resonance Spectroscopy
MS	Multiple Sclerosis
NBS	Neuro–Behcet Syndrome
NIfTI	Neuroimaging Informatics Technology Initiative
NMOSD	Neuromyelitis Optica Spectrum Disorder
PACS	Picture Archiving and Communication System
PCA	Principal Component Analysis
RA	Relative Anisotropy
RD	Radial Diffusivity
ROC	Receiver Operating Characteristic
SVM	Support Vector Machine
SWI	Susceptibility–Weighted Imaging
TE	Echo Time
TR	Repetition Time
T1WI	T1–Weighted Imaging
T2WI	T2–Weighted Imaging
VOI	Volume of Interest
V	Sample variance
WM	White Matter

1. INTRODUCTION

1.1 Behcet Syndrome

Behcet disease (BD) or Behcet Syndrome is a rare disease with a prevalence in multi–inflammatory systemic vasculitis in which all vessel sizes are affected in the venous and arterial system, small or large. It is associated with recurrent oral and genital ulcers, skin lesions, and uveitis. Although mainly characterized by recurrent oral and genital ulcers, it can affect any part of the organ system, including the gastrointestinal system (GI) and the central nervous system (CNS) [1]. Other symptoms may include vascular, gastrointestinal, and neurological involvement. Genetic factors, infections, and immune system processes are thought to contribute to the development of BD. As there is no definitive test for BD, its diagnosis is primarily based on identifying physical manifestations of the disease. BD is also called the 'Silk Route' disease, as it is mostly dominant in areas stretching along the Mediterranean, the Middle East, and the Far East. Although the exact cause of the disease is still not known and several factors in the emergence of the disease have been identified, the International Study Group currently uses up–to–date criteria to detect BD, which was proposed in 1990 and reviewed in 2006 [2].

1.2 Neuro Behcet Syndrome

Neurological involvement of Behcet's syndrome is called NBS (Neuro–Behcet Syndrome). It occurs in about 10-30% of individuals with BD, but the prevalence can be influenced by several factors, such as regional or ethnic differences, diagnostic tools, and the treatment methods used. It is more commonly seen in young men and typically develops within 5 years of the disease onset [3].

NBS is classified into two main phenotypes; a CNS inflammatory parenchy-

mal condition that is intra–axial and non–parenchymal which is extra–axial and primarily involves intracranial vascular structures, particularly the venous dural sinuses. Parenchymal NBS (p-NBS) is more common than the other form and commonly affects the brainstem, spinal cord, and cerebral hemisphere, while vascular structures in the brain are affected and cause cerebral venous thrombosis (CVT) and aneurysm in non–parenchymal NBS (np–NBS) form [4]. The most common findings are headache and papilledema which is optic disc swelling [5].

1.3 Central nervous system symptoms

1.3.1 Parenchymal NBS

The brainstem and basal ganglia are the primary regions affected, showing diverse presentations depending on whether the onset is acute, subacute, or progressive, and whether lesions are focal or multifocal. Common symptoms of parenchymal NBS are paresthesia, dysesthesia, numbness, and hemiparesis, with pyramidial signs [6]. Clinical signs include central cranial nerve palsy, cerebellar ataxia, ophthalmoplegia, and hearing loss [7]. Chronic progressive NBS may also lead to pseudobulbar palsy. Seizures are another common feature, often arising from parenchymal irritation or cerebral thrombosis, including tonic–clonic seizures, partial seizures, myoclonus, and epilepsia. In a cohort of 200 Turkish patients with NBS, only 6% had movement disorders, though hemichorea, hemiballismus, and hemidystonial [8].

1.3.2 Neuro–psychiatric BD

Several reports have highlighted neuropsychological impairments in BS patients, with frontal lobe involvement primarily affecting memory and executive functions. Psychiatric symptoms, including psychosis, manic episodes, and depression, have also been reported, although they are rarely linked to limbic system involvement [9]. Therefore, it is sometimes challenging to differentiate between the psychiatric manifestations of

BS and reactive symptoms [10].

1.3.3 Myelitis

Myelitis in NBS is associated with the spinal involvement. Spinal involvement is less common but associated with a poor prognosis. The lesions typically involve transverse myelitis, often spanning more than three vertebrae. Anterior spinal artery syndrome is barely seen in NBS. While isolated transverse myelitis can occur in BD. It is crucial to differentiate these patients from those with neuromyelitis optica spectrum disorders (NMOSD). Clinical manifestations of spinal NBS include paraparesis, bladder dysfunction, and sensory impairments [11],[12].

1.3.4 Headaches

Headaches are highly prevalent symptom for NBS patients. The headache may result from parenchymal NBD, vascular NBD, increased intracranial pressure without sinus venous thrombosis, meningitis, ocular inflammation or uveitis, or even isolated headaches unrelated to NBD. Severe headaches or any recent changes in pain patterns should be treated as warning signs and require further examination [13],[14].

1.3.5 Non–parenchymal NBS

CVT is the most frequent manifestation of the non–parenchymal type, and it can occur in the superior sagittal sinus, transverse sinuses, and basal vein of Rosenthal which are the most commonly affected areas [15]. Multiple sinus involvement, as well as cerebral infarction, may also occur [16]. Interestingly, BD patients with CVT demonstrate a higher prevalence of extracerebral vascular involvement. CVT symptoms are often non–specific, with headache being the most common, along with nausea/vomiting, papilledema, focal neurological deficits, and seizures [17].

1.3.6 Peripheral nervous system symptoms

Although peripheral neuropathy is uncommon, focal myositis in BD seems to be associated with the systemic vasculitis process and shows a distinct clinical and histological pattern [18]. BD patients with focal myositis experience significant pain, normal creatine kinase levels, vasculitis–related lesions with intact muscle structure, and frequent relapses [19].

1.4 Diagnosis of Neuro–Behcet Syndrome

The diagnosis and monitoring of NBS primarily depend on clinical assessment and MRI findings. The widely accepted diagnostic criteria for NBS [20], outlined in the 2014 International Consensus Recommendations (ICR) for Neuro–Behcet's Disease, emphasize the importance of integrating clinical symptoms with additional diagnostic tests, including MRI scans and cerebro spinal fluid (CSF) Interleukin (IL)–6 levels. Elevated levels of IL–6 in CSF is the most studied aspect of NBS, which are linked to disease activity, treatment results, and prognosis [21]. CSF IL-6 is a laboratory biomarker for NBS, however, it is not a specific indicator for NBS, therefore, CSF IL–6 is insufficient in terms of disease diagnosis itself. Although conventional MRI detects brain atrophy and WM abnormalities in neurological disorders like NBS, it can be mistaken in detecting for other neurological diseases which mimics NBS, especially MS. Accordingly, specific biomarker to aid in NBS diagnosis has yet to be identified [22],[23].

1.5 Neuroimaging in Neuro–Behcet Syndrome

Magnetic resonance imaging (MRI) of brain and spinal cord with the gadolinium contrast agent is the prevalent neuroimaging method to assess both forms of the disease. MRI protocol includes T1W imaging, 3D fluid–attenuated inversion recovery (FLAIR) with fat suppression, susceptibility–weighted imaging (SWI), diffusion–

weighted imaging (DWI), and contrast enhanced T1 series following gadolinium contrast injection. For vascular assessment, magnetic resonance venography (MRV) and magnetic resonance angiography (MRA) are performed using time–of–flight techniques and/or gadolinium–based contrast injection. Magnetic resonance spectroscopy (MRS) can help in differential diagnosis by highlighting a vascular pattern [24]. Additionally, clinical and laboratory findings such as positive HLA–B51 (human leukocyte antigen) in the blood test and possible positive CSF findings are supportive diagnostic tools. However, neurological diseases such as multiple sclerosis (MS), NMOSD, could mimic NBS due to pathological similarities. The lack of sufficient discrimination in MRI, combined with the disease pathology similarities between diseases in terms of lesion location and tissue properties, has led to confusion and misdiagnosis in several cases [25]. In conventional MRI, NBS mostly mimics MS in certain ways more than other neurological conditions. It is particularly prevalent among young people and may show a relapsing–remitting or progressive course. In addition, patients often show abnormal evoked potentials and white matter (WM) lesions with hyperintense areas on T2WI (T2–weighted images). Over the years researchers have been focused on finding a disease specific imaging marker such as; demyelination and immunocyte–infiltrated lesions. However, these findings are lacking and retards the clinical diagnosis [26].

1.6 Diffusion Tensor Imaging

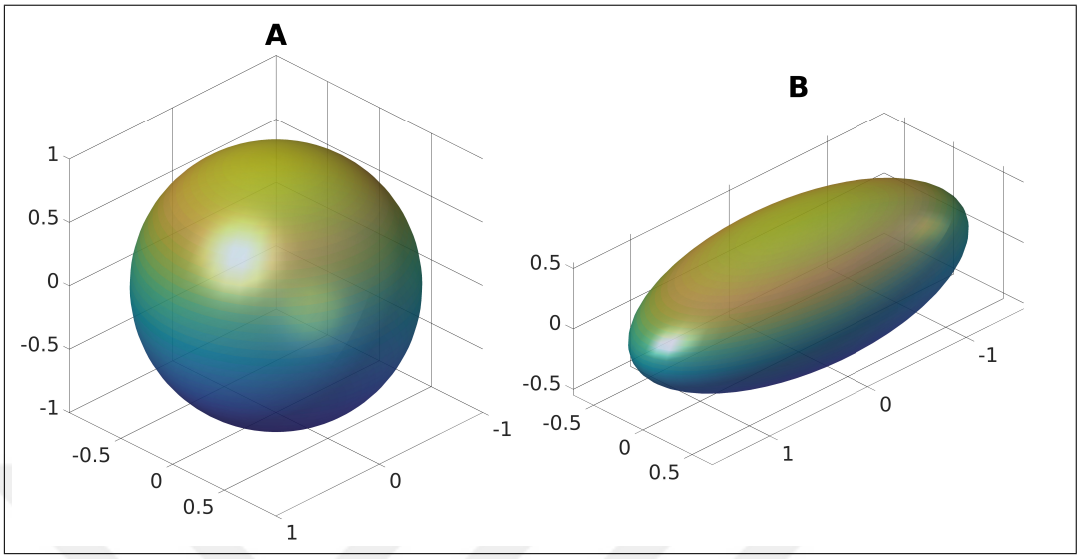


Figure 1.1 The depiction of water molecules diffusing equally in all directions in an isotropic environment (A), and more along one direction in an anisotropic environment (B).

Diffusion is a random transport process that explains the movement of material (e.g., water molecules) from one location to others over time. Diffusion tensor imaging (DTI) is a an advanced MRI technique for CNS imaging to evaluate WM changes based on diffusion characteristics of water molecules [27],[28],[29]. The principles of DTI are linked to diffusion weighted imaging (DWI). DWI is a non–enhanced and non–invasive MRI method based on the motion of water molecules and tissue water diffusion rate. DWI can be applied without any chemical tracers and contrast agents. The movement of water molecules reflect the diffusivity in the related area of the tissue and this enables to demonstrate biophysical properties and cellularity of the tissue. DTI is an extension of DWI and it not only measures the water diffusion but also detects the direction of the water diffusion along the tissues relying on the tissue characteristics such as tissue type and architecture [30],[31],[32],[33].

When water molecules move equally in all directions which means all eigenvalues are equal, the diffusion is referred to as isotropic.

$$\lambda_1 = \lambda_2 = \lambda_3 \tag{1.1}$$

where λ_1, λ_2, and λ_3 represents the three eigenvalues of diffusion tensor.

However, if the environment has a specific structural organization or contains barriers that restrict movement more in certain directions, the diffusion becomes anisotropic and in this case one eigenvalue dominates the others.

$$\lambda_1 \geq \lambda_2 \geq \lambda_3 \tag{1.2}$$

where λ_1, dominates other eigenvalues.

As shown in Figure 1.1. This variation in anisotropy is seen due to differences in the structural and compositional properties of various environments, such as CSF, WM, and GM. DTI analysis provides quantitative information which is not possible in conventional MRI. In WM, diffusion is less constrained along the axons and tends to be anisotropic whereas in CSF, diffusion occurs freely in all directions, making it isotropic [34],[35],[36].

The diffusion process is modeled as an ellipsoid, which can be mathematically described using a 3 x 3 symmetric matrix [37]. It is commonly referred to as the tensor as showed in Eq. 1.3.

$$\mathbf{D} = \begin{bmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{xy} & D_{yy} & D_{yz} \\ D_{xz} & D_{yz} & D_{zz} \end{bmatrix} \tag{1.3}$$

The diagonal elements ($D_{ii} > 0$) represent the diffusion variances along the x, y, and z axes, while the off–diagonal elements represent the covariance terms, which are symmetric about the diagonal ($D_{ij} = D_{ji}$). Upon diagonalizing the diffusion tensor. The resulting eigenvalues ($\lambda_1, \lambda_2, \lambda_3$) and corresponding eigenvectors ($\epsilon_1, \epsilon_2, \epsilon_3$) provide information about the directions and apparent diffusivity along the principal diffusion

axes [38]. As showed in Eq. 1.4:

$$\mathbf{E} = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} \tag{1.4}$$

The quantitative measures derived from diffusion tensor which models how water diffuses in tissue (especially in WM). While the three positive eigenvalues of the tensor (λ_1 (axial diffusivity, AD), the largest value eigenvalue, λ_2, λ_3) indicate the magnitude of water diffusion along the three principal axes of the diffusion tensor, apparent diffusion coefficient (ADC; or mean diffusivity, MD) showed in Eq. 1.5. that is the mean of the eigenvalues, radial diffusivity (RD) showed in Eq. 1.6. which measures the diffusion perpendicular to to axons and measures of anisotropy degree; fractional anisotropy (FA), Eq. 1.7 and relative anisotropy represented by Eq. 1.8.

$$ADC = \frac{\lambda_1 + \lambda_2 + \lambda_3}{3} \tag{1.5}$$

$$RD = \frac{\lambda_2 + \lambda_3}{2} \tag{1.6}$$

$$FA = \sqrt{\frac{(\lambda_1 - \lambda_2)^2 + (\lambda_2 - \lambda_3)^2 + (\lambda_1 - \lambda_3)^2}{2(\lambda_1^2 + \lambda_2^2 + \lambda_3^2)}} \tag{1.7}$$

$$RA = \sqrt{\frac{3\big((\lambda_1 - \lambda_2)^2 + (\lambda_2 - \lambda_3)^2 + (\lambda_1 - \lambda_3)^2\big)}{(\lambda_1 + \lambda_2 + \lambda_3)}} \tag{1.8}$$

DTI is a valuable imaging tool for examining WM abnormalities and structure in vivo due to its sensitivity to microstructural tissue properties. It provides diffusion measures and tissue fiber orientation, making it widely used in clinical settings to investigate conditions with expected WM abnormalities, such as stroke, brain tumors, neurodegenerative disorders like MS and Alzheimer's disease, neuropsychiatric disorders like schizophrenia and developmental disorders like attention deficit hyperactivity Disorder (ADHD) and autism, movement disorders such as Parkinson's disease, and genetic disorders. DTI is also used to track WM changes during aging and neurodevelopment. Its variables, help detect structural alterations related to myelination and axonal injury. With increasing applications and consistent results, DTI is now to become an integral part of disease treatment planning, preclinical marker detection, and routine clinical imaging [39],[40],[41],[42],[43],[44]. To assess the microstructural integrity of brain white matter tracts in NBS patients and BD patients who do not exhibit neurological symptoms, DTI has the potential to be an important non–invasive quantitative tool in clinical practice.

1.7 Machine Learning in Neuro–Behcet Syndrome

The role of machine learning (ML) and artificial intelligence (AI) in medicine has become a major area of focus, especially for their potential to enhance the diagnostic and predictive analysis of medical images in radiology [45]. A wide range of brain–related diseases, including MS, brain tumors, traumatic brain injuries, epilepsy, stroke and rare diseases such as NBS, are commonly diagnosed and monitored using MRI in conjunction with clinical symptoms. Medical imaging technologies offer various opportunities for disease prediction and prognosis. Imaging studies are essential in the differential diagnosis and evaluation of rare diseases, often helping to avoid the need for invasive procedures such as biopsy. Imaging biomarkers, in particular, allow the quantification of phenotypes across a spectrum of neurological conditions, thus enhancing the precision of predictive models. ML techniques can be used to extract and analyze these biomarkers, facilitating the development of predictive models for clinical applications. An increasing number of studies are examining imaging biomarkers for

their prognostic and predictive capabilities, as well as their consistency across various clinical environments [46],[47],[48],[49],[50].

MRI is a crucial tool for most of the neurological disorders in terms of diagnosis and monitoring. Traditional biomarkers include lesion count, volume, and location, while advanced biomarkers encompass diffusely abnormal white matter (DAWM), normal–appearing white matter (NAWM), and cortical lesions. These advanced biomarkers, often derived from DWI or DTI, provide insights into the disease's diffuse effects beyond focal lesions. The potential of integrating multimodal data and advanced ML techniques to develop personalized prognostic models for the patients suffers from neurological conditions is promising [51].

1.8 Literature Review

Kocer et al. (1999) conducted some of the earliest studies on CNS involvement in BD using conventional MRI, revealing common lesion sites such as the brainstem and basal ganglia. These findings set the stage for more advanced imaging methods to characterize WM pathology in detail in the future [52]. In 2005 Anik et al. further explored microstructural brain changes by measuring magnetization transfer ratios, which demonstrated tissue integrity loss even in regions appearing normal on conventional MRI. This implied that NBS could involve widespread subtle brain changes beyond focal lesions [53]. In 2003, Kunimatsu et al. analyzed ADC values in patients with NBS using DWI. They reported increased ADC values in both supratentorial and infratentorial white matter regions, indicating microstructural integrity loss. These findings were associated with cognitive impairment in NBS patients [54]. It is important to note that during the early 2000s, DTI was not as commonly available or utilized as DWI, which may explain the use of DWI in this study. Additionally, Baysal et al. (2005) reported a significant increase in ADC values in BD patients with neurological involvement compared to those without, based on DWI analysis [55]. Yildiz et al. (2013) demonstrated the utility of DWI in distinguishing between vasogenic and cytotoxic edema in brainstem lesions associated with NBS. Based on a case study of a

28–year–old male diagnosed with BD who exhibited neurological abnormalities, both MRI and DWI revealed notable findings, including increased ADC values. These observations provided insights into the underlying pathophysiology and contributed to more informed treatment decisions [56]. The potential impact of DTI in NBS began to be systematically investigated in the mid–2010s [57]. The earliest known application of DTI in the study of NBS appears to be the 2019 study by Alis et al. This retrospective study evaluated the ADC characteristics of parenchymal NBS lesions, highlighting that both acute and chronic lesions had significantly higher ADC values compared to contralateral normal brain tissue. This suggested that ADC measurements might provide substantial information about the pathology of parenchymal NBS lesions [58]. In 2021 Atasoy et al. mentioned that even in BD patients without neurological involvement, DTI revealed altered WM integrity, particularly in regions such as the corpus callosum, internal capsule, and frontal lobes. FA was significantly decreased in several WM regions of BD patients whereas RD and ADC ere elevated compared to healthy controls [59]. Aslan et al. investigated microstructural WM changes in patients with BD with and without neurological involvement.

In this thesis, a decrease in FA and an increase in ADC and RD, in larger regions of major WM tracts, while AD values were observed increased in fewer tracts of the patients with NBS [60]. In 2024, Gunduz et al. also expressed that high ADC value is an indicate for non–ischemic tissue damage whereas AD and RD presents axonal degeneration and demyelination, respectively [61]. In one of the most recent studies conducted in 2025 by Uygunoglu et al. has mentioned that in NBS patients, DTI revealed decreased FA and increased RD, ADC, and AD in both supratentorial and infratentorial WM regions, indicating widespread microstructural damage. Similar, though less pronounced, alterations were observed in BD patients without neurological symptoms. In NBS patients, alterations in specific WM tracts, including the cingulum bundle, were associated with poorer performance across multiple cognitive domains and higher disease severity [62].

1.9 Research Gap

Previous studies on NBS have primarily focused on basic statistical comparisons of DTI variables or have been limited to case reports [52],[53],[54],[55],[56],[57],[58],[59],[60],[61],[62]. These investigations have explored ADC, FA, and other diffusion metrics to characterize WM abnormalities in NBS, often demonstrating alterations associated with microstructural damage and cognitive impairment. Although the field has seen substantial progress in characterizing NBS with DTI, most of these studies have compared NBS patients either with healthy individuals or with BD patients without neurological involvement.

Only a limited number have directly analyzed WM abnormalities in NBS patients versus healthy controls, and few have used advanced image analysis or classification methods. For instance, a recent study by Zhan et al.(2025) utilized tractography to compare DTI metrics between 12 p–NBS patients and 12 healthy individuals [22]. However, tractography–based analysis can introduce bias due to dependence on diffusion anisotropy metrics (FA, RA) for tract selection, potentially creating a methodological loop. Despite advancements in imaging technologies, there are currently no published studies that integrate ML techniques with DTI data specifically for NBS. While several ML applications have been successfully developed for other BD manifestations, such as ocular and gastrointestinal involvement [63],[64],[65] and their potential in neurological forms of the disease remains unexplored.

This lack of ML–based approaches in NBS is especially significant given proven capabilities of ML in uncovering subtle patterns in neuroimaging data, improving diagnostic accuracy, and supporting prognosis in other neuroinflammatory and neurodegenerative disorders.

This study addresses this critical gap by applying supervised ML algorithms to classify DTI metrics between NBS patients and healthy individuals. Given the global rarity of NBS and the difficulty in acquiring large datasets, this study presents a valuable cohort from Türkiye, comprising 14 NBS patients, including two with suspected

non–parenchymal involvement. The integration of ML with DTI represents a novel contribution to the field. This approach has the potential to enable earlier and more accurate identification of CNS involvement in BD, enhance differential diagnosis between NBS and other neurologic disorders with overlapping features, and support the development of tools for risk stratification and individualized monitoring of neurological progression in BD.

2. MATERIALS AND METHODS

2.1 Medical Data

Table 2.1
Patient Demographics.

Demographics	**Value**
Gender	6 Female / 9 Male
Average Age ± SD	40,4 ± 13,28

Table 2.2
Types of NBS Observed in the Cohort.

Type of NBS	**Value**
Parenchymal	13
Parenchymal + Suspicious Extra–axial	1
Parenchymal + Extra–axial	1

In this Institutional Review Board (IRB) approved study, the data gathered from 14 NBS patients (9 males and 5 females, average age 40.40 ± 13.28 years) and 38 healthy volunteers (16 males and 22 females, average age 45.05 ± 13.70 years), with written consent. All patients exhibited the parenchymal subtype, the most prevalent form. One patient additionally showed a suspicious extra–axial (non–parenchymal NBS) form, and another presented with both parenchymal and extra–axial involvement. All patient data included in this study were obtained during the acute phase of the disease. However, one NBS patient was excluded due to the existence of a large lesion in the frontal lobe. Cohort characteristics including patient demographics and the amount of patient of specific NBS type were demonstrated respectively, in Table 2.1 and Table 2.2.

2.2 Imaging Data

Retrospective MRI data of 14 NBS patients and 38 healthy individuals with given written consents were collected from medical imaging database, PACS (Picture Archival and Communication Systems) of Acibadem Altunizade Hospital. The anatomical T1WI, T2WI, FLAIR and nonprocessed DTI data were collected with Siemens 3T Magnetom Tim Trio MR imaging scanner using a 32 channel head coil. The scanning protocol involved acquiring T2WI using a 2D turbo spin–echo sequence with a pixel size of 0.245 $\times$ 0.245 mm across 35 axial slices. The imaging parameters were TE/TR = 101/6750 ms, slice thickness = 3 mm, and a flip angle of 130°. DTI data were obtained using a 2D diffusion echo planar imaging (EPI) sequence with 2.05 $\times$ 2.05 mm pixels, 52 axial slices, TE/TR = 69/5500 ms, slice thickness = 2 mm, flip angle = 90°, and a b–value of 1000 ms/mm^2 with 20 diffusion gradient directions. The console of the scanner computed the three eigenvalue maps ($\lambda_1 \geq \lambda_2 \geq \lambda_3$). All the data were obtained in DICOM (Digital Imaging and Communications in Medicine) format, and anonymized. Using MIPAV (Medical Image Processing, Analysis and Visualization) software, anonymized DICOM data were converted to Neuroimaging Informatics Technology Initiative (NIfTI) to be used in this study.

2.3 White Matter Segmentation

White mater segmentation process was conducted on a cohort of 14 NBS and 38 control axial T2WIs. The segmentation process was using the MIPAV®'s semi–automatic segmentation tool selecting WM regions.

Primarily, the segmentation process started by defining the Volume of Interest (VOI) within the anatomical brain images, using 'Draw Levelset VOI' feature available in MIPAV®. Within the context of this study, the VOI encompassed the WM regions of the brain. Selected regions, i.e., the VOIs, were manually demarcated for each slice across all NBS patients and healthy individuals, which were then converted into binary

mask images. WM segmentation in MR images was conducted in a systematic manner under the supervision of Dr. Abdullah Arcan, Department of Neurology, Cerrahpaşa Faculty of Medicine. For both patient and control groups, segmentation commenced at the axial T2WI from the slice where the lateral ventricles became visible. This approach effectively minimized the inclusion of non–white matter regions, thereby reducing the potential for errors in subsequent diffusion calculations and optimizing the overall processing time. Owing to the pathological and anatomical complexities inherent in NBS, some brain images demonstrated scattered or indistinct lesions, necessitating adjustments in contrast levels and iterative segmentation procedures. During the segmentation process T2WI offered enhanced contrast among various tissue types, such as gray matter, white matter, and CSF.

As the objective of this study was to identify the most discriminative DTI metrics for distinguishing pathological WM in NBS, evaluate the classification accuracy of ML algorithms in detecting WM abnormalities, and establish a bias free, whole brain framework for WM assessment without relying on selective tractography. WM masks were extracted from T2W images and accurately coregistered onto the non diffusion weighted (B_0) images. Subsequently, DTI metrics were analyzed across the entire WM volume, thus eliminating regional selection bias and avoiding circular dependencies introduced by tractography based methods as demonstrated in Figure 2.1.

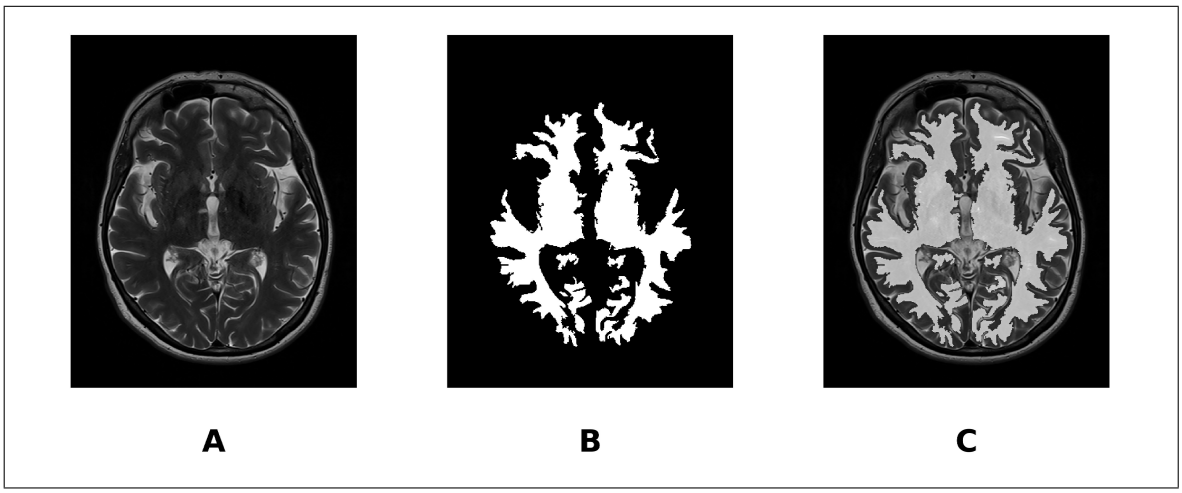


Figure 2.1 T2W image and WM mask derived from T2W: (A) T2W image; (B) WM mask obtained from T2W; (C) WM mask overlaid on the T2W anatomical image.

2.4 Coregistration

The coregistration procedure was performed using Advanced Normalization Tools Software (ANTs) [66] by an undergraduate electrical engineering student from Bogazici University, Hamit Alp Cömert with one year of experience in medical imaging. The ANTs parameter values optimized by Alpay Özcan for earlier neuroimaging studies were used. The aim was adjusting the intensity values of the images into a common scale by coregistering B_0 images onto T2WI. This helped to enhance contrast and reduce variability between different scans, making it easier to analyze anatomical structures and detect abnormalities. First, T2WIs underwent N4 bias field correction, and all T2WIs along with B_0 images were subjected to skull stripping. N4 bias field correction was performed to compensate the biases caused from MRI scanner such as artifacts and some tissue properties. The B_0 images served as a structural reference to identify and align anatomical features, including WM regions, before applying diffusion specific analysis. Using affine transforms and mutual information cost function, T2WIs were coregistered onto B_0 images. The resulting transformations were applied to the WM masks of T2WIs, using linear interpolation.

To binarize the outcomes of the linear interpolation in the range of [0,1], a two step thresholding process was created. Initially, the results were subjected to a simple thresholding process with a threshold value of 0.6. Subsequently, since the WM should appear darker in DWIs, a threshold based on the B_0 intensity was set. For this, masks selecting the lateral ventricles on the slice where they are most prominent. The threshold was set according to the following Eq. 2.1.

$$t = \mathrm{Mean}_{\mathrm{vent}} - \left(2 \times \mathrm{SD}_{\mathrm{brain}}\right) \tag{2.1}$$

where t is the threshold, $\mathrm{Mean}_{\{\mathrm{vent}\}}$ is the mean value of the intensities of the masked lateral ventricles area, $\mathrm{SD}_{\{\mathrm{brain}\}}$ refers to the standard deviation of the intensities of the non–zero voxels within the B_0 volumes as shown in Figure 2.2.

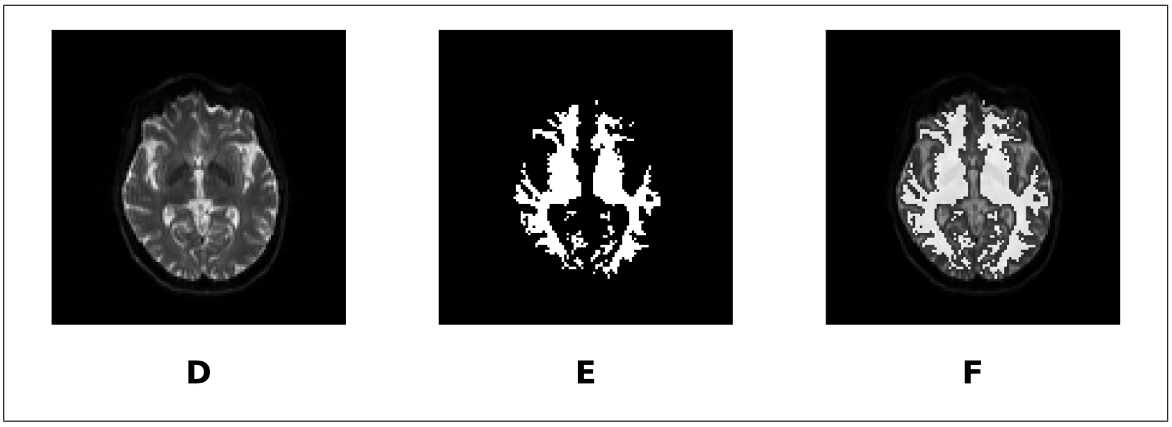


Figure 2.2 B_0 and coregistered masks: B_0 image (D); T2 mask overlaid on B_0 (E); coregistered T2 mask on T2WI (F).

2.5 DTI Variables

DTI data were processed using MATLAB® to extract voxel based diffusion metrics within segmented WM masks for both NBS ($n = 14$) and control ($n = 38$) groups. After loading the WM masks and eigenvalue images (λ_1, λ_2, λ_3), derived diffusion metrics including ADC, RD, FA, and RA were computed. For each subject, the mean and standard deviation values of λ_1, λ_2, λ_3, ADC, RD, FA, and RA within the WM masks were calculated, and subjects were labeled according to their respective groups (NBS = 1, Control = 0). These statistical summaries were subsequently organized into structured tables, facilitating further classification analyses and comparisons between the NBS and control cohorts.The mean, Eq. 2.2 and standard deviation, Eq. 2.4 were calculated as in the following equations.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{2.2}$$

$$V = \frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2 \tag{2.3}$$

$$\mathrm{SD} = \sqrt{V} \tag{2.4}$$

where $\bar{x}$ is the sample average of variable x and n is the number of values in the sample [67].

2.6 Statistical Analysis

The statistical analysis was performed using MATLAB®. The statistical analysis was performed by comparing the mean diffusion values. Diffusion differences between NBS and healthy WM compared with Mann–Whitney U test with a p–value of 0.05.

2.7 Classification

The 1000 bin fixed histograms of the representative non–healthy and healthy subjects provided a high–resolution approximation of the underlying distributions, enabling a more precise assessment of their shape and skewness. The visual progression from low to high bin counts underscored how finer binning captured more nuanced distributional features, which may enhance the discriminative power of the resulting feature vectors when used as inputs to ML models. For the voxel–wise analysis higher–resolution histograms (e.g., 30, 60, 90, 120 bins) were plotted to extract meaningful distributional features for ML classification. In the figures below, there is a depiction of the histograms for each DTI variable distribution in NBS WM and healthy WM in each bin sizes to be used in ML algorithms.

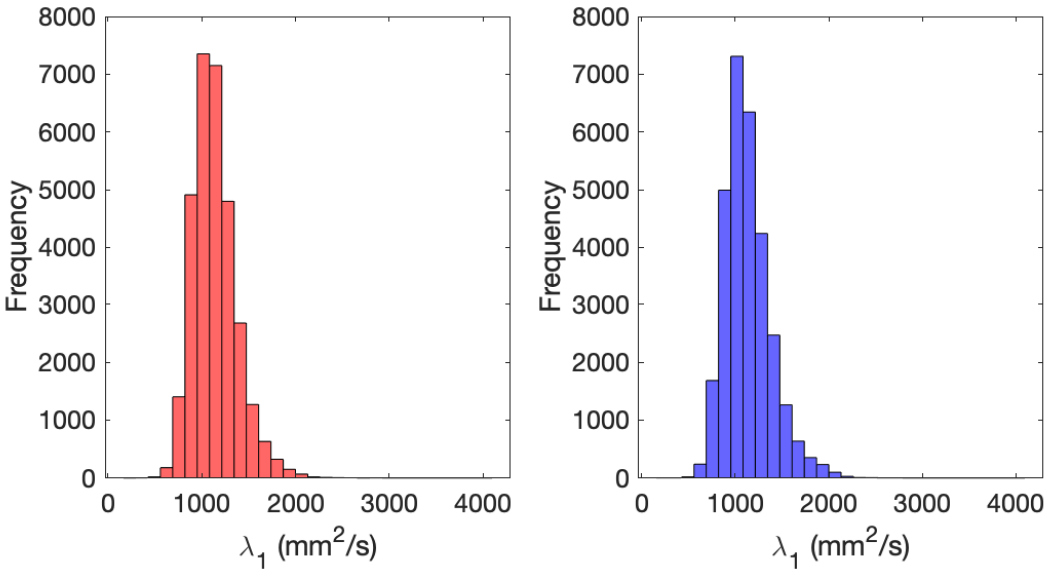


Figure 2.3 30–bin histogram distribution of λ_1 in NBS and Healthy WM.

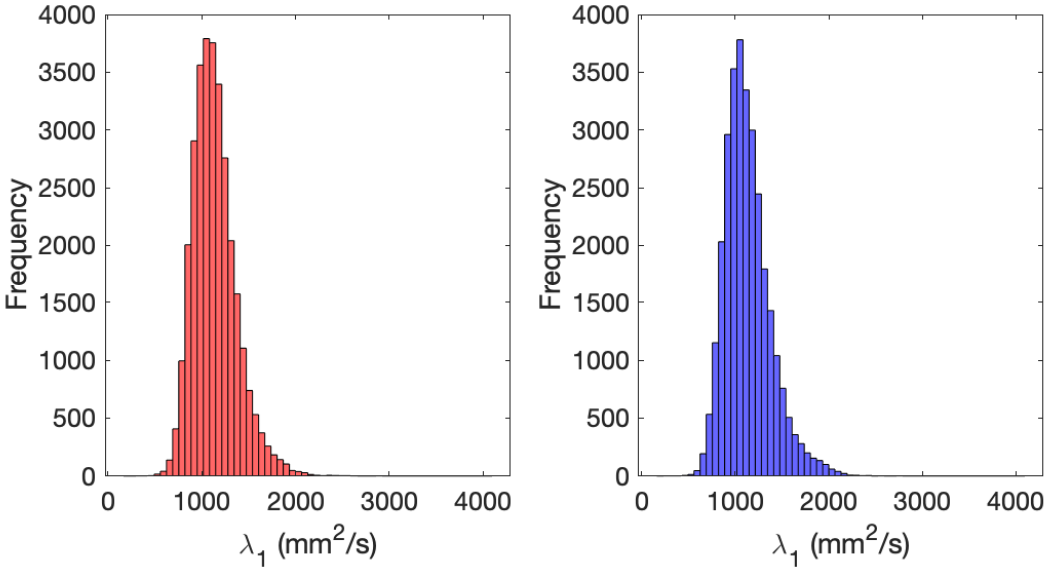


Figure 2.4 60–bin histogram distribution of λ_1 in NBS and Healthy WM.

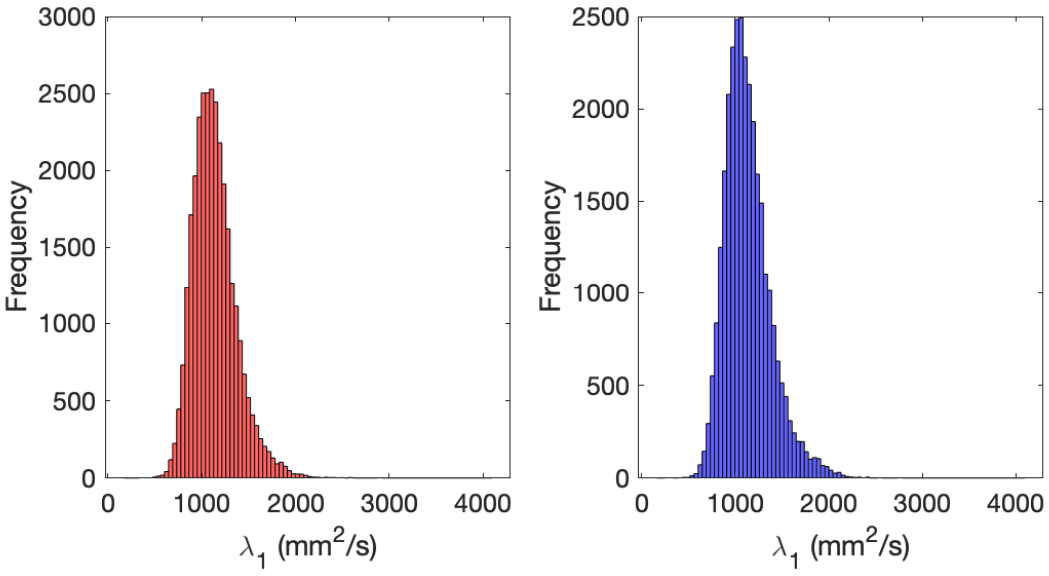


Figure 2.5 90–bin histogram distribution of λ_1 in NBS and Healthy WM

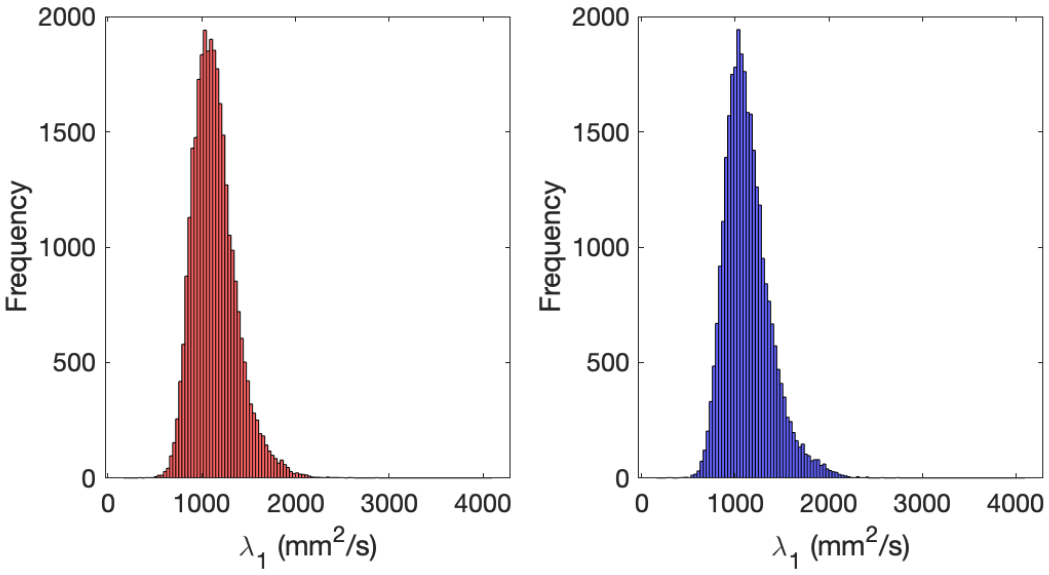


Figure 2.6 120–bin histogram distribution of λ_1 in NBS and Healthy WM.

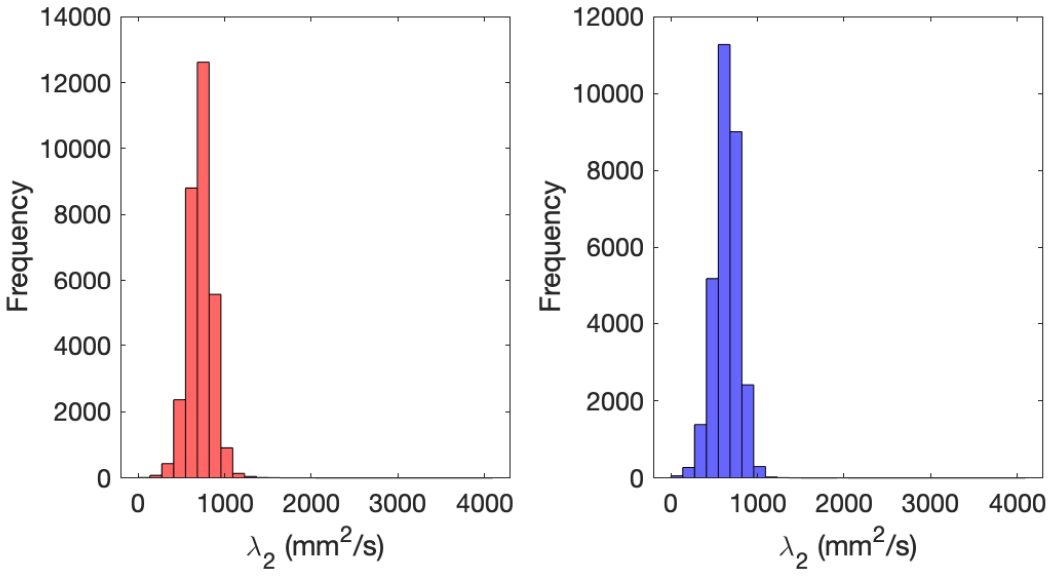


Figure 2.7 30–bin histogram distribution of λ_2 in NBS and Healthy WM.

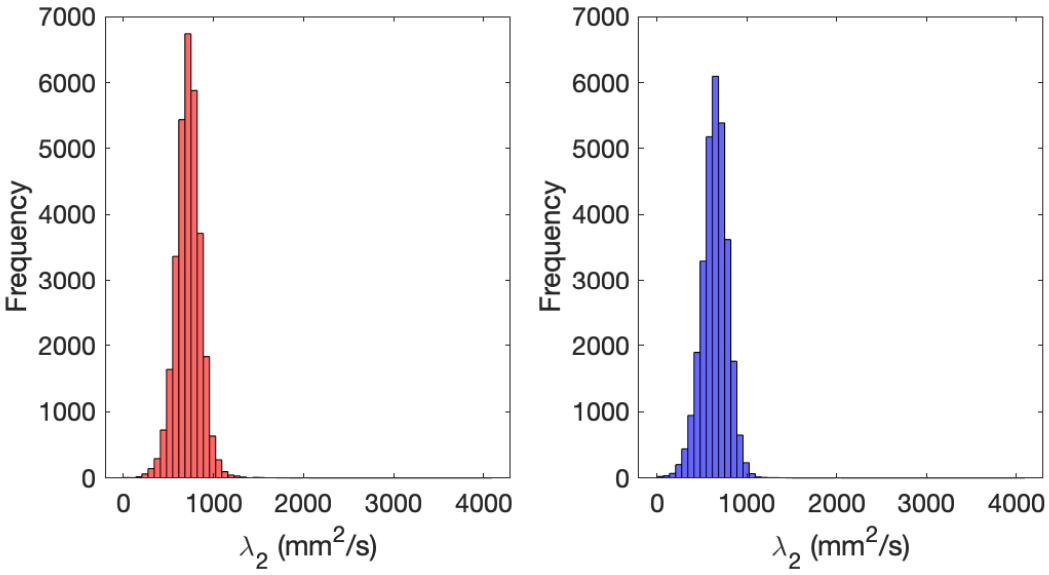


Figure 2.8 60–bin histogram distribution of λ_2 in NBS and Healthy WM.

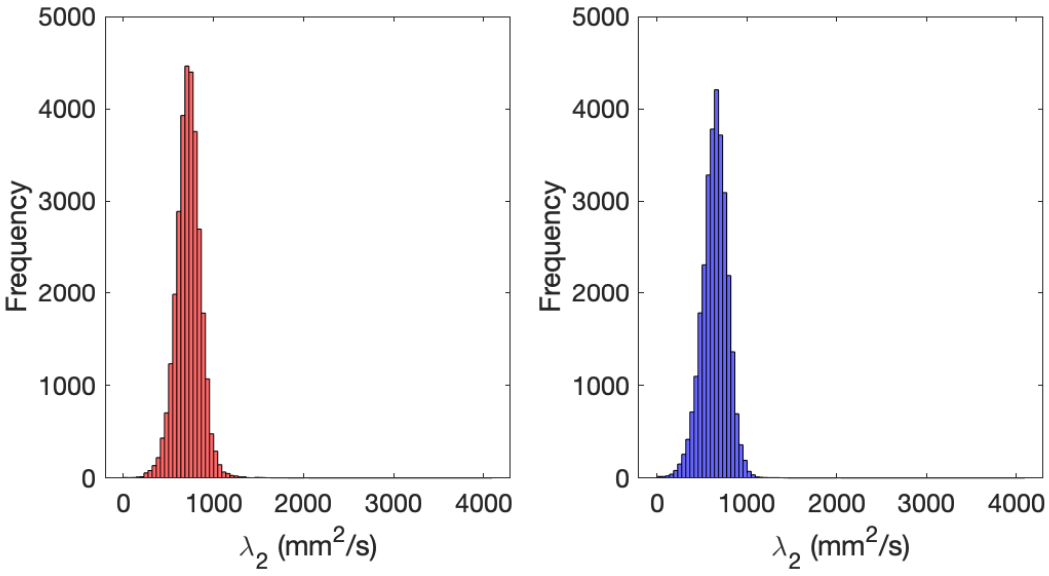


Figure 2.9 90–bin histogram distribution of λ_2 in NBS and Healthy WM.

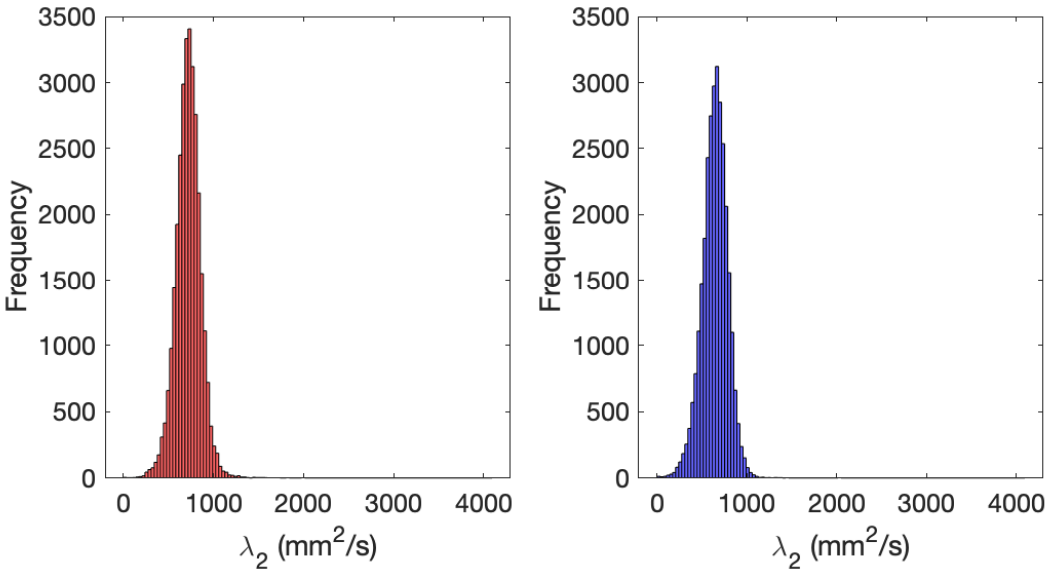


Figure 2.10 120–bin histogram distribution of λ_2 in NBS and Healthy WM.

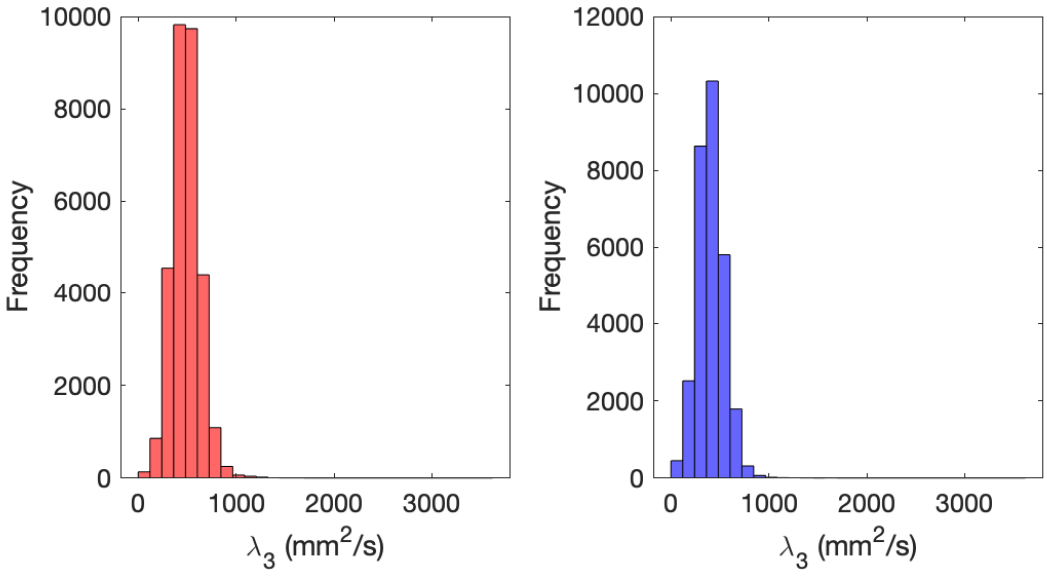


Figure 2.11 30–bin histogram distribution of λ_3 in NBS and Healthy WM.

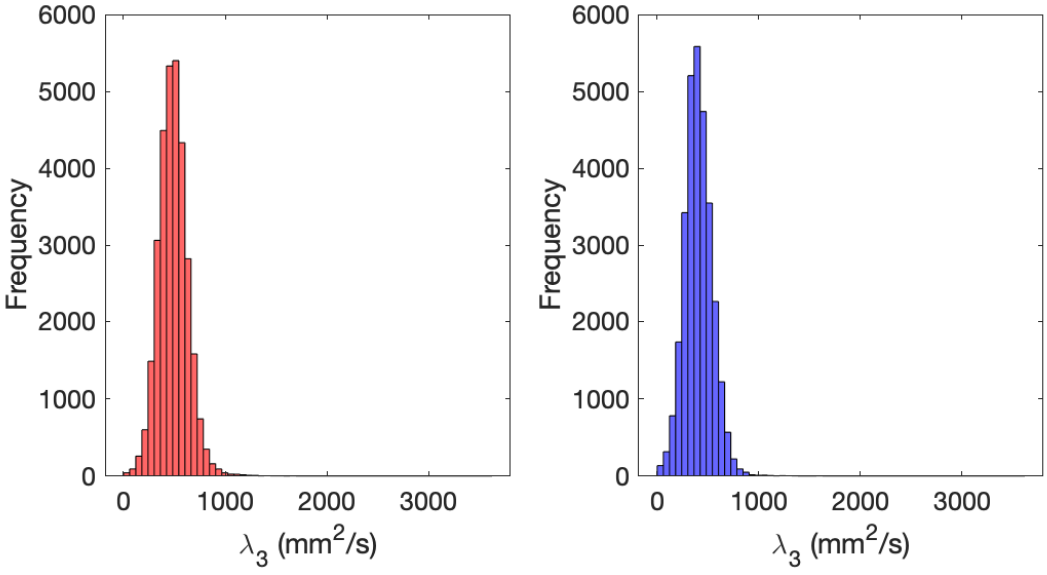


Figure 2.12 60–bin histogram distribution of λ_3 in NBS and Healthy WM.

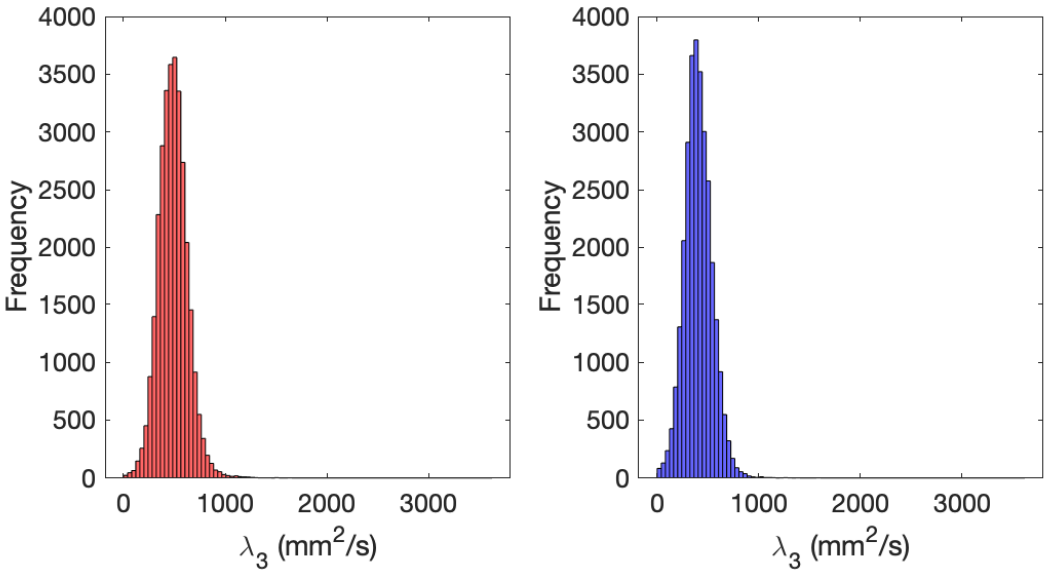


Figure 2.13 90–bin histogram distribution of λ_3 in NBS and Healthy WM.

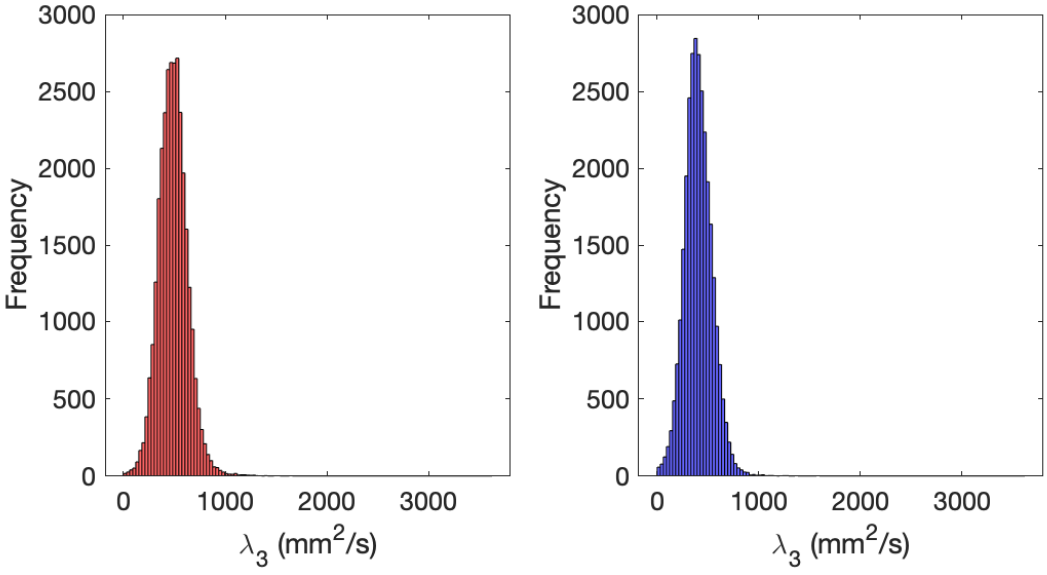


Figure 2.14 120–bin histogram distribution of λ_3 in NBS and Healthy WM.

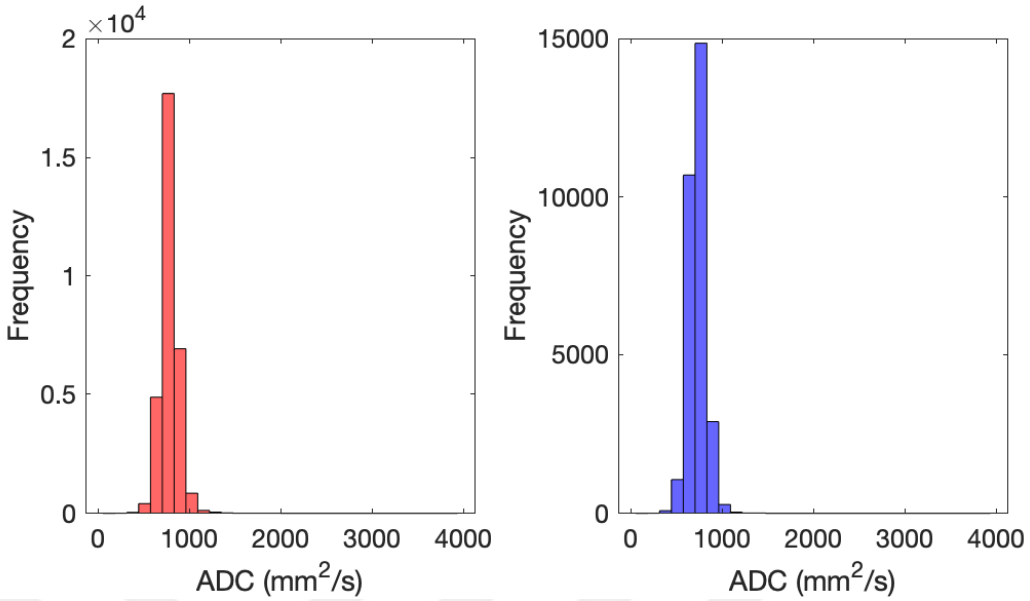


Figure 2.15 30–bin histogram distribution of ADC in NBS and Healthy WM.

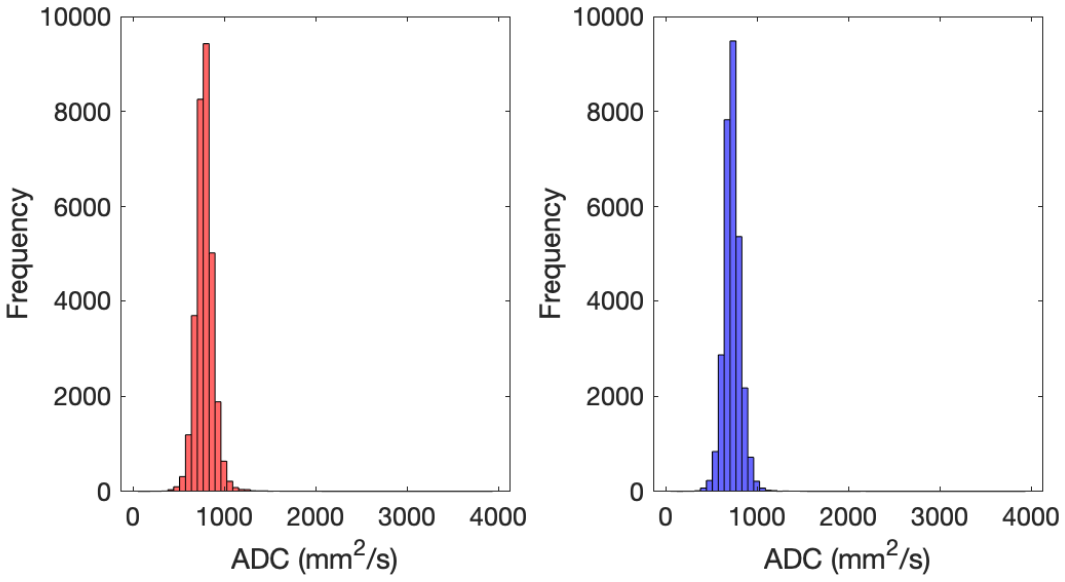


Figure 2.16 60–bin histogram distribution of ADC in NBS and Healthy WM.

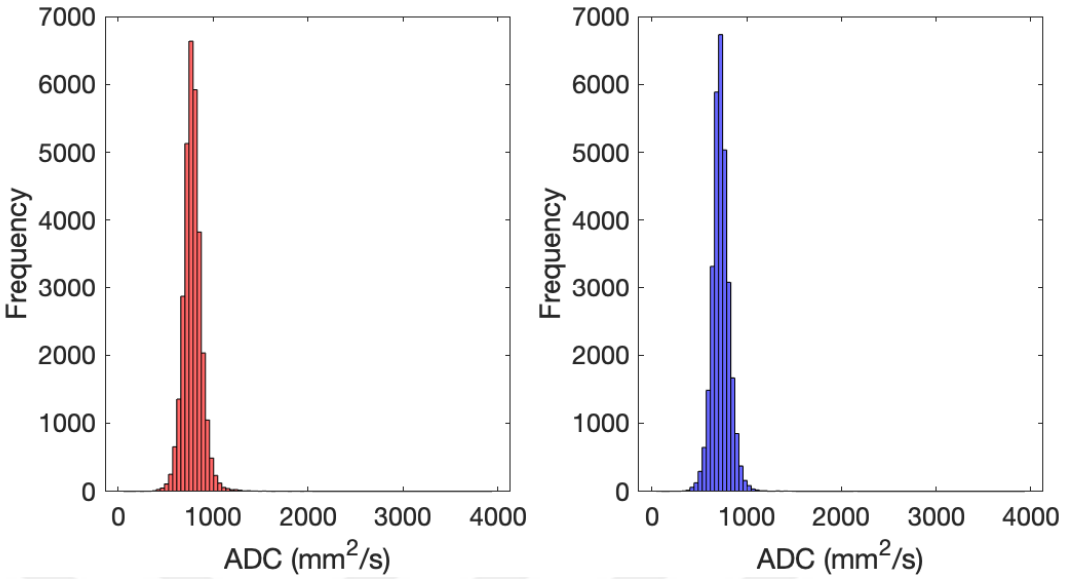


Figure 2.17 90–bin histogram distribution of ADC in NBS and Healthy WM.

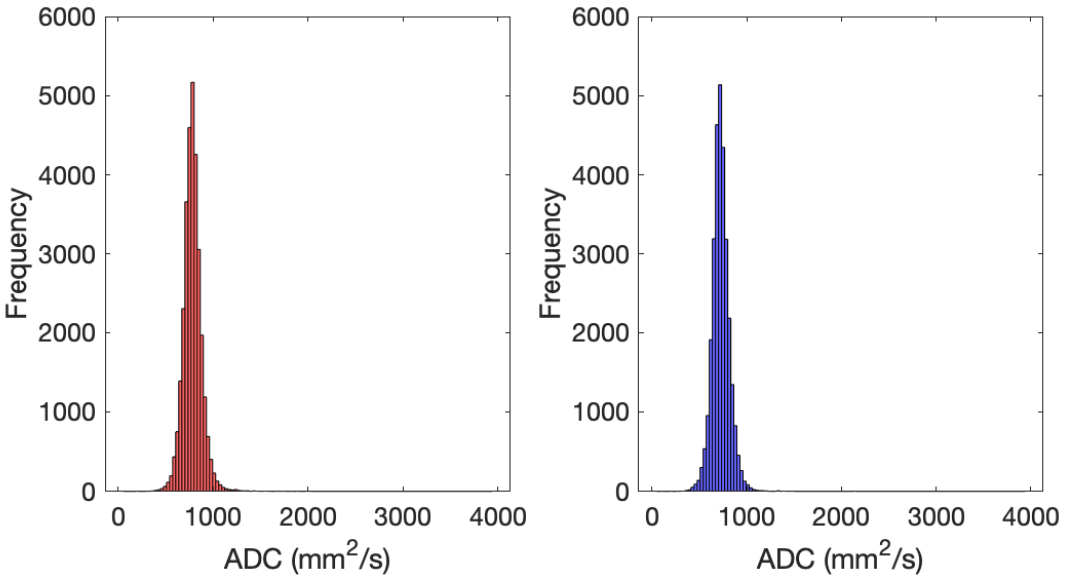


Figure 2.18 120–bin histogram distribution of ADC in NBS and Healthy WM.

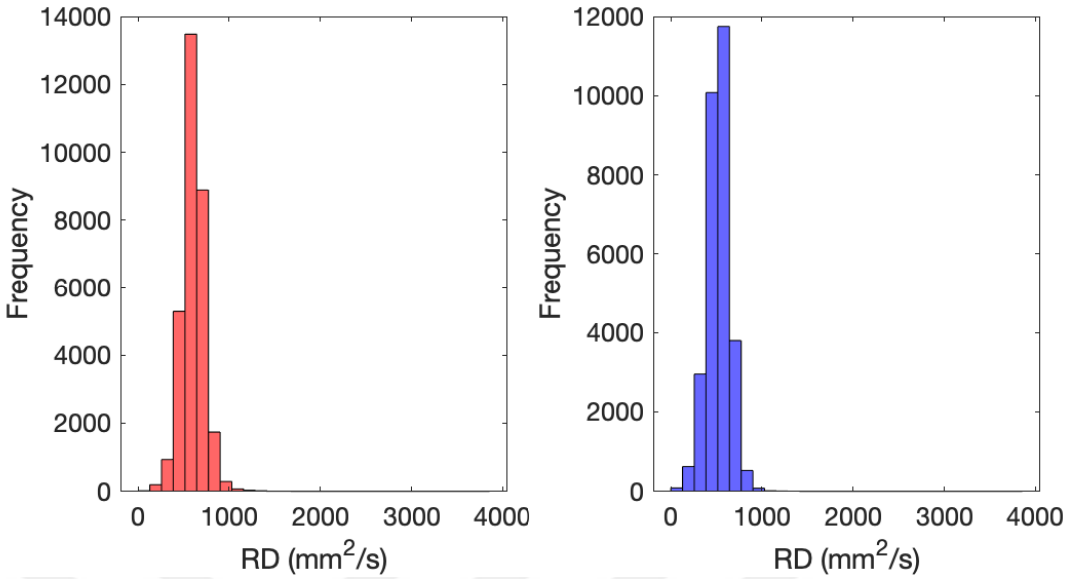


Figure 2.19 30–bin histogram distribution of RD in NBS and Healthy WM.

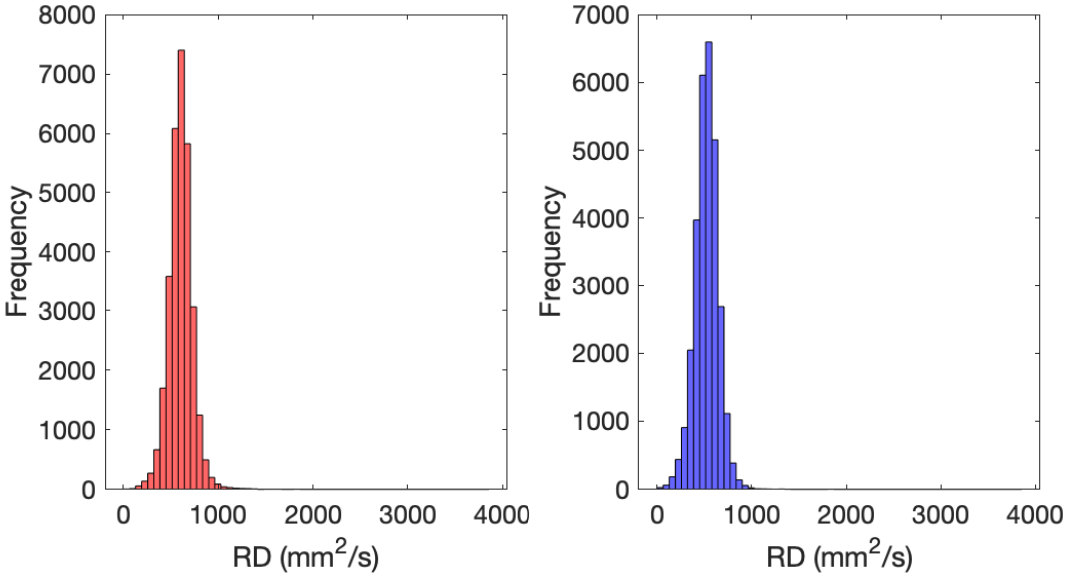


Figure 2.20 60–bin histogram distribution of RD in NBS and Healthy WM.

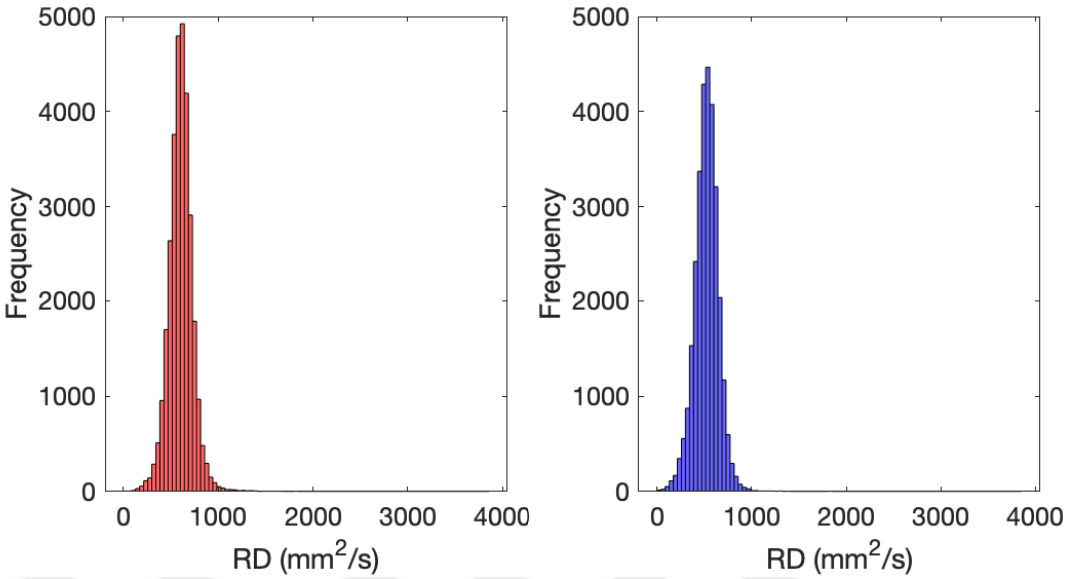


Figure 2.21 90–bin histogram distribution of RD in NBS and Healthy WM.

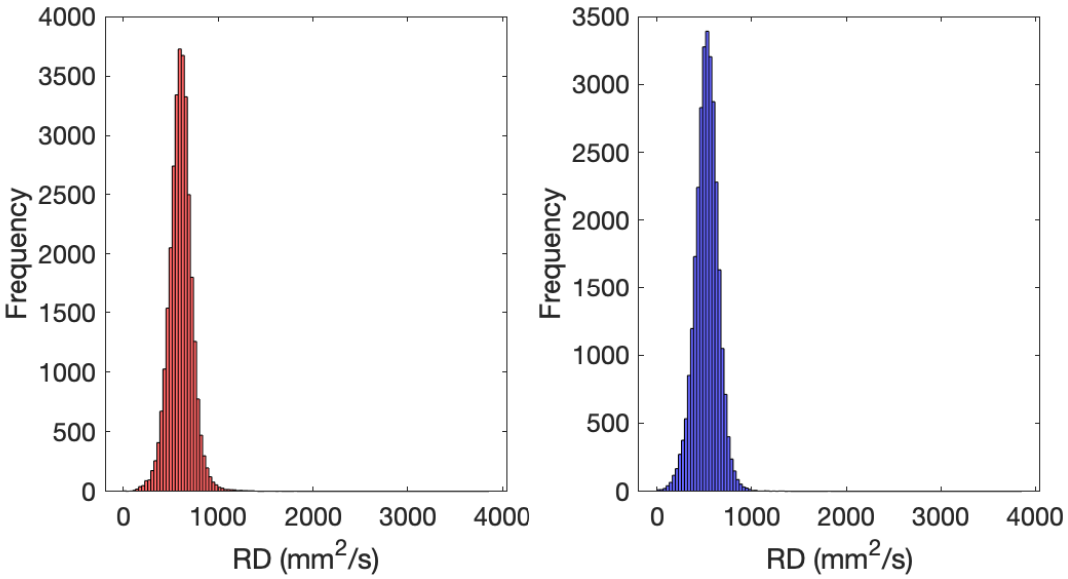


Figure 2.22 120–bin histogram distribution of RD in NBS and Healthy WM.

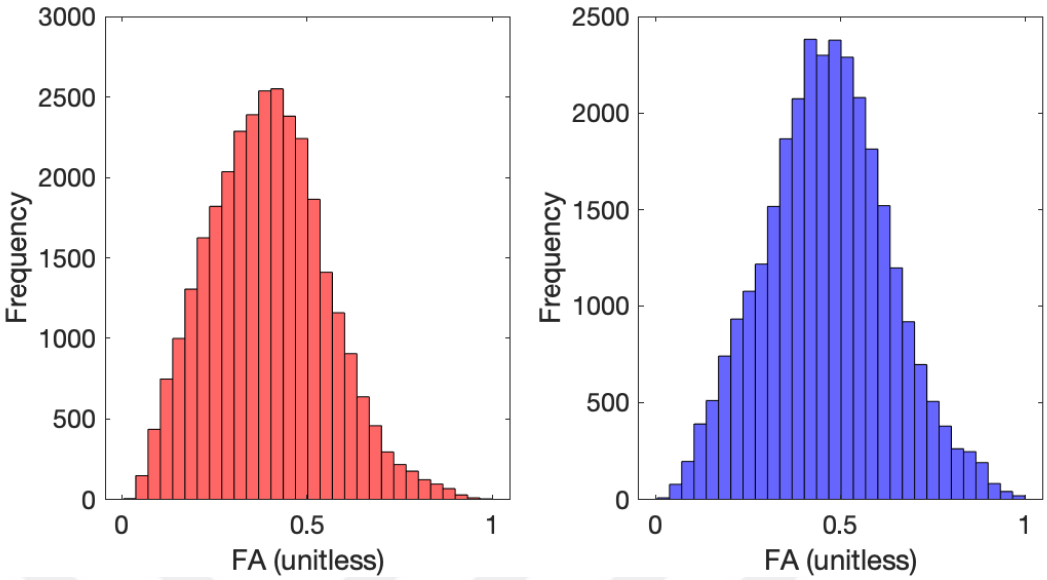


Figure 2.23 30–bin histogram distribution of FA in NBS and Healthy WM.

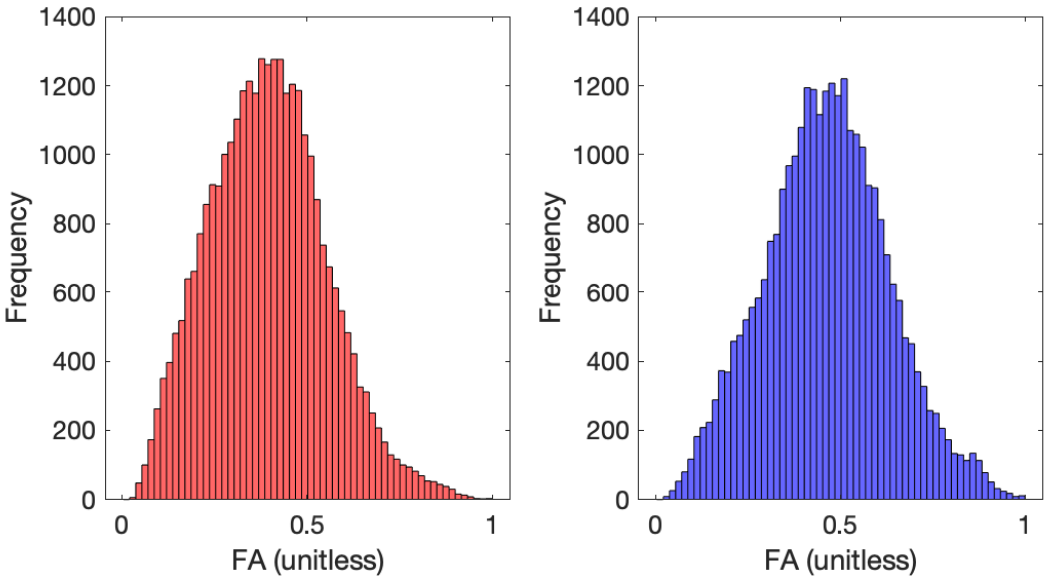


Figure 2.24 60–bin histogram distribution of FA in NBS and Healthy WM.

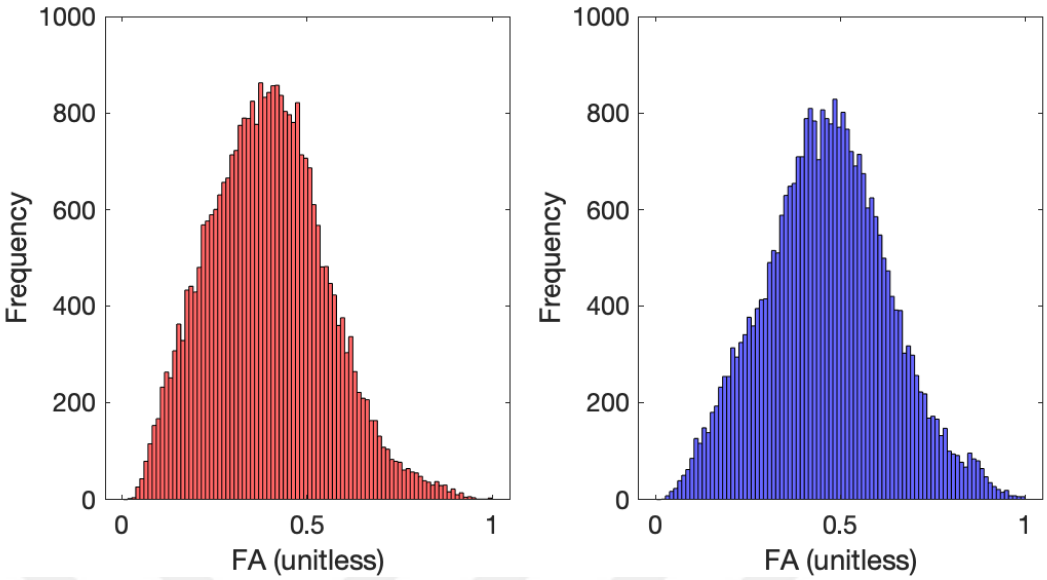


Figure 2.25 90–bin histogram distribution of FA in NBS and Healthy WM.

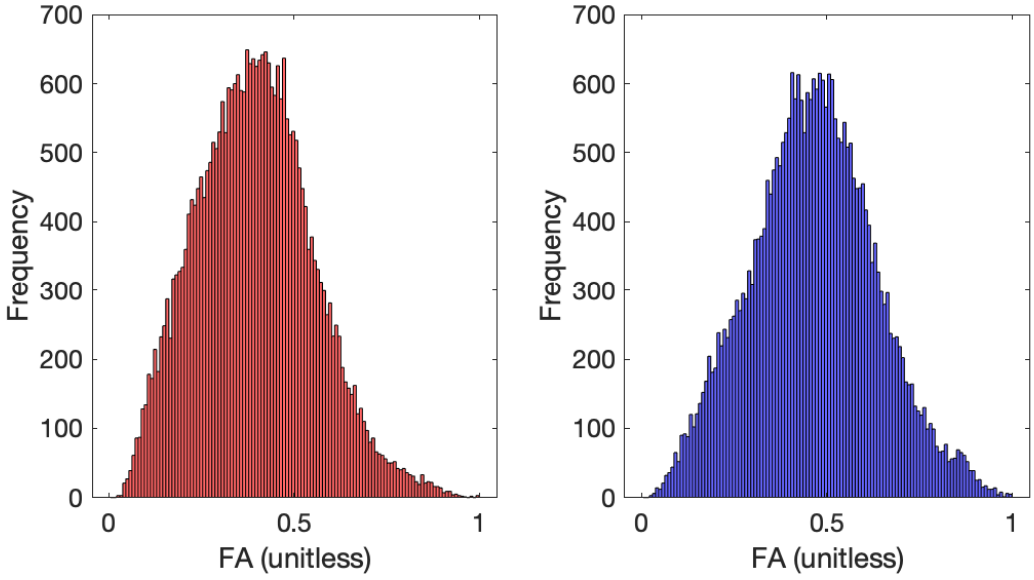


Figure 2.26 120–bin histogram distribution of FA in NBS and Healthy WM.

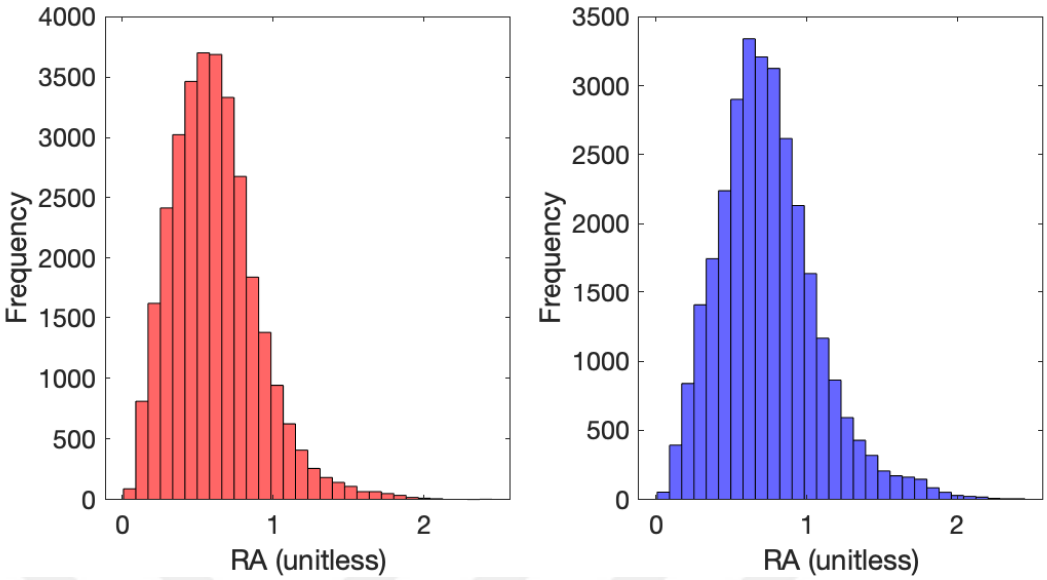


Figure 2.27 30–bin histogram distribution of RA in NBS and Healthy WM.

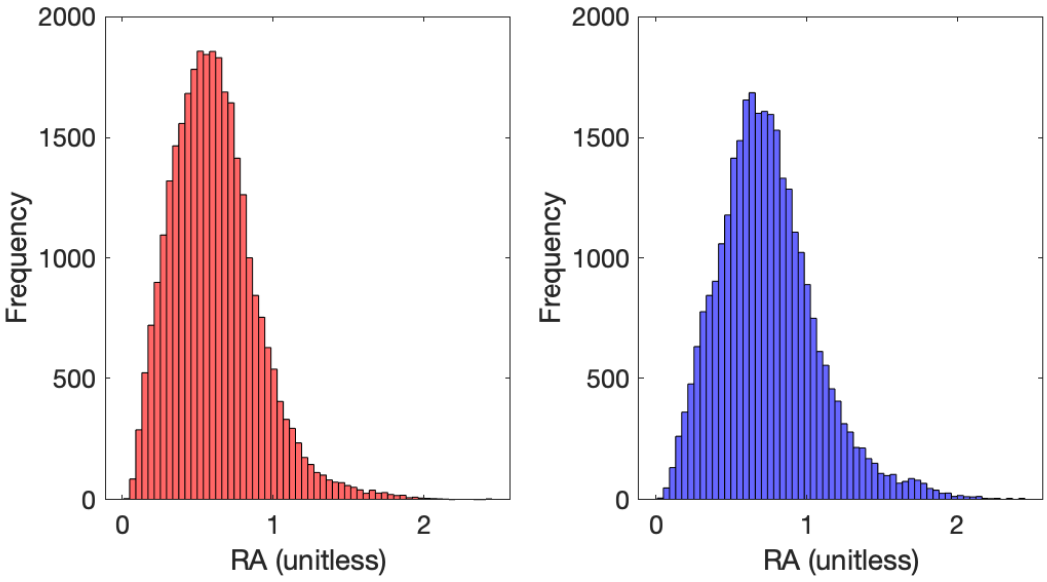


Figure 2.28 60–bin histogram distribution of RA in NBS and Healthy WM.

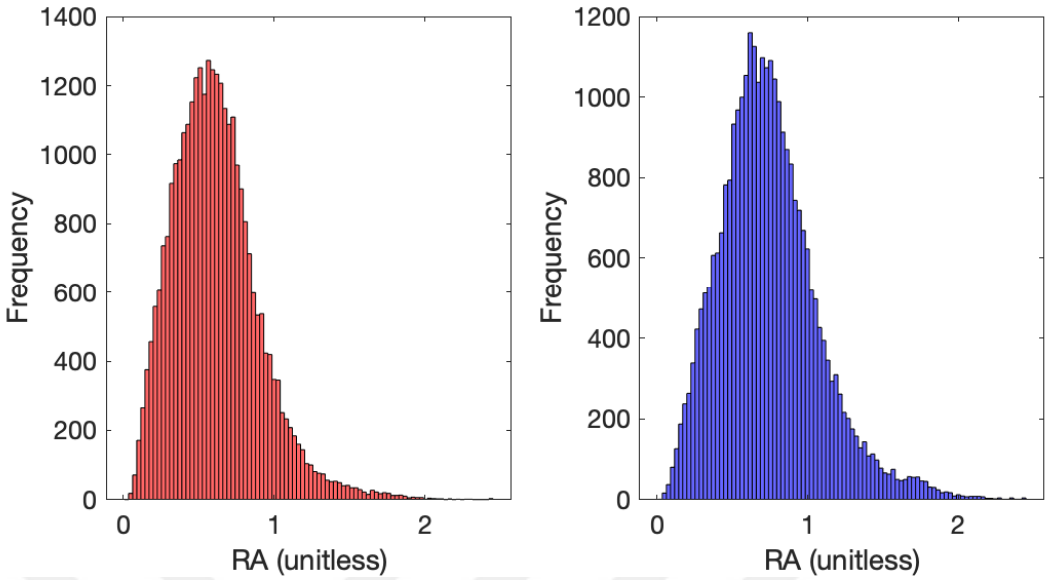


Figure 2.29 90–bin histogram distribution of RA in NBS and Healthy WM.

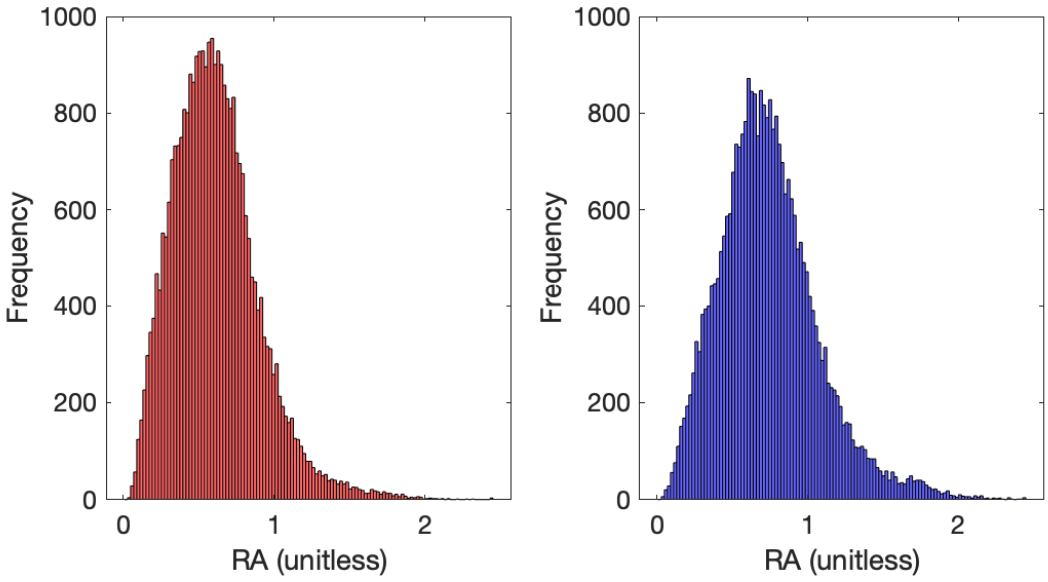


Figure 2.30 120–bin histogram distribution of RA in NBS and Healthy WM

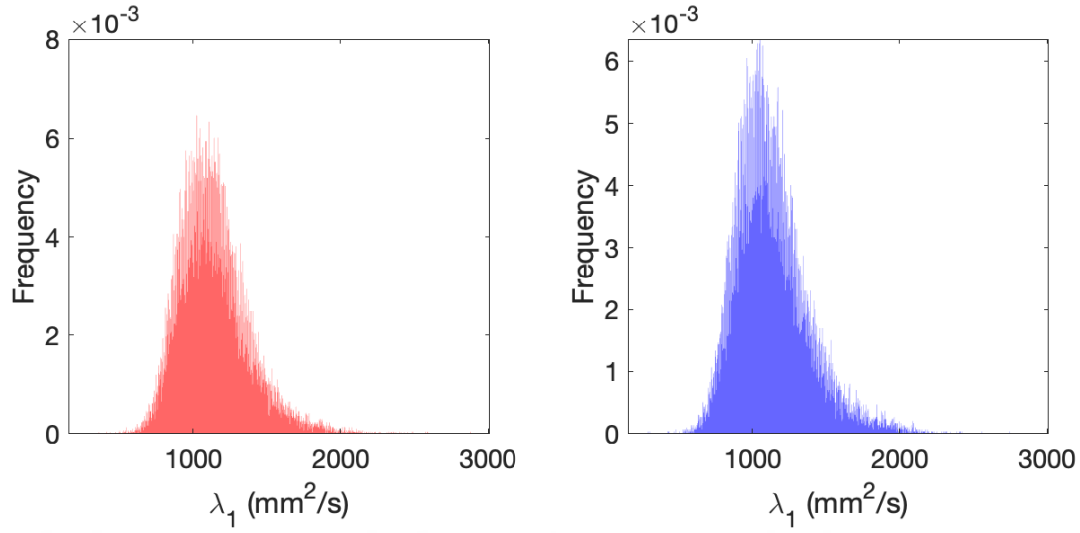


Figure 2.31 1000–bin histogram distribution of λ_1 in NBS and Healthy WM.

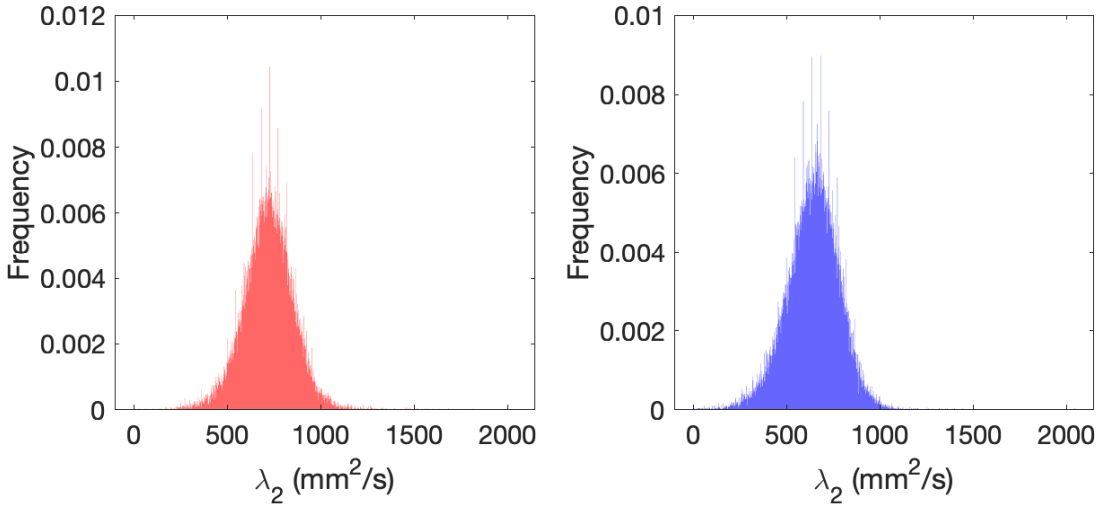


Figure 2.32 1000–bin histogram distribution of λ_2 in NBS and Healthy WM.

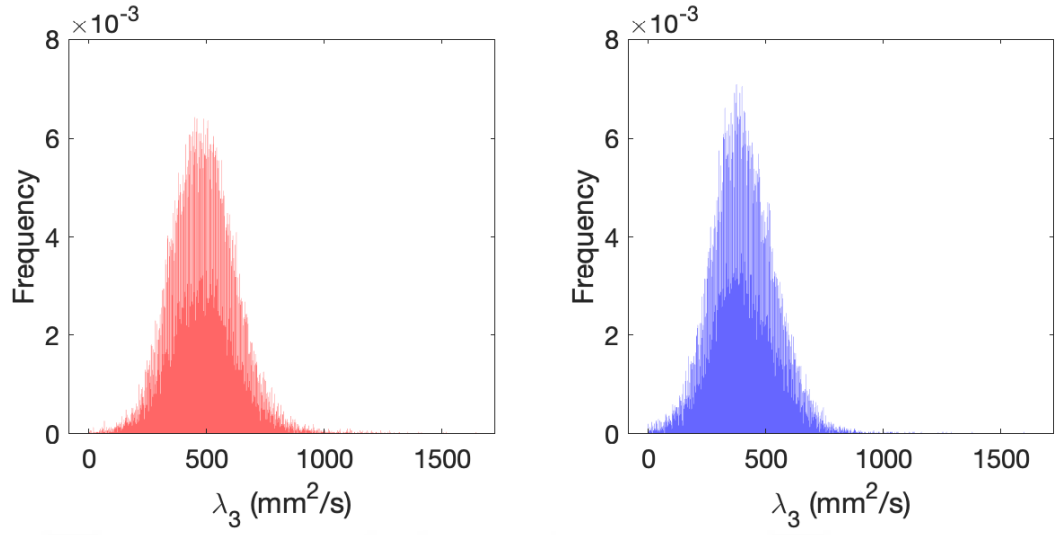


Figure 2.33 1000–bin histogram distribution of λ_3 in NBS and Healthy WM.

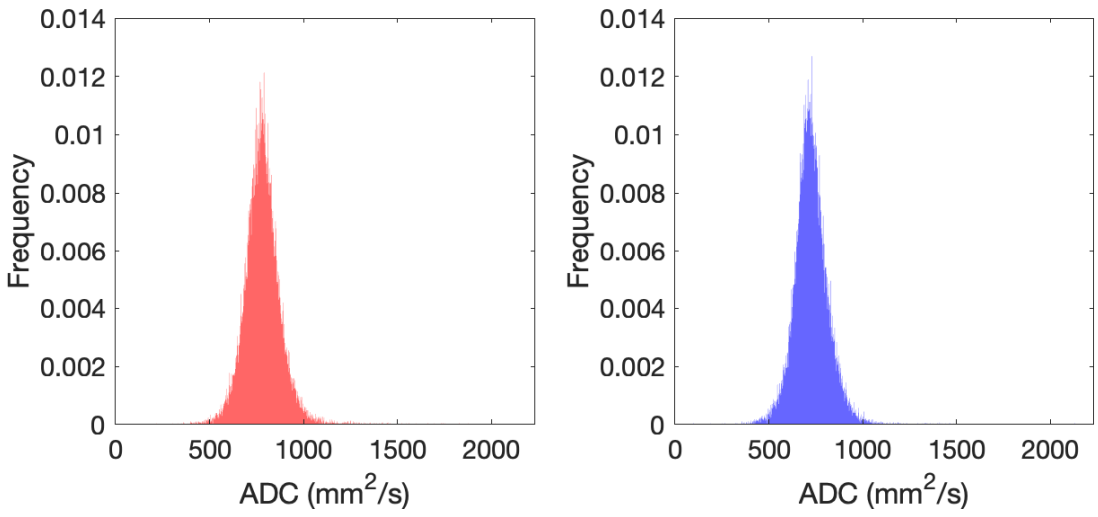


Figure 2.34 1000–bin histogram distribution of ADC in NBS and Healthy WM.

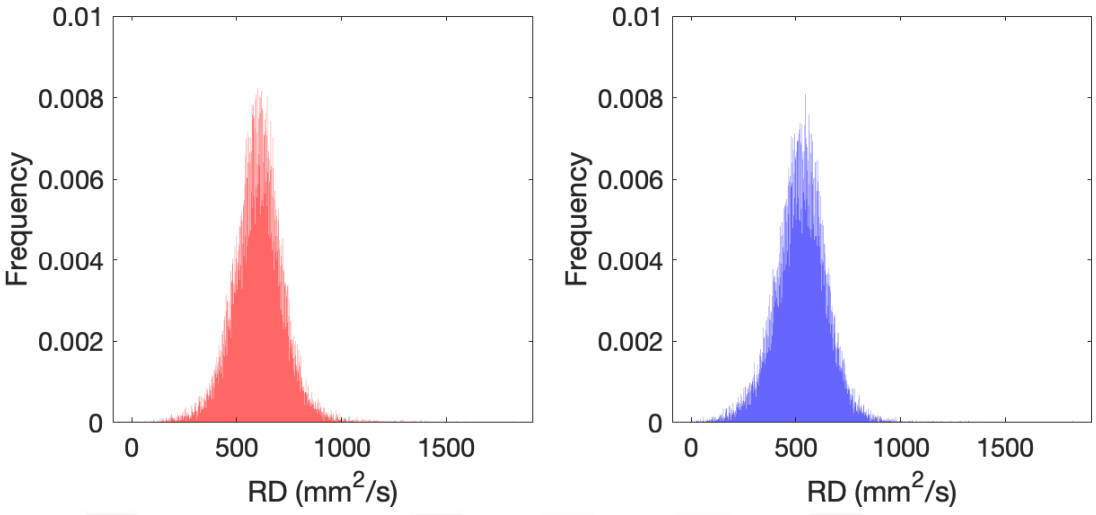


Figure 2.35 1000–bin histogram distribution of RD in NBS and Healthy WM.

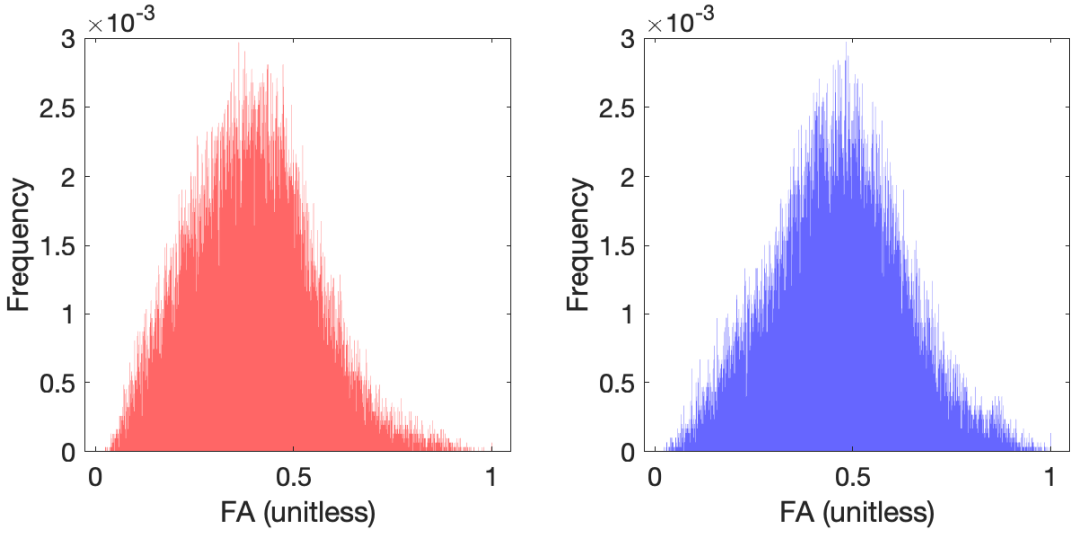


Figure 2.36 1000–bin histogram distribution of FA in NBS and Healthy WM.

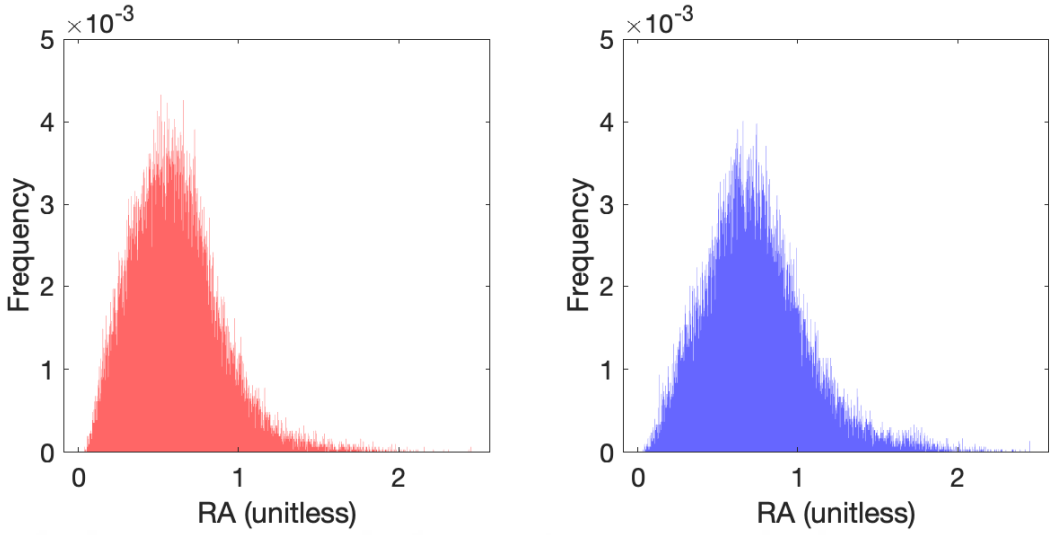


Figure 2.37 1000–bin histogram distribution of RA in NBS and Healthy WM.

2.8 Histogram Based Feature Extraction

Each DTI derived scalar metric (λ_1, λ_2, λ_3, ADC, RD, FA, RA) was structured as a cell array for both the NBS and control groups, where each cell contained the white matter voxel values corresponding to an individual subject. For the calculations,Eq. 2.5.,Eq. 2.6.,Eq. 2.7. let,

$$x \in \{\lambda_1, \lambda_2, \lambda_3, \text{ADC}, \text{RD}, \text{FA}, \text{RA}\} \tag{2.5}$$

$$i \in \{1, 2, \ldots, N\} \tag{2.6}$$

$$k \in \{1, \ldots, K_i\} \tag{2.7}$$

where x denotes a DTI metric; i indexes subjects (N total, including NBS and healthy controls); k indexes WM pixels, and K_i is the total number of WM pixels for subject i.

For histogram based feature extraction, subject level minimum and maximum values were computed. For each subject i, let $x_i(k)$ denote the value of DTI metric x at the kth pixel and, the minimum,Eq. 2.8. and maximum Eq. 2.9, values over the WM mask be:

$$X_i = \max_{k \in \{1,\ldots,K_i\}} x_i(k) \tag{2.8}$$

$$Y_i = \min_{k \in \{1,\ldots,K_i\}} x_i(k) \tag{2.9}$$

Afterwards, group–level minimum, Eq. 2.10. and maximum values, Eq. 2.11. were obtained by defining the minimum and maximum values over all subjects as

$$X = \max_{i \in \{1,\ldots,N\}} X_i \tag{2.10}$$

$$Y = \min_{i \in \{1,\ldots,N\}} Y_i \tag{2.11}$$

For each DTI metric, the interval $[Y\, X]$ was used as the histogram range. Using these global bounds, bin edges were defined for each metric to generate 30, 60, 90, and 120 equal width bins. The same interval was uniformly subdivided according to the number of bins ($n_{bins} = 30\,, 60\,, 90\,, 120$) for determining the bin locations and the histograms were calculated as specified by these locations.

The histogram bin values formed n_{bins} dimensional row vectors, one for each subject, i.e. for the ith subject, h_i. These row vectors were stacked up to form the input data as represented by Eq. 2.12.

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}_1^\top \\ \mathbf{h}_2^\top \\ \vdots \\ \mathbf{h}_N^\top \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} & \dots & h_{1\,n_{\text{bins}}} \\ h_{21} & h_{22} & \dots & h_{2\,n_{\text{bins}}} \\ \vdots & \vdots & \ddots & \vdots \\ h_{N1} & h_{N2} & \dots & h_{N\,n_{\text{bins}}} \end{bmatrix} \tag{2.12}$$

2.9 Histogram Normalization

As the number of voxels are different for each subject, subjects with higher voxel counts may become disproportionately represented in the analysis, thereby introducing potential bias and creating the false impression of greater importance solely due to increased data volume. Moreover, if one feature has large values and another has small ones, the large one dominates the distance or decision function which negatively impacts on the ML model. Without normalization, biased data size is inevitable, and this causes poor ML performance models.

For each histogram matrix corresponding to a DTI metric and bin size, the raw bin counts were first normalized row–wise to the [0, 1] range by dividing each row by its sum.

2.10 Mean Centering of Histogram Features

For each row of the feature matrix Eq.2.12, the mean value across all subjects was computed and represented in Eq. 2.13. and Eq. 2.14., respectively.

$$\mathbf{h}_{\text{mean}} = \frac{1}{N} \sum_{l=1}^{N} \mathbf{h}_l \tag{2.13}$$

$$\mathbf{h}_{c,i} = \mathbf{h}_i - \mathbf{h}_{\text{mean}}, \qquad i = 1, \ldots, N \tag{2.14}$$

thereby centering each feature column. This step resulted in values within the range $[-0.5, 0.5]$.

Finally, the feature matrix was constructed as shown in Eq. 2.15. Eq. 2.15.

$$\mathbf{Z} = \begin{bmatrix} \mathbf{h}_{c,1} & y_1 \\ \mathbf{h}_{c,2} & y_2 \\ \vdots & \vdots \\ \mathbf{h}_{c,N} & y_N \end{bmatrix}, \qquad y_i = \text{label}(i) \in \{0, 1\} \text{ (NBS} = 1,\ \text{Control} = 0). \tag{2.15}$$

2.11 DTI Variable Combination

For further expanding the dimension of the feature space, the features from individual DTI metrics were concatenated with the following combinations: (λ_1–λ_2, λ_1–λ_3, λ_2–λ_3, λ_1–λ_2–λ_3). A total of 11 features, comprising both individual and combined diffusion metrics, were provided as input for the ML analysis. As an example, λ_1–λ_2 combination input matrix is formed as in Eq. 2.16. Eq. 2.16.

$$\mathbf{Z}_{(\lambda_1,\lambda_2)} = \begin{bmatrix} \mathbf{h}_{c,1}^{\lambda_1} & \mathbf{h}_{c,1}^{\lambda_2} & y_1 \\ \mathbf{h}_{c,2}^{\lambda_1} & \mathbf{h}_{c,2}^{\lambda_2} & y_2 \\ \vdots & \vdots & \vdots \\ \mathbf{h}_{c,N}^{\lambda_1} & \mathbf{h}_{c,N}^{\lambda_2} & y_N \end{bmatrix}, \qquad y_i = \text{label}(i) \in \{0, 1\} \text{ (NBS} = 1,\ \text{Control} = 0). \tag{2.16}$$

where $h_{ci}^{\lambda_1}$, $h_{ci}^{\lambda_2}$ denote the feature row vectors of subject i for λ_1 and λ_2 respectively.

2.12 Classification Models

ML models were assessed using MATLAB® Statistics and Machine Learning Toolbox, the standard setup. To achieve the highest accuracy, features were initially evaluated individually, and then combined based on their performance results. The classification learner application within the toolbox was utilized with the provided input data. Prior to running all ML models, the cross–validation strategy was selected. Leave–One–Out Cross–Validation (LOOCV) was employed, which corresponds to setting the number of folds equal to the number of samples. This choice maximizes the use of available data for model evaluation. The classification learner enabled to explored supervised ML models; Fine Tree, Medium Tree, Coarse Tree, Linear Discriminant, Quadratic Discriminant, Binary GLM Logistic Regression, Efficient Logistic Regression, Efficient Linear SVM, Gaussian Naive Bayes, Kernel Naive Bayes, Linear SVM, Quadratic SVM, Cubic SVM, Fine SVM, Medium Gaussian SVM, Coarse Gaussian SVM, Fine KNN, Medium KNN, Coarse KNN, Cosine KNN, Cubic KNN, Weighted KNN, Ensemble (Boosted Trees), Ensemble (Bagged Trees), Ensemble (Subspace Discriminant), Ensemble (Subspace KNN), Ensemble (RUSBoosted Trees), Narrow Neural Network, Medium Neural Network, Wide Neural Network, Bilayered Neural Network, Trilayered Neural Network, SVM Kernel, and Logistic Regression Kernel algorithms to classify the data. The accuracy for each classification model obtained with and without principal component analysis (PCA) as well as adjusting the component number, explained variance value and cross validation.

2.13 Classifier Performance

To compare the performance of the models, the Receiver Operating Characteristic (ROC) curve and the Area Under the Curve (AUC) were used. The model's accuracy was calculated by comparing the predicted labels with the true labels, and a confusion matrix summarized the classification performance across the two classes. When selecting the optimal model for each variable, priority was given to the vari-

able combination achieving the highest accuracy with the fewest components. Models achieving an accuracy below 85% were classified as unsuccessful, represented in Figure 3.1.

3. RESULTS

3.1 Statistical Results

Quantitative analysis of DTI metrics revealed significant differences between the NBS and control cohorts in several WM parameters. All statistical results were summarized in Table 3.1.

λ_2 (p=0.000001), axial diffusivity (λ_1) (p=0.000007), (λ_3) (p=0.002343) with ADC (p=0.000001) and RD (p=0.000027) showed statistically significant differences between the NBS and control cohort. λ_2 showed the highest significant difference, followed by λ_1. Among the individual eigenvalues of the diffusion tensor, mean values of λ_1, λ_2, and λ_3 all showed statistically significant differences between groups. Specifically, λ_1 was significantly higher in the NBS group , as were λ_2 and λ_3, suggesting widespread microstructural alterations. Accordingly, among three raw eigenvalues of the diffusion tensor λ_2 was the most significant one.

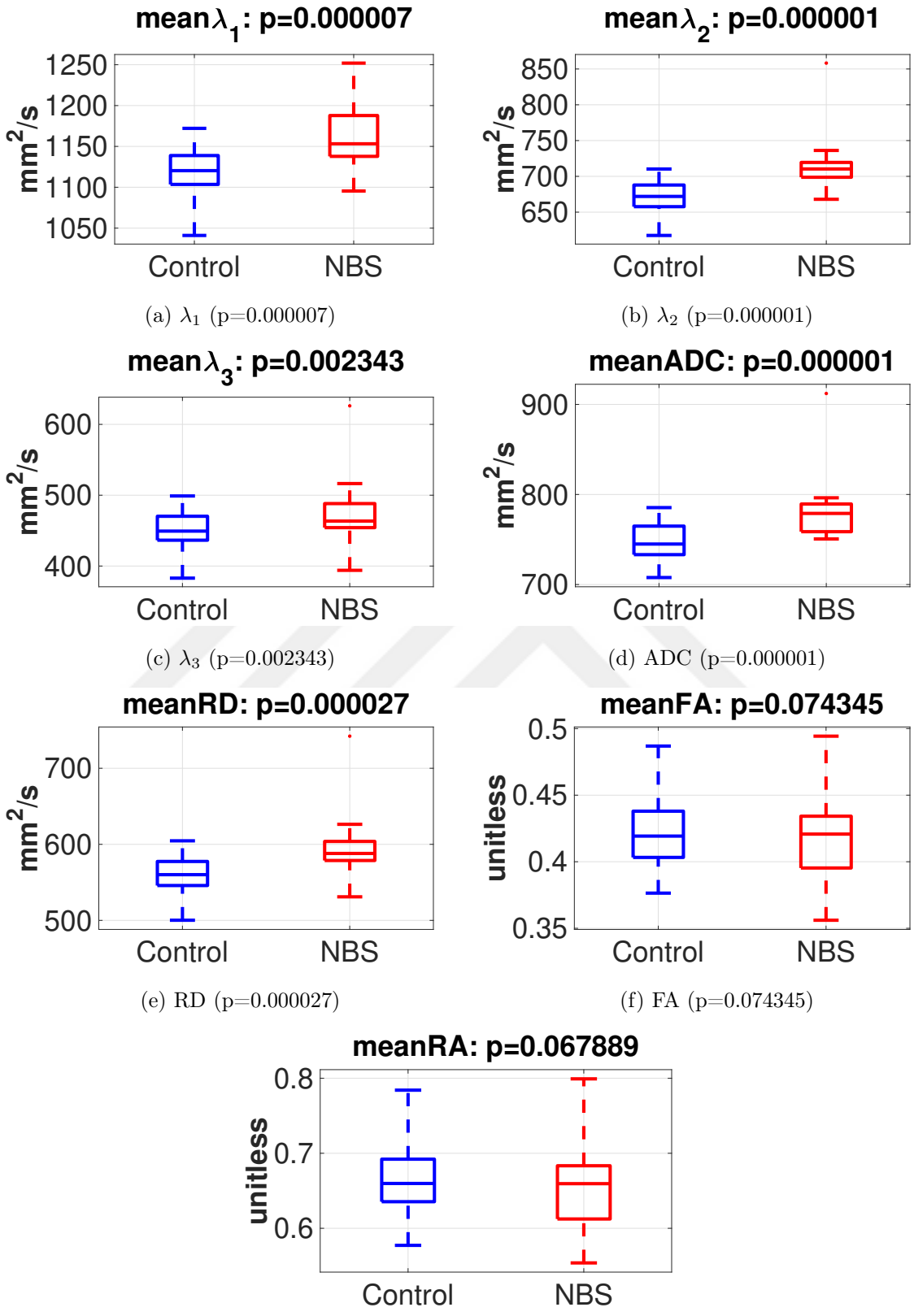
The significance of ADC is driven by the weight of λ_2 and λ_1, indicating greater overall diffusivity, and the significance of RD is affected by λ_3, which is slightly higher than ADC. Furthermore, the standard reference indices RA and FA became completely insignificant; they did not show statistically significant group differences, with p-values of 0.074345 and 0.067889, respectively.

Table 3.1 summarizes the results of the Mann–Whitney U test where the threshold level was $p<0.05$ to compare WM differences from NBS WM and healthy WM based on the mean DTI metrics in the WM. These observations suggest that diffusivity–based metrics, especially λ_2 and λ_1, may serve as more sensitive indicators of microstructural abnormalities associated with NBS. While the distributions of λ_1 and λ_2 are well separated, contributing to the distinct profiles seen in ADC and RD, the overlapping distribution of λ_3 dampens the separation in RD and leads to near–complete overlap

between the distributions of FA and RA.

Table 3.1
Statistical significance of group differences for DTI metrics between NBS and control cohorts.

DTI variable	p**-value**
Mean λ_1	0.000007
Mean λ_2	0.000001
Mean λ_3	0.002343
Mean ADC	0.000001
Mean RD	0.000027
Mean FA	0.074345
Mean RA	0.067889



(a) λ_1 (p=0.000007)

(b) λ_2 (p=0.000001)

(c) λ_3 (p=0.002343)

(d) ADC (p=0.000001)

(e) RD (p=0.000027)

(f) FA (p=0.074345)

(g) RA (p=0.067889)

Figure 3.1 Box plots comparing the mean values of DTI metrics between the NBS and control cohorts, extracted from white matter regions. Panels (a)–(c) show the eigenvalues $\lambda_1, \lambda_2, \lambda_3$; (d) ADC; (e) RD; and (f), (g) the anisotropy indices FA and RA, respectively. Reported p-values indicate the statistical significance of the observed group differences.

3.2 Machine Learning Results

Validation accuracies for the most discriminative DTI variables were obtained using the Classification Learner app in MATLAB® (Statistics and Machine Learning Toolbox). Models with validation accuracy $\geq 85\%$ were retained, and the corresponding optimal histogram bin sizes, with and without PCA dimensionality reduction, are summarized below.

Table 3.2
Validation accuracy for λ_1.

Best model(s)	**Accuracy (%)**	$n_{\mathbf{bins}}$	**PCA components**
Fine Tree Coarse Tree Medium Tree Linear SVM	96.2	30	2
Fine Tree Coarse Tree Medium Tree Quadratic Discriminant	94.2	60	4
Linear Discriminant Trilayered Neural Network	94.2	90	6
Bilayered Neural Network	94.2	120	8

For λ_1, the highest validation accuracy of 96.2% was achieved using 30 bins and 2 PCA components by Fine Tree, Coarse Tree, Medium Tree, and Linear SVM. Increasing the bin count to 60 (with 4 PCA components) yielded 94.2% accuracy for the same tree-based models plus a Quadratic Discriminant classifier. At 90 bins (6 components), both Linear Discriminant Analysis and the Trilayered Neural Network attained 94.2%. Finally, with 120 bins and 8 PCA components, the Bilayered Neural Network reached 94.2% accuracy.

Table 3.3
Validation accuracy for λ_2.

Best model(s)	**Accuracy (%)**	$n_{\mathbf{bins}}$	**PCA components**
Quadratic SVM	94.2	30	2
Linear Discriminant Quadratic Discriminant	94.2	60	4
Linear Discriminant	94.2	90	6
Subspace KNN Ensemble	98.1	120	8

For λ_2, the Quadratic SVM achieved 94.2% validation accuracy using 30 histogram bins and 2 PCA components. With 60 bins and 4 components, both Linear and Quadratic Discriminant reached 94.2%. At 90 bins (6 components), Linear Discriminant again attained 94.2%. Finally, the Subspace KNN Ensemble peaked at 98.1% with 120 bins and 8 PCA components.

Table 3.4
Validation accuracy for λ_3.

Best model(s)	**Accuracy (%)**	$n_{\mathbf{bins}}$	**PCA components**
Subspace KNN	86.5	30	2
Binary GLM Logistic Regression Medium Gaussian SVM Medium Neural Network Bilayered Neural Network	94.2	60	4
Narrow Neural Network Medium Neural Network	100	90	6
Binary GLM Logistic Regression Medium Neural Network Wide Neural Network	98.1	120	8

For λ_3, the Subspace KNN classifier achieved 86.5% validation accuracy using 30 bins and 2 PCA components. Increasing the bin count to 60 (with 4 PCA components), Binary GLM Logistic Regression, Medium Gaussian SVM, Medium Neural Network, and Bilayered Neural Network all attained 94.2% accuracy. At 90 bins (6 components), both the Narrow Neural Network and Medium Neural Network reached 100% accuracy. Finally, with 120 bins and 8 PCA components, Binary GLM Logistic Regression, Medium Neural Network, and Wide Neural Network achieved 98.1% accuracy.

Table 3.5
Validation accuracy for ADC.

Best model(s)	**Accuracy (%)**	$n_{\mathbf{bins}}$	**PCA components**
Efficient Linear SVM; Quadratic SVM; Cubic SVM; Coarse Gaussian SVM; Cubic KNN	100	30	2
Linear Discriminant; Quadratic Discriminant; Linear SVM; Binary GLM Logistic Regression; Gaussian Naive Bayes; Kernel Naive Bayes; Quadratic SVM; Cubic SVM; Medium Gaussian SVM; Fine KNN; Weighted KNN; Narrow Neural Network; Medium Neural Network; Wide Neural Network; Bilayered Neural Network; Trilayered Neural Network; Kernel SVM	100	60	4
Linear Discriminant; Quadratic Discriminant; Cubic SVM; Binary GLM Logistic Regression; Gaussian Naive Bayes; Kernel Naive Bayes; Linear SVM; Quadratic SVM; Fine KNN; Weighted KNN; Subspace Discriminant Ensemble; Narrow Neural Network; Medium Neural Network; Wide Neural Network; Bilayered Neural Network; Trilayered Neural Network; Kernel SVM	100	90	6
Linear Discriminant; Quadratic Discriminant; Binary GLM Logistic Regression; Gaussian Naive Bayes; Kernel Naive Bayes; Linear SVM; Quadratic SVM; Cubic SVM; Fine KNN; Subspace Discriminant Ensemble; Subspace KNN Ensemble; Wide Neural Network; Trilayered Neural Network	100	120	8

For ADC, every evaluated classifier attained 100% validation accuracy across all bin sizes. With 30 bins and 2 PCA components, Efficient Linear SVM, Quadratic SVM, Cubic SVM, Coarse Gaussian SVM, and Cubic KNN reached 100%. At 60 bins (4 PCA components), all tested models—Linear and Quadratic Discriminant, Linear SVM, Binary GLM Logistic Regression, Gaussian and Kernel Naive Bayes, Quadratic and Cubic SVM, Medium Gaussian SVM, Fine and Weighted KNN, and neural networks from Narrow to Trilayered—achieved 100%. The same complete set of classifiers maintained 100% accuracy at 90 bins with 6 components. Finally, at 120 bins (8 components), Linear and Quadratic Discriminant, Binary GLM Logistic Regression, Naive Bayes variants, SVMs, KNN ensembles, and Wide and Trilayered Neural Networks also sustained 100% accuracy.

Table 3.6
Validation accuracy for RD.

Best model(s)	**Accuracy (%)**	$n_{\mathbf{bins}}$	**PCA components**
Kernel Naive Bayes	78.8	30	2
Coarse Tree Kernel Naive Bayes	84.6	60	4
Coarse Tree	84.6	90	6
RUSBoosted Trees Ensemble	88.5	120	8

For RD, Kernel Naive Bayes achieved 78.8% validation accuracy with 30 bins and 2 PCA components. With 60 bins (4 components), Coarse Tree and Kernel Naive Bayes reached 84.6%. At 90 bins (6 components), Coarse Tree again attained 84.6%. Finally, with 120 bins and 8 components, the RUSBoosted Trees Ensemble peaked at 88.5%.

Table 3.7
Validation accuracy for FA.

Best model(s)	**Accuracy (%)**	$n_{\mathbf{bins}}$	**PCA components**
Gaussian Naive Bayes	78.8	30	2
Quadratic SVM	86.5	60	4
Bilayered Neural Network	84.6	90	6
Quadratic SVM	84.6	120	8

For FA, Gaussian Naive Bayes achieved 78.8% validation accuracy with 30 bins and 2 PCA components. Increasing to 60 bins (4 components), Quadratic SVM reached 86.5%. At 90 bins (6 components), the Bilayered Neural Network attained 84.6%. Finally, with 120 bins and 8 components, Quadratic SVM achieved 84.6% accuracy.

Table 3.8
Validation accuracy for RA.

Best model(s)	**Accuracy (%)**	$n_{\mathbf{bins}}$	**PCA components**
Fine Tree Coarse Tree Medium Tree	84.6	30	2
Fine Tree Medium Tree Coarse Tree RUSBoosted Trees Ensemble	80.8	60	4
Coarse Tree	88.5	90	6
Quadratic SVM RUSBoosted Trees Ensemble	80.8	120	8

For RA, Fine Tree, Coarse Tree, and Medium Tree achieved 84.6% validation accuracy

with 30 bins and 2 PCA components. With 60 bins (4 components), Fine Tree, Medium Tree, Coarse Tree, and the RUSBoosted Trees Ensemble attained 80.8%. At 90 bins (6 components), Coarse Tree reached 88.5%. Finally, with 120 bins and 8 components, Quadratic SVM and the RUSBoosted Trees Ensemble achieved 80.8% accuracy.

Table 3.9
Validation accuracy for $\lambda_1,\ \lambda_2$.

Best model(s)	**Accuracy (%)**	$n_{\mathbf{bins}}$	**PCA components**
Gaussian Naive Bayes Linear SVM Quadratic SVM	98.1	30	3
Fine Tree Medium Tree Coarse Tree Gaussian Naive Bayes Kernel Naive Bayes Linear SVM Bagged Trees Ensemble Medium Neural Network Wide Neural Network	96.2	60	6
Linear Discriminant Binary GLM Logistic Regression Kernel Naive Bayes Linear SVM Medium Gaussian SVM Subspace Discriminant Ensemble Subspace KNN Ensemble Narrow Neural Network Medium Neural Network Wide Neural Network Bilayered Neural Network	98.1	90	9
Kernel Naive Bayes Cubic SVM Medium Gaussian SVM Subspace Discriminant Ensemble Subspace KNN Ensemble	98.1	120	12

For the combined features λ_1 and λ_2, Gaussian Naive Bayes, Linear SVM, and Quadratic SVM achieved 98.1% validation accuracy with 30 bins and 3 PCA components. Increasing to 60 bins (6 components), Fine Tree, Medium Tree, Coarse Tree, Gaussian and Kernel Naive Bayes, Linear SVM, Bagged Trees Ensemble, and Medium and Wide Neural Networks attained 96.2%. At 90 bins (9 components), Linear Discriminant Analysis, Binary GLM Logistic Regression, Kernel Naive Bayes, Linear SVM, Medium Gaussian SVM, Subspace Discriminant and Subspace KNN Ensembles, plus Narrow, Medium, Wide, and Bilayered Neural Networks reached 98.1%. Finally, with 120 bins and 12 PCA components, Kernel Naive Bayes, Cubic SVM, Medium Gaussian SVM, and Subspace Discriminant and Subspace KNN Ensembles maintained 98.1% accuracy.

Table 3.10
Validation accuracy for $\lambda_1,\ \lambda_3$.

Best model(s)	**Accuracy (%)**	$n_{\mathbf{bins}}$	**PCA components**
Fine KNN	92.3	30	3
Quadratic SVM Medium Gaussian SVM	100	60	9
Linear Discriminant Binary GLM Logistic Regression Kernel Naive Bayes Linear SVM Medium Gaussian SVM Subspace Discriminant Ensemble Subspace KNN Ensemble Narrow Neural Network Medium Neural Network Wide Neural Network Bilayered Neural Network	98.1	90	9
Linear SVM Quadratic SVM Cubic SVM Narrow Neural Network Medium Neural Network Wide Neural Network Bilayered Neural Network	100	120	12

For the combined features λ_1 and λ_3, Fine KNN achieved 92.3% validation accuracy with 30 bins and 3 PCA components. At 60 bins (9 components), Quadratic SVM and Medium Gaussian SVM both attained 100%. Increasing to 90 bins (9 components), Linear Discriminant Analysis, Binary GLM Logistic Regression, Kernel Naive Bayes, Linear SVM, Medium Gaussian SVM, Subspace Discriminant Ensemble, Sub-

space KNN Ensemble, and Narrow/Medium/Wide/Bilayered Neural Networks reached 98.1%. Finally, with 120 bins and 12 PCA components, Linear/Quadratic/Cubic SVMs and Narrow/Medium/Wide/Bilayered Neural Networks maintained 100% accuracy.

Table 3.11
Validation accuracy for $\lambda_2,\ \lambda_3$.

Best model(s)	**Accuracy (%)**	$n_{\mathbf{bins}}$	**PCA components**
Subspace KNN Ensemble	94.2	30	3
Linear Discriminant Kernel Naive Bayes Quadratic SVM Subspace KNN Ensemble Bilayered Neural Network	94.2	60	6
Quadratic SVM Medium Gaussian SVM	100	90	9
Fine Tree Medium Tree Coarse Tree Linear Discriminant Binary GLM Logistic Regression Gaussian Naive Bayes Kernel Naive Bayes Linear SVM Quadratic SVM Cubic SVM Bagged Trees Ensemble Subspace Discriminant Ensemble Narrow Neural Network Medium Neural Network Wide Neural Network Bilayered Neural Network	100	120	12

For the combined features λ_2 and λ_3, Subspace KNN Ensemble achieved 94.2% validation accuracy with 30 bins and 3 PCA components. At 60 bins (6 components), Linear Discriminant Analysis, Kernel Naive Bayes, Quadratic SVM, Subspace KNN Ensemble, and Bilayered Neural Network also attained 94.2%. Increasing to 90 bins (9 components), Quadratic SVM and Medium Gaussian SVM reached 100%. Finally, with 120 bins and 12 components, Fine Tree, Medium Tree, Coarse Tree, Linear Discriminant Analysis, Binary GLM Logistic Regression, Gaussian Naive Bayes, Kernel Naive Bayes, Linear SVM, Quadratic SVM, Cubic SVM, Bagged Trees Ensemble, Subspace Discriminant Ensemble, and Narrow/Medium/Wide/Bilayered Neural Networks all maintained 100% accuracy.

Table 3.12
Validation accuracy for $\lambda_1,\ \lambda_2,\ \lambda_3$.

Best model(s)	**Accuracy (%)**	$n_{\mathbf{bins}}$	**PCA components**
Linear Discriminant; Quadratic Discriminant; Binary GLM Logistic Regression; Linear SVM; Quadratic SVM; Cubic SVM; Medium Gaussian SVM; Coarse Gaussian SVM; Fine KNN; Weighted KNN; Subspace Discriminant Ensemble; Narrow Neural Network; Medium Neural Network; Wide Neural Network; Bilayered Neural Network; Trilayered Neural Network; Kernel SVM	100	30	3
Linear SVM	96.2	60	9
Fine Tree; Medium Tree; Coarse Tree; Linear Discriminant; Binary GLM Logistic Regression; Gaussian Naive Bayes; Kernel Naive Bayes; Linear SVM; Quadratic SVM; Cubic SVM; Bagged Trees Ensemble; Subspace Discriminant Ensemble; Subspace KNN Ensemble; Narrow Neural Network; Medium Neural Network; Wide Neural Network; Bilayered Neural Network; Trilayered Neural Network	100	90	12
Linear Discriminant; Binary GLM Logistic Regression; Linear SVM; Quadratic SVM; Cubic SVM; Bagged Trees Ensemble; Subspace Discriminant Ensemble; Subspace KNN Ensemble; Narrow Neural Network; Medium Neural Network; Wide Neural Network	100	120	15

For the combined features λ_1, λ_2, and λ_3, at 30 bins with 3 PCA components, a broad set of classifiers—including Linear and Quadratic Discriminant Anal-

ysis, Binary GLM Logistic Regression, Linear/Quadratic/Cubic SVM, Medium and Coarse Gaussian SVM, Fine and Weighted KNN, the Subspace Discriminant Ensemble, Kernel SVM, and neural networks from Narrow through Trilayered—achieved 100% validation accuracy. With 60 bins and 9 PCA components, Linear SVM attained 96.2%. At 90 bins with 12 components, Fine/Medium/Coarse Trees, Linear Discriminant Analysis, Binary GLM Logistic Regression, Gaussian and Kernel Naive Bayes, Linear/Quadratic/Cubic SVM, Bagged Trees Ensemble, Subspace Discriminant and Subspace KNN Ensembles, and Neural Networks (Narrow through Trilayered) again reached 100%. Finally, at 120 bins with 15 components, Linear Discriminant Analysis, Binary GLM Logistic Regression, Linear/Quadratic/Cubic SVM, Bagged Trees Ensemble, Subspace Discriminant and Subspace KNN Ensembles, and neural networks (Narrow, Medium, and Wide) maintained 100% accuracy.

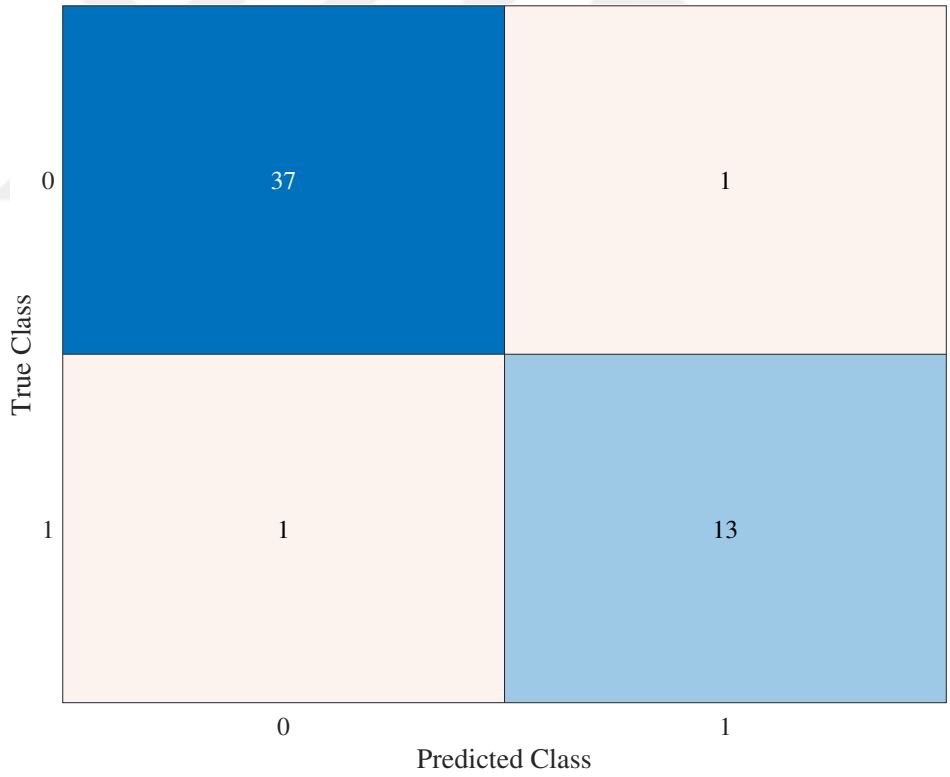


Figure 3.2 Confusion matrix for λ_1 (Linear SVM).

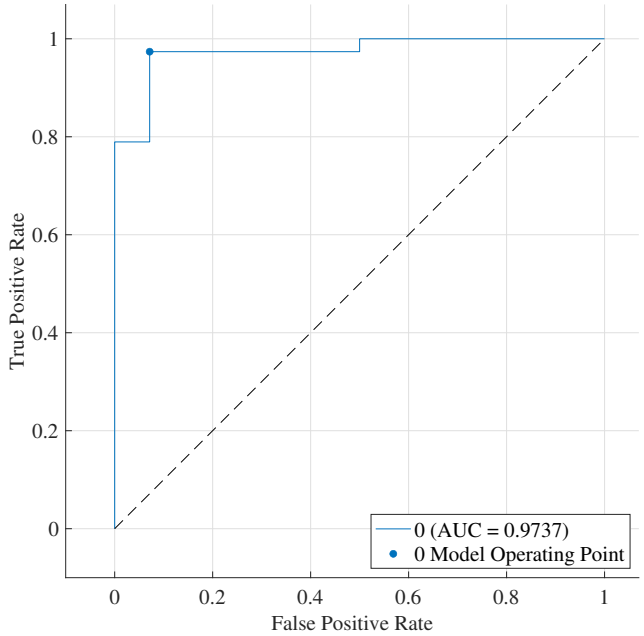


Figure 3.3 ROC curve for λ_1 (Linear SVM).

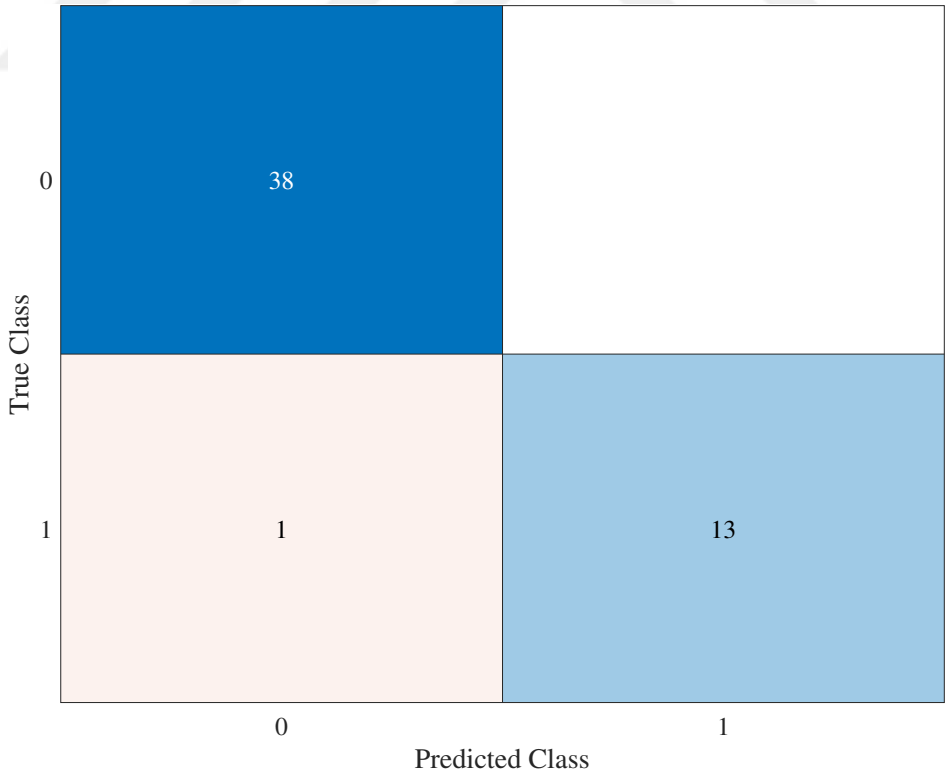


Figure 3.4 Confusion matrix for λ_2 (Linear SVM).

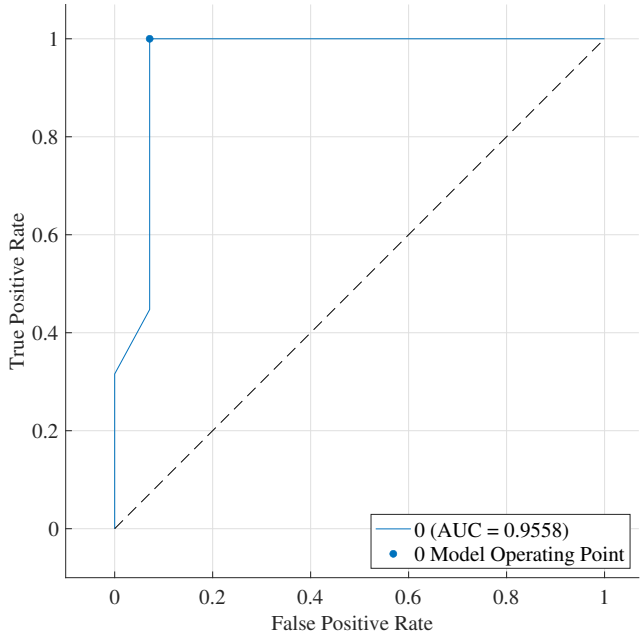


Figure 3.5 ROC curve for λ_2 (Linear SVM).

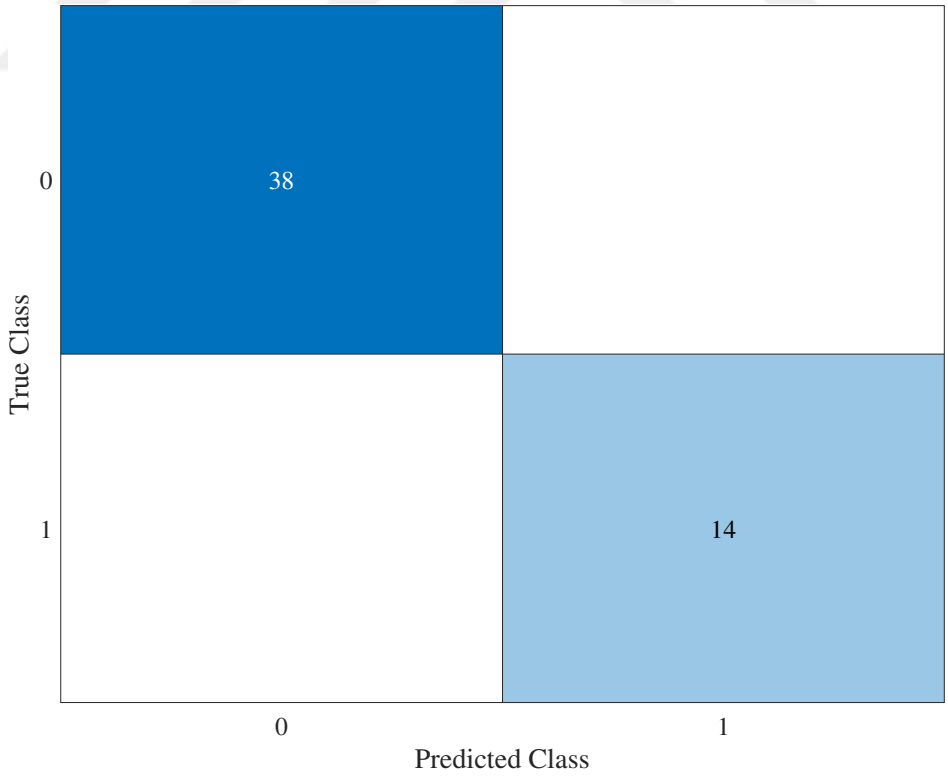


Figure 3.6 Confusion matrix for λ_3 (Narrow Neural Network).

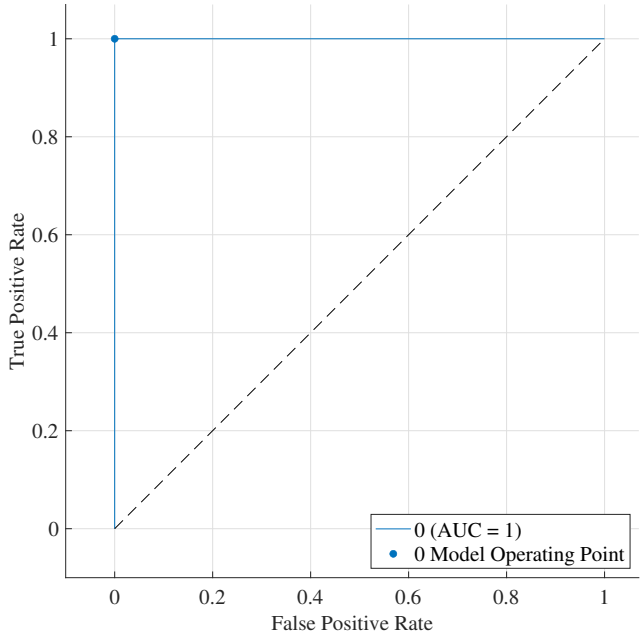


Figure 3.7 ROC curve for λ_3 (Narrow Neural Network).

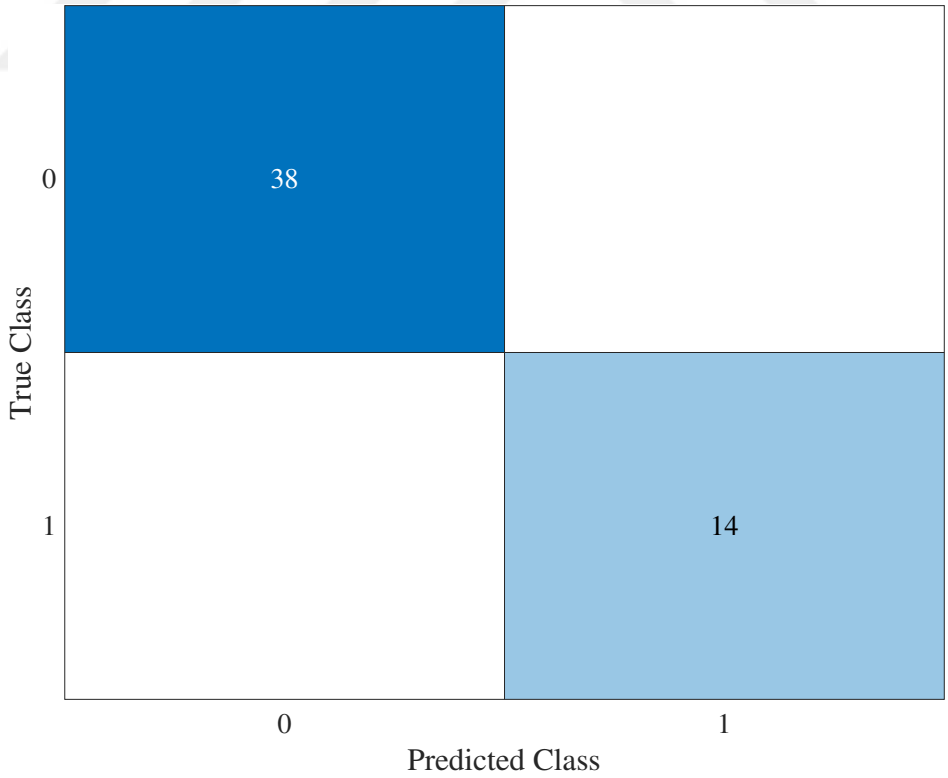


Figure 3.8 Confusion matrix for ADC (Linear SVM).

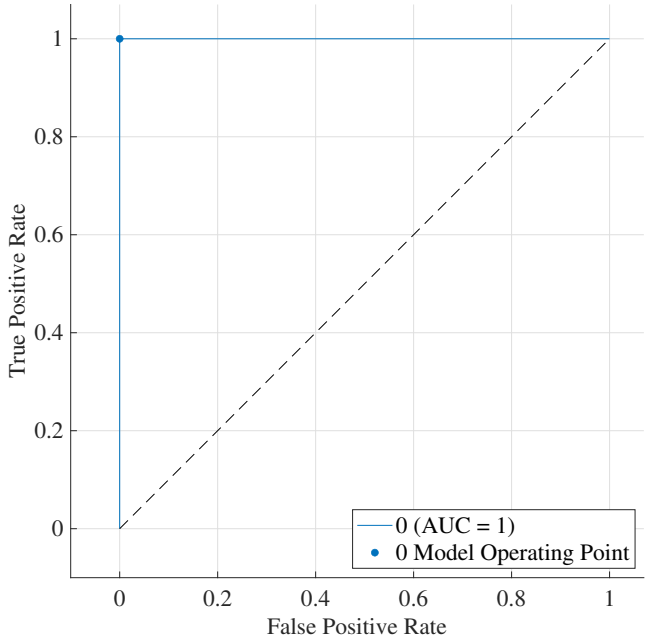


Figure 3.9 ROC curve for ADC (Linear SVM).

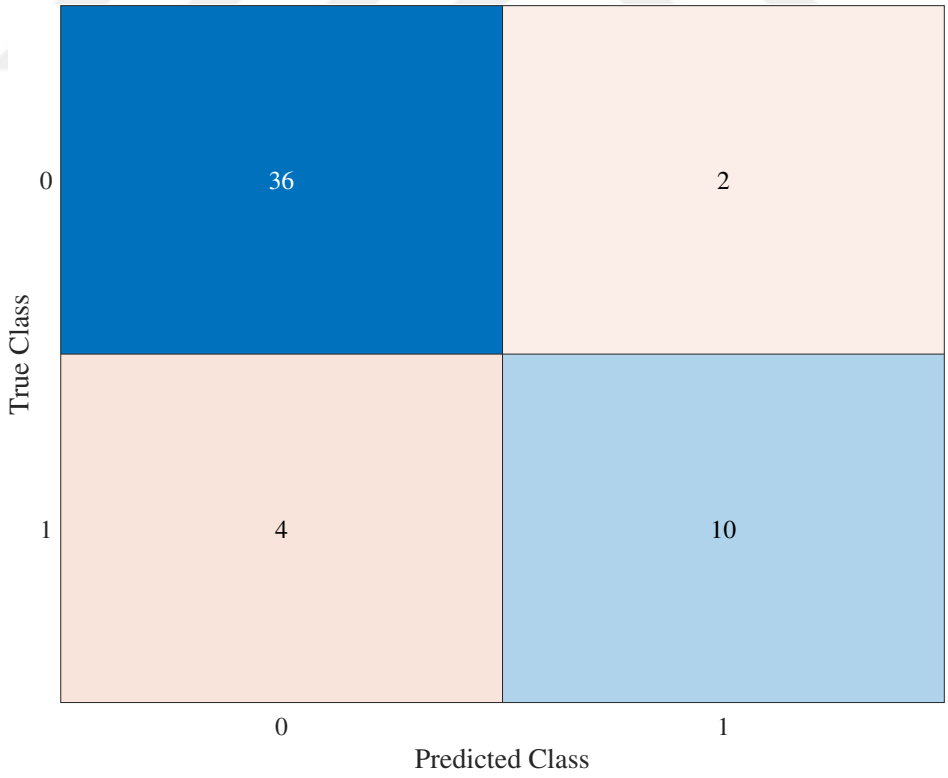


Figure 3.10 Confusion matrix for RD (RUSBoosted Trees).

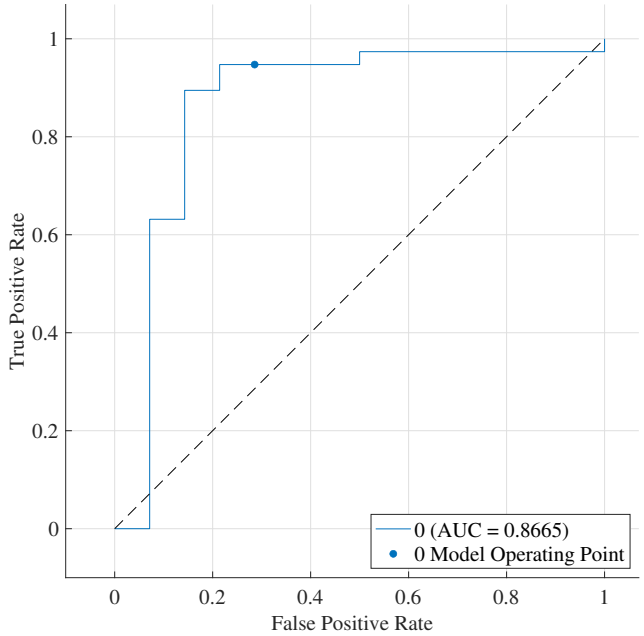


Figure 3.11 ROC curve for RD (RUSBoosted Trees).

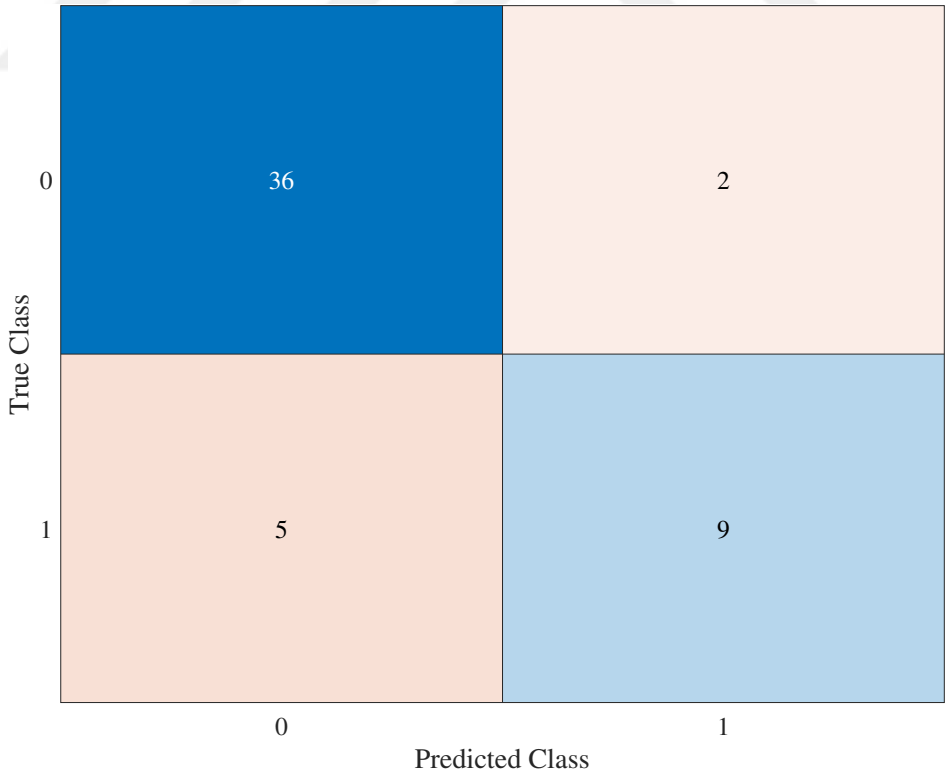


Figure 3.12 Confusion matrix for FA (Quadratic SVM).

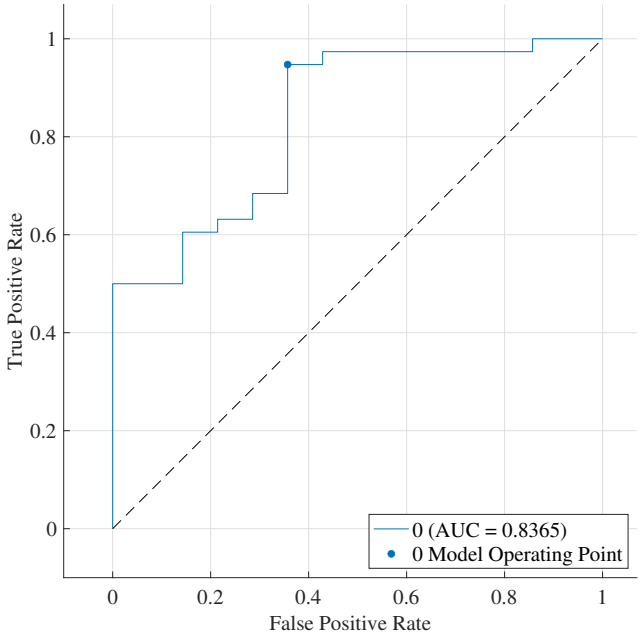


Figure 3.13 ROC curve for FA (Quadratic SVM).

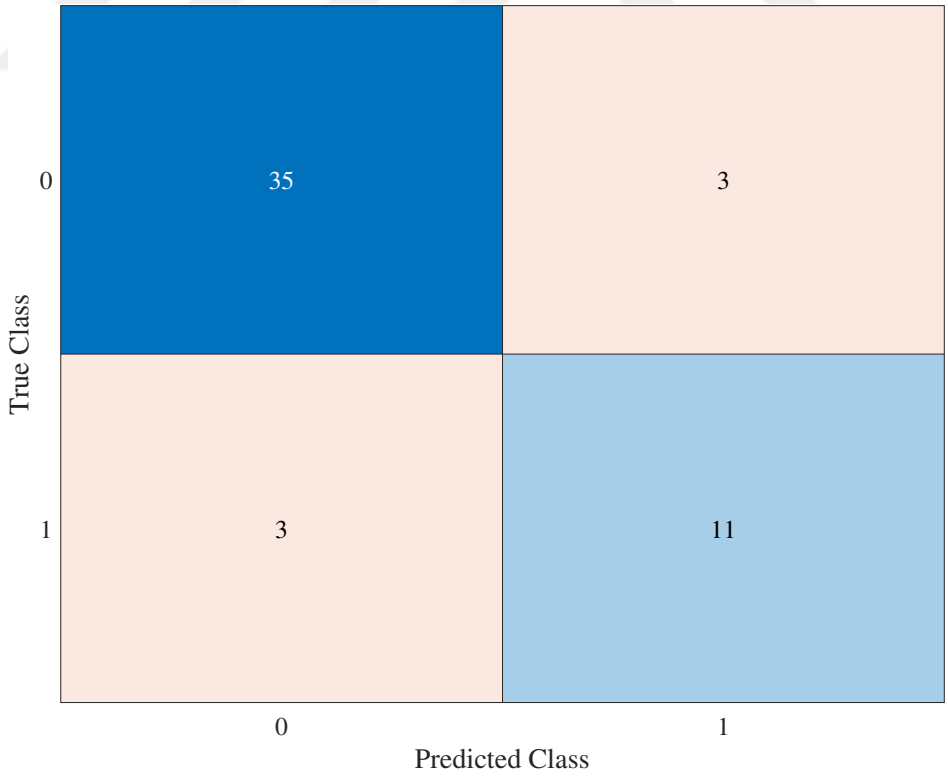


Figure 3.14 Confusion matrix for RA (Coarse Tree).

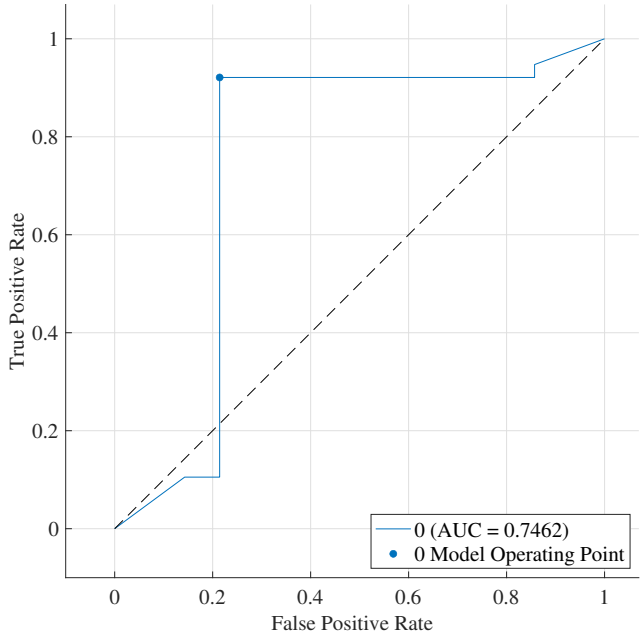


Figure 3.15 ROC curve for RA (Coarse Tree).

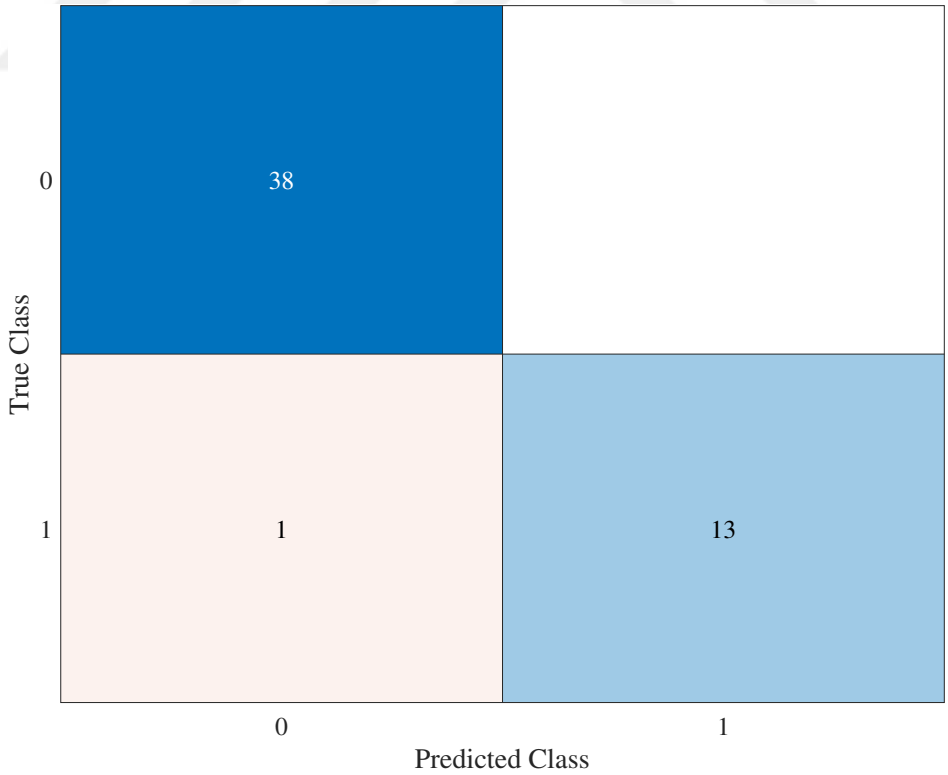


Figure 3.16 Confusion matrix for λ_1, λ_2 (Linear SVM).

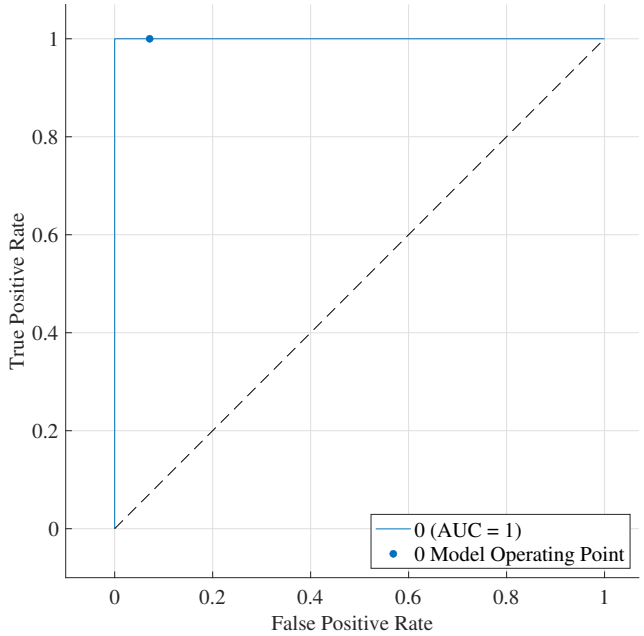


Figure 3.17 ROC curve for $\lambda_1,\ \lambda_2$ (Linear SVM).

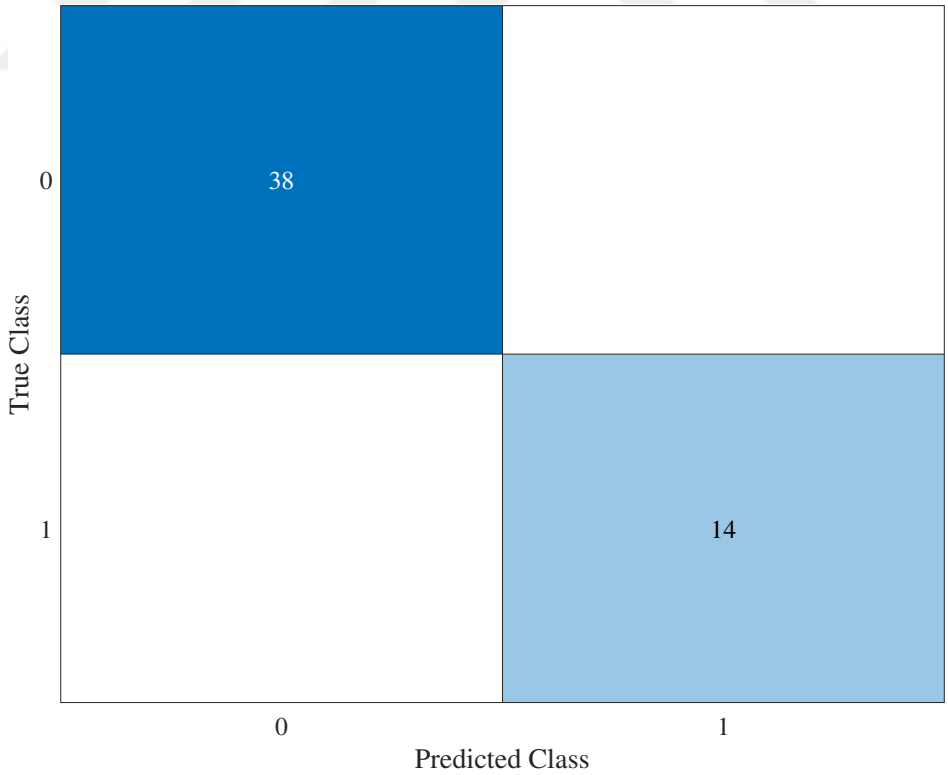


Figure 3.18 Confusion matrix for $\lambda_1,\ \lambda_3$ (Linear SVM).

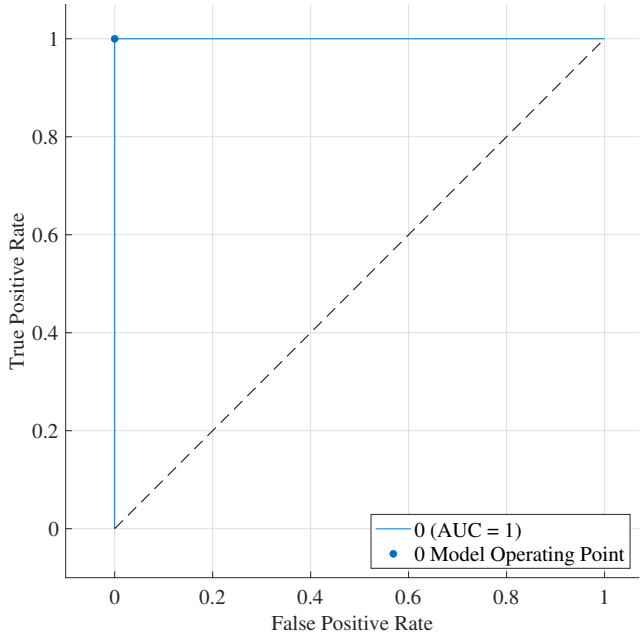


Figure 3.19 ROC curve for $\lambda_1,\ \lambda_3$ (Linear SVM).

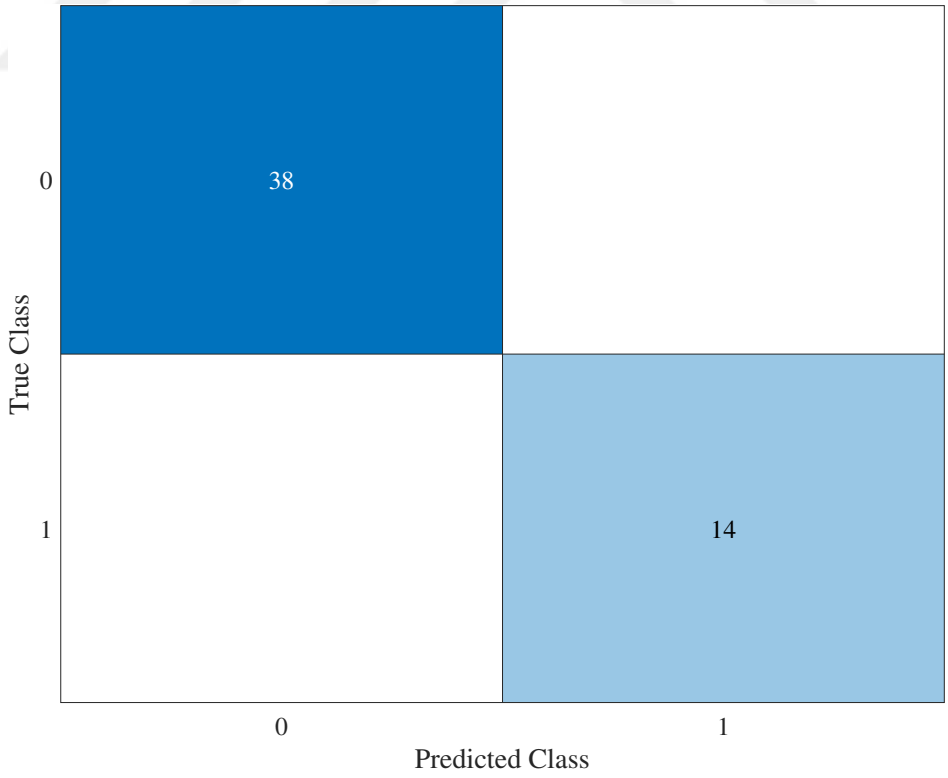


Figure 3.20 Confusion matrix for $\lambda_2,\ \lambda_3$ (Linear SVM).

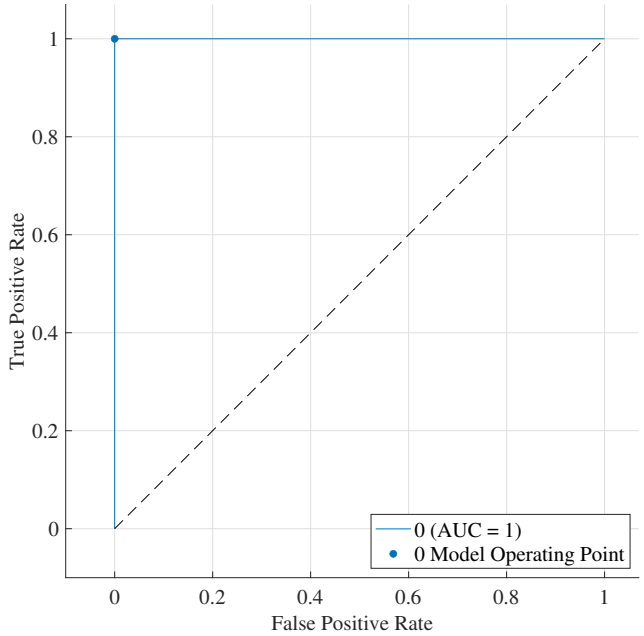


Figure 3.21 ROC curve for $\lambda_2,\ \lambda_3$ (Linear SVM).

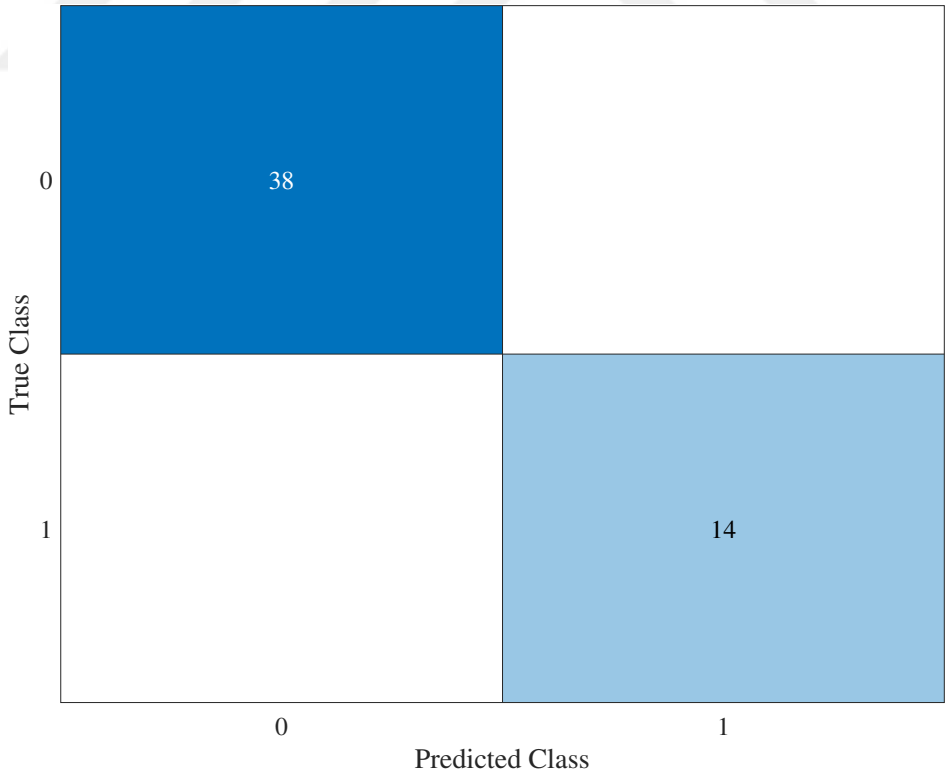


Figure 3.22 Confusion matrix for $\lambda_1,\ \lambda_2,\ \lambda_3$ (Linear SVM).

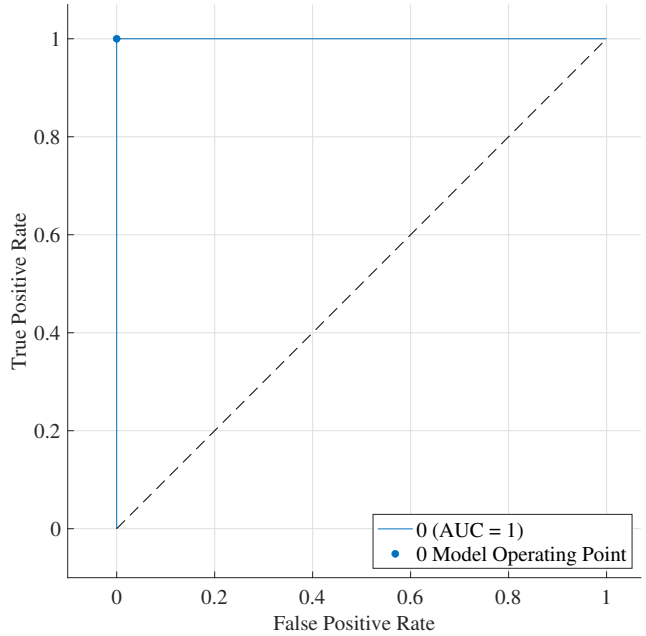


Figure 3.23 ROC curve for λ_1, λ_2, λ_3 (Linear SVM).

The performances of the DTI metrics and corresponding best ML classifier has shown via confusion matrices and ROC curves.

In Figure 3.2 The confusion matrix (A) demonstrates the effectiveness of the Linear SVM classifier in distinguishing NBS patients from healthy control subjects based on histogram features derived from the λ_1 DTI metric. The model achieved a high rate of correct classifications for both groups, as evidenced by the strong diagonal dominance in the confusion matrix, with minimal misclassification. Additionally, the corresponding ROC curve (B) confirmed the model's robustness, displaying a high true positive rate across varying thresholds. The AUC, which approaches 1.0, suggests discriminatory performance of the classifier.

The classification results based on the λ_2 metric and a Linear SVM classifier in Figure 3.4 and Figure 3.5 demonstrated discriminatory performance between NBS patients and healthy controls. The confusion matrix (A) indicated that the model classified the majority of both groups, with minimal misclassification. Furthermore,

the ROC curve (B) illustrates high separability, with the curve approaching the top–left corner, suggesting a good sensitivity and specificity.

The classification performance based on the λ_3 metric using a Narrow Neural Network achieved outstanding results as shown in Figure 3.6 and Figure 3.7. As evidenced by the confusion matrix (A), the classifier attained 100% sensitivity and specificity, indicating that all subjects were correctly classified without any misclassification. Furthermore, the ROC curve (B) corresponding to λ_3 demonstrated an ideal shape with an area under the curve (AUC) of 1.00, underscoring the model's robustness.

In Figure 3.8 and Figure 3.9 the classification performance based on the ADC using a Linear SVM classifier demonstrated strong discriminative power between NBS patients and healthy controls. The confusion matrix (A) indicates a balanced and accurate classification, while the ROC curve (B) shows a sharp increase and high AUC, reflecting excellent sensitivity and specificity.

The classification performance of the RD feature using the RUSBoosted Trees classifier. The confusion matrix demonstrates a strong classification capability, with a high number of correctly identified NBS and control subjects. Notably, the ROC curve for this model exhibits a sharp increase near the origin and quickly approaches the upper–left corner, indicating excellent discriminative ability. This steep rise reflects a high true positive rate achieved with minimal false positives, confirming the effectiveness of RD in capturing the group–level differences in white matter integrity.

The classification performance based on FA using a Quadratic SVM model was moderate compared to other DTI metrics. By the confusion matrix , although the classifier correctly identified a majority of samples, it exhibited a less balanced distribution between true positives and true negatives. This is further reflected in the ROC curve, which presented a more gradual slope and a lower AUC, suggesting reduced discriminative capability between the NBS and control groups.

The RA metric, when classified using the Coarse Tree model, demonstrated

modest discriminative ability between the NBS and control groups. The confusion matrix suggests moderate classification accuracy, with a tendency toward misclassification in one of the classes. This is further supported by the ROC curve, which exhibits a relatively gradual rise and does not reach ideal sensitivity–specificity trade–offs.

The combination of λ_1 and λ_2 as input features to a Linear SVM classifier yielded a reasonably balanced classification performance. The confusion matrix, the model demonstrated fair sensitivity and specificity, correctly identifying a substantial proportion of both NBS and control subjects. The ROC curve supports this observation, showing a moderately sloped increase in the true positive rate across a range of thresholds.

The combination of λ_1 and λ_3 features using a Linear SVM classifier showed a moderate classification performance. The confusion matrix revealed an acceptable level of discrimination between NBS and control groups, with a balanced true positive and true negative rate. The ROC curve for this combination indicated a gradual increase in the true positive rate across various threshold values, reflecting a steady, but less sharp, classifier confidence.

The combination of λ_2 and λ_3 was evaluated using a Linear SVM classifier. The confusion matrix revealed a strong separation between the NBS and control groups, with only a few misclassifications. The model achieved high sensitivity and specificity, indicating its robustness in detecting both true positives and true negatives. The ROC curve further supported this finding, showing a consistently rising curve and a high AUC value.

The combination of all three eigenvalue metrics, λ_1, λ_2, and λ_3 evaluated using a Linear SVM classifier, yielded highly favorable classification performance. The confusion matrix demonstrates near–perfect separation between the NBS and control groups, and the ROC curve exhibits a sharp initial rise with an extensive plateau approaching the top–left corner, suggesting both high sensitivity and specificity.

4. DISCUSSION

In this study, WM differences between individuals with NBS and healthy controls were investigated using both statistical analysis and ML approaches.

The presented study identified significant differences in WM microstructure between the NBS and healthy individuals using DTI metrics. Statistical analysis with the Mann–Whitney U test revealed that several DTI metrics demonstrated highly significant group differences, while anisotropy indices did not reach conventional thresholds for statistical significance. Among all investigated parameters, λ_2 ($p = 0.000001$) exhibited the strongest statistical distinctive between groups, followed closely by axial diffusivity λ_1 ($p = 0.000007$). These findings highlighted the potential sensitivity of the individual eigenvalues of the diffusion tensor, particularly λ_2, in detecting microstructural abnormalities associated with NBS. Additionally, λ_3 ($p = 0.002343$) also showed a statistically significant difference, though to a lesser extent. The increase in eigenvalues (λ_1, λ_2, and λ_3) in the NBS group suggests a pattern of widespread diffusivity alterations, potentially indicative of abnormalities in WM development and microstructural organization in patients with NBS. The derived scalar measures ADC ($p = 0.000001$) and RD ($p = 0.000027$) also demonstrated significant group differences, consistent with the elevated eigenvalue measures.

The significance of ADC appears to be predominantly driven by changes in λ_1 and λ_2, reflecting an increase in overall diffusivity. RD which is derived from λ_2 and λ_3, was also significantly elevated but to a slightly lesser extent. This is likely due to λ_3's comparatively weaker significant difference in separating NBS and the control cohort. Interestingly, traditional anisotropy measures FA ($p = 0.074345$) and RA ($p = 0.067889$) did not reach statistical significance compared to other DTI metrics. Although there was a trend toward lower FA and RA values in the NBS group, the differences did not meet the $p < 0.05$ threshold. This may be due to overlapping distributions between groups, as observed in Figure 3.1, particularly for λ_3, which

strongly influences anisotropy computations. The reduced sensitivity of FA and RA in this context aligns with prior findings suggesting that anisotropy based measures may be less effective in detecting subtle or diffuse WM abnormalities, especially in neurological conditions as NBS.

These findings highlighted the enhanced sensitivity of the primary diffusivity metrics, particularly λ_2 and λ_1, in detecting white matter abnormalities within the NBS cohort. Owing to their pronounced statistical distinction and their substantial contribution to derived indices such as ADC and RD, these eigenvalues may represent more reliable and informative biomarkers for evaluating white matter integrity in both clinical and research settings. Moreover, the clearer separation observed in the distributions of the eigenvalues compared to those of FA and RA reinforces the notion that anisotropy based measures alone may fail to capture critical microstructural alterations. In summary, the statistical analysis demonstrated that diffusivity based DTI metrics offer superior sensitivity in detecting WM abnormalities associated with NBS. The findings advocate for the incorporation of individual eigenvalue analysis in particular, λ_2 in future neuroimaging studies of developmental disorders, as these metrics may better capture the underlying microstructural disruptions than traditional anisotropy indices.

In evaluating the DTI variables in terms of classification performance, the variables were investigated both individually and combined. Remarkably, ADC consistently reached 100.00% classification accuracy across all bin settings. These findings reinforce the statistical results from the Mann-Whitney U test, where λ_3 and ADC were among the most significantly different metrics between the two groups. The repeated top performance of λ_2-λ_3 and λ_1-λ_3 combinations both achieving 100.00% accuracy across most of the bin levels highlights the effectiveness of integrating multiple eigenvalue components in multivariate models. Such combinations likely capture a broader representation of underlying microstructural abnormalities. Among individual metrics, λ_3 demonstrated robust classification performance, with accuracies consistently reaching up to 100% at 90 bins with 6 PCA components. Therefore, the best discriminative eigenvalue classifier is λ_3. The effectiveness of λ_3 when combined with other eigenvalues

positions as sensitive and stable biomarker for NBS–related white matter abnormalities. In contrast, anisotropy based metrics such as FA and RA consistently showed lower classification accuracy relative to the diffusivity metrics as they demonstrated low significance in statistical analysis. Despite minor improvements at higher bin resolutions, their best performances (maximum 86.5% for FA and 84.6% for RA) remained below those of other metrics. This performance outcomes may reflect the reduced sensitivity of FA and RA to subtle or isotropic microstructural disruptions, particularly when the overall magnitude of diffusion changes without substantially altering directional anisotropy. Furthermore, overlapping distributions of FA and RA values between groups, as suggested in earlier statistical plots, likely contributed to their reduced separability. An important observation is that classification accuracy generally improved with increasing bin size, particularly for eigenvalue–based combinations. The higher dimensional histograms at 90 and 120 bins provided richer representations of distributional properties, which may have enabled ML models to more effectively differentiate patterns between groups. However, the increase in bin size also coincided with an increase in the number of components used in the feature space. This suggests that the imputable of feature extraction plays a critical role in capturing subtle group differences–especially when using flexible classifiers such as neural networks and ensemble methods. Across all experiments, neural network architectures, SVMs and ensemble classifiers emerged as the most effective ML strategies. Their ability to model non–linear boundaries and adapt to high dimensional input spaces likely contributed to the high classification performance observed.

The classification performance of the top three selected feature sets was further evaluated using confusion matrices and ROC curves. Each feature set was chosen based on its ability to achieve 100.00% classification accuracy with a minimal number of components, reflecting both high discriminative power and model simplicity. λ_3 (90 bins, 6 components) showed excellent performance, with a perfectly separable confusion matrix and an ROC curve indicating ideal sensitivity and specificity. This confirms the strong complementarity between λ_3 in capturing microstructural variations relevant to NBS pathology, even with a limited number of features and a relatively coarse histogram resolution. Similarly, the ADC metric at all bins also produced flawless

classification. Its confusion matrix showed no misclassifications, and the ROC curve approached the ideal top–left corner, indicating high diagnostic potential. Given that ADC integrates multiple eigenvalue contributions, its performance further validates the significance of diffusivity driven metrics in distinguishing between the NBS and control cohorts.

The classifier performance underscore that both focused eigenvalue pairs and strategically combined DTI variables can serve as predictors for white matter abnormalities in NBS.

In summary, ML results underscore the discriminative power of diffusivity–based DTI metrics, especially λ_3 and ADC in detecting WM alterations in NBS. Multivariate combinations involving these metrics further enhanced classification accuracy, often achieving perfect separation between cohorts. Finally, the findings also suggested that bin size and component dimensionality should be carefully considered to optimize classification performance in histogram based DTI feature extraction approaches.

5. CONCLUSION

The statistical analysis demonstrated that inter–diffusivity indices, i.e., FA and RA, were insignificant for NBS, which carries potentially cautionary information for clinical practice as well as regarding the biology of NBS. The high differentiating significance of the λ_2 followed by the largest diffusivity significance indicates perhaps that more relevant than the damage along the axons, the WM integrity is more significantly compromised in the axons transversally. The significance of ADC is comparable to that of λ_2, and this may be largely attributed to its dependence on both λ_2 and λ_1. Further pathological studies might reveal the content of microstructural damage, such as the changes in the myelin layers for NBS and/or loss thereof and/or changes in the interstitial space.

This study demonstrated the effectiveness of DTI metrics in classifying white matter abnormalities associated with NBS using a ML–based histogram analysis approach. Both individual and combined evaluations of DTI variables revealed that diffusivity based metrics particularly λ_3, and ADC consistently outperformed traditional anisotropy measures such as FA and RA in distinguishing between the NBS and control groups. The most discriminative feature sets, including λ_3 and ADC alone, achieved perfect classification accuracy using only two components at moderate bin sizes, highlighting their efficiency and strong diagnostic potential. Moreover, results showed that increasing the histogram bin size generally improved classification accuracy by providing richer feature representations, though this came with a trade–off in feature dimensionality. Classifiers capable of capturing complex, non–linear patterns particularly neural networks, SVMs, and ensemble models exhibited superior performance.

The results highlighted the discriminative power of λ_3 and ADC which emerged as the most effective features, each achieving 100.00% classification accuracy with only six components at moderate bin sizes (90 bins). These findings were in line with

statistical significance results from the Mann–Whitney U test and further validated by confusion matrices and ROC curve analyses, confirming both the accuracy and robustness of the classification performance.

In addition to λ_3, the classifiers uniformly identified the ADC as a top feature achieving peak accuracies in Efficient Linear SVM, Quadratic SVM, Cubic SVM, Coarse Gaussian SVM, Cubic KNN, Linear Discriminant, Quadratic Discriminant, Linear SVM, Binary GLM Logistic Regression, Gaussian Naive Bayes, Kernel Naive Bayes, Medium Gaussian SVM, Fine KNN, Weighted KNN, Narrow Neural Network, Medium Neural Network, Wide Neural Network, Bilayered Neural Network, Trilayered Neural Network, Kernel SVM, Subspace Discriminant Ensemble, Subspace KNN Ensemble. Biologically, elevated ADC indicates increased free water content and expanded extracellular space, as occurs with vasogenic edema, perivascular inflammation, and tissue loss. In NBS, vasculitis and chronic inflammation disrupt blood-brain barrier integrity, promoting interstitial fluid accumulation and gliosis; these processes manifest as raised ADC. Thus, while λ_3, the slowest diffusivity as indicative of disrupted barriers to perpendicular water movement, such as myelin damage or extracellular space expansion. ADC captures the global burden of edema and neurodegeneration. The importance of both λ_3 and ADC across diverse ML models underscores a dual–pathology signature in NBS: focal myelin injury coupled with widespread extracellular changes.

Furthermore, the novelty of this study lies in its methodological approach: rather than relying on conventional DTI metrics such as FA or RA, the study demonstrated that integrating individual eigenvalue components and optimizing histogram–based feature extraction significantly improved classification performance. Moreover, by emphasizing models that achieved perfect accuracy with minimal components, the approach offers a practical path toward automated biomarker discovery. Although the sample size of 14 subjects is relatively limited and may have influenced the ML results, it nonetheless represents a considerably larger cohort compared to the majority of studies in the literature on NBS.

In summary, this work presents promising statistical and classification results

when comparing NBS and healthy white matter.

From a statistical perspective, λ_2 and ADC emerged as the most discriminative metrics, whereas in ML analyses, λ_3 and ADC demonstrated the strongest discriminative power. For future work, the complementary findings–where λ_2, representing perpendicular diffusivity, provided the best statistical separation, and λ_3, the smallest eigenvalue, achieved the best ML performance–will be further investigated to better understand their underlying biological and methodological implications as well as increasing patient data size.

5.1 List of publications produced from the thesis

1. *Machine Learning–Based Classification of Neuro Behcet Syndrome Using DTI Eigenvalue Metrics*, Y. Tuzun, H. A. Comert, I. Cevikcan, M. N. Bayer, A. Aderemi, A. Arcan, U. Uygunoglu, A. Dincer, A. Siva, A. Ozcan *33rd IEEE Conference on Signal Processing and Communications (SIU 2025), Extended Abstract*, 2025.

2. *Machine Learning–Based Classification of Neuro Behcet Syndrome Using DTI Eigenvalue Metrics*, Y. Tuzun, H. A. Comert, I. Cevikcan, M. N. Bayer, A. Aderemi, A. Arcan, U. Uygunoglu, A. Dincer, A. Siva, A. Ozcan *ESMRMB 2025, 41st Annual Scientific Meeting*, 2025.

APPENDIX A. MACHINE LEARNING MODELS

A.1 Support Vector Machines

Support vector machines (SVMs) is a supervised ML algorithm which was introduced by Vapnik et al. in 1990s, are a group of algorithms designed to learn how to distinguish between two classes using a set of training data [37].

They have been used in a variety of fields, such as text classification, handwriting recognition, face detection, gene expression analysis, protein homology prediction, and more [68],[69],[70],[71],[72],[73]. A large body of research has expanded on SVM method, leading to versions used for regression, clustering, and factor analysis, all built on the same foundational ideas. The primary goal of the SVM algorithm is to identify a hyperplane that optimally separates data points belonging to different classes. In a two–dimensional space, this hyperplane is a line, while in an n–dimensional space (where n represents the number of features for each observation), it takes the form of a plane. Although there may be several potential hyperplanes capable of separating the classes, the optimal hyperplane is one that maximizes the margin between the two classes. The margin refers to the greatest possible width of the region parallel to the hyperplane that does not contain any data points. The data points that lie closest to this separating hyperplane and define the boundary of this margin are known as support vectors. These support vectors form a subset of the training data that plays a crucial role in determining the position of the optimal hyperplane.

The goal of support vector classification is to efficiently identify a separating hyperplane in a high–dimensional space that offers strong generalization performance. This performance is often evaluated using the concept of the margin, which is the shortest distance between the separating hyperplane and the nearest data points from each class. The optimal hyperplane is the one that maximizes this margin, thereby improving the classifier's ability to generalize [74]. SVM aims to determine the optimal

separating hyperplane that maximizes the margin between two classes, typically labeled as positive and negative, thereby ensuring the greatest possible separation between the two groups.

A.2 Decision Trees

A decision tree is one of the most widely utilized tools for decision–making and analysis. It is typically constructed by mapping out a tree–like structure consisting of branches and leaves, where each branch represents a decision influencing the situation, and each leaf represents a possible outcome. As a form of decision support system, a decision tree provides a visual and analytical means of evaluating various courses of action based on their associated outcomes, resource requirements, and potential utility. It also serves as a method for representing algorithms in a structured format. Depending on the nature of the problem and the specific objectives, there are different types of decision trees which are classification tree, regression tree, decision tree forests, classification and regression tree and K–means clustering.

A classification tree enables the derivation of the most probable outcome based on the available calculated information. This type of decision tree is particularly used in the fields of probability and statistical analysis, where it is used to model and predict categorical outcomes. A regression tree is employed to determine a single, continuous outcome based on multiple input variables. It is particularly useful in domains requiring numerical predictions, such as real estate calculations and market analysis. Decision tree forests are constructed to improve the accuracy and robustness of outcome prediction. This ensemble approach is commonly known as a random forest and is widely used in scenarios requiring reliable classification or regression results. Classification and regression tree is used to derive the most logical prediction, outcomes are estimated based on a set of dependent variables or influencing factors. K–Means Clustering is the least accurate among decision tree methods, as it simply combines all previously identified factors without leveraging advanced optimization or ensemble techniques.

A.3 K–Nearest Neighbors

The K–Nearest Neighbors (KNN) algorithm is a nonparametric and supervised learning method primarily used for classification tasks. It operates by comparing new, unlabeled data points to a labeled training dataset and predicting class membership based on the majority class among the K closest neighbors. The algorithm is favored for its simplicity and low implementation complexity. In classification, KNN uses features of data points to determine their class by measuring proximity–commonly via Euclidean distance for continuous variables. Although straightforward, the choice of the parameter 'K' significantly influences performance, and it is often optimized empirically by testing multiple values. KNN is considered a lazy learning algorithm, as it defers computation until the prediction phase. It does not involve model training in the traditional sense but instead memorizes the training data. As a result, the entire training set must be retained for use during classification or regression [75].

A.4 Naive Bayes

Naive Bayes (NB) is a statistical classification method based on Bayes Theorem. It operates under the assumption of feature independence within a class which means each feature contributes independently to the probability of a given outcome. As a supervised learning algorithm, NB is notable for its simplicity and computational efficiency, particularly in handling categorical data. It calculates prior and likelihood probabilities to determine the conditional probability of a target class. Despite its advantages, NB faces limitations such as the 'zero frequency' problem, where unseen categories in the training data lead to zero probability estimates. This issue is commonly addressed using smoothing techniques such as Laplace estimation. Additionally, the assumption of predictor independence is often unrealistic in practical applications, which may impact the model's accuracy [76].

A.5 Ensemble Methods

Ensemble methods involve training multiple models to address the same problem. Unlike traditional learning approaches that build a single model from the training data, ensemble methods create a collection of models and combine their outputs. This approach is also known as committee–based learning or learning with multiple classifier systems.

An ensemble consists of multiple models referred to as base learners. These base learners are typically generated from training data using a base learning algorithm, such as decision trees, neural networks, or other ML methods. Most ensemble approaches employ a single base learning algorithm to produce base learners of the same type, resulting in homogeneous ensembles. However, some methods utilize multiple learning algorithms to generate diverse base learners of different types, forming heterogeneous ensembles. In such cases, since there is no singular base learning algorithm, the term 'individual learners' or 'component learners' is often used instead. The predictive performance of an ensemble is generally superior to that of its individual base learners. This is one of the key advantages of ensemble methods. In particular, these methods can transform weak learners, which perform only marginally better than random guessing, into highly accurate strong learners. For this reason, base learners are often referred to as weak learners [77].

A.6 Kernel Methods

Kernel methods (KMs) represent a widely used and established set of techniques in ML and pattern recognition. They are grounded in a rigorous mathematical framework which solves nonlinear problems while leveraging the foundational principles of linear algebra. KMs have become a preferred approach across various applied scientific domains (e.g. signal and image processing, computer vision, and geosciences). Since their emergence in the 1990s most notably through the introduction of SVMs and ker-

nel methods have expanded into a broad class of algorithms capable of addressing a wide range of tasks beyond classification. In particular, kernel–based approaches have demonstrated strong performance in regression, interpolation, and function approximation, with methods such as Gaussian Processes (GPs) and support vector regression achieving notable success across diverse application areas [78].

A.7 Neural Network

The fundamental building block of artificial neural networks and deep learning architectures is the perceptron (or neuron), conceptually inspired by the biological neuron. A basic perceptron consists of weights, a bias (or threshold), and a nonlinear activation function, and it can perform binary classification on a given set of inputs. Each input is multiplied by a corresponding weight, the weighted inputs are summed, and this sum is compared with a bias value; the result is then passed through a nonlinear activation function. If the weighted sum exceeds the bias (threshold), the perceptron produces a positive output; otherwise, it produces a negative output.

The activation function is typically a nonlinear step function in the original perceptron, but many alternatives are used in practice, such as the sigmoid, *tanh*, and ReLU (rectified linear unit). The nonlinearity introduced by these functions is essential for learning complex decision boundaries. During training—typically via supervised learning—the perceptron's weights, including the bias, are adjusted to minimize classification error, allowing the model to emphasize the most relevant input features by assigning them higher weights.

While a single perceptron is limited and cannot model complex patterns, stacking multiple neurons across several layers yields artificial neural networks (ANNs), which can handle a wide range of learning tasks [79].

A.8 Logistic Regression

Logistic regression is a widely used supervised learning algorithm for binary classification tasks. It predicts the probability that a given input belongs to a particular class by applying a sigmoid function to a weighted sum of input features. This transformation maps outputs to a probability range between 0 and 1, facilitating threshold based decision making. The model is trained using optimization techniques such as gradient descent, aiming to minimize a loss function that quantifies the error between predicted and actual labels [80].

REFERENCES

1. Hadj Taieb, M. A., H. Slimane, M. Mhiri, R. Ben Dhia, and et al., "Pseudotumoral Neuro-Behcet's disease: case series and review of literature," *Acta Neurologica Belgica*, pp. 1–15, 2024.

2. Sahin, M., J. Correia, A. Campar, C. Ferrão, and et al., "Infections and Behcet disease," in *Infection and Autoimmunity*, pp. 807–822, Elsevier, 2024.

3. Belfeki, N., N. Ghriss, M. Fourati, D. Leclercq, and et al., "Neuro-Behcet's disease: A review," *La Revue de Médecine Interne*, 2024.

4. Pardo-Villamizar, C. A., "Chapter 96 - neurosarcoidosis and Neuro-Behcet's disease,," in *Neurology and Clinical Neuroscience* (Schapira, A. H. V., E. Byrne, S. DiMauro, R. S. J. Frackowiak, and et al., eds.), pp. 1303–1311, Philadelphia: Mosby, 2007.

5. Gündüz, T., M. Kürtüncü, Z. Matur, E. Tüzün, and et al., "Laboratory and clinical correlates of brain atrophy in Neuro-Behcet's disease," *Journal of the Neurological Sciences*, Vol. 414, p. 116831, 2020.

6. Benamour, S., T. Naji, F. Z. Alaoui, H. El-Kabli, and et al., "Neurological involvement in Behcet's disease. 154 cases from a cohort of 925 patients and review of the literature," *Revue Neurologique*, Vol. 162, no. 11, pp. 1084–1090, 2006.

7. Alrefai, A., K. El-Salem, R. Saadeh, and H. Al-Balas, "A case report of cranial polyneuropathy in Behcet's disease," *Neurological Sciences*, Vol. 31, pp. 83–85, 2010.

8. Akman-Demir, G., P. Serdaroglu, B. Tasçi, and Neuro-Behcet Study Group*, "Clinical patterns of neurological involvement in Behcet's disease: evaluation of 200 patients," *Brain*, Vol. 122, no. 11, pp. 2171–2182, 1999.

9. Altunkaynak, Y., Ş. Usta, D. H. Ertem, A. Köksal, and et al., "Cognitive functioning and silent neurological manifestations in Behcet's disease with ocular involvement," *Archives of Neuropsychiatry*, Vol. 56, no. 3, p. 173, 2018.

10. Deniz, O., A. Çayköylü, G. Vural, Y. Albayrak, and et al., "A case study of Neuro-psycho-Behcet's syndrome presenting with psychotic attack," *Clinical Neurology and Neurosurgery*, Vol. 111, no. 10, pp. 877–879, 2009.

11. Yesilot, N., M. Mutlu, O. Gungor, B. Baykal, and et al., "Clinical characteristics and course of spinal cord involvement in Behcet's disease," *European Journal of Neurology*, Vol. 14, no. 7, pp. 729–737, 2007.

12. Lee, H. S., D. Y. Kim, H. Y. Shin, Y.-C. Choi, and et al., "Spinal cord involvement in Behcet's disease," *Multiple Sclerosis Journal*, Vol. 22, no. 7, pp. 960–963, 2016.

13. Beton, O. E., and C. Yucesan, "Primary headache in Behcet's and Neuro-Behcet's disease," *Neurology India*, Vol. 70, no. 1, pp. 254–257, 2022.

14. Yoon, D.-L., Y. J. Kim, B. S. Koo, Y.-G. Kim, and et al., "Neuro-Behcet's disease in south korea: clinical characteristics and treatment response," *International Journal of Rheumatic Diseases*, Vol. 17, no. 4, pp. 453–458, 2014.

15. Oumerzouk, J., R. Klevor, B. Slioui, M. Chraa, and et al., "Cerebral venous thrombosis in Behcet's disease. a report on 24 cases," *Revue Neurologique*, Vol. 178, no. 3, pp. 213–218, 2022.

16. Shi, J., X. Huang, G. Li, L. Wang, and et al., "Cerebral venous sinus thrombosis in Behcet's disease: a retrospective case-control study," *Clinical Rheumatology*, Vol. 37, pp. 51–57, 2018.

17. Aguiar de Sousa, D., T. Mestre, and J. M. Ferro, "Cerebral venous thrombosis in Behcet's disease: a systematic review," *Journal of Neurology*, Vol. 258, pp. 719–727, 2011.

18. Hammad, M. A. H., D. M. Sharaf, A. M. El-Shafey, and M. M. Nasr, "Peripheral nerve involvement in Behcet's disease; an electrophysiological study," *The Egyptian Rheumatologist*, Vol. 36, no. 4, pp. 195–199, 2014.

19. Gallay, L., A. Hot, P. Petiot, F. Thivolet-Bejui, and et al., "Focal myositis: new insights on diagnosis and pathology," *Neurology*, Vol. 90, no. 12, pp. e1013–e1020, 2018.

20. Kalra, S., A. Silman, G. Akman-Demir, S. Bohlega, and et al., "Diagnosis and management of Neuro-Behcet's disease: international consensus recommendations," *Journal of Neurology*, Vol. 261, pp. 1662–1676, 2014.

21. Kchaou, M., K. Bahrini, M. Belghith, A. Cherif, and et al., "Il-6 increase in csf as additional diagnostic criteria for definite Neuro-Behcet's disease," *IL-6 Increase in CSF as Additional Diagnostic Criteria for Definite Neuro-Behcet's Disease*, 2022.

22. Zhan, H., L. Cheng, and Y. Li, "Neuro-Behcet's disease: an update of clinical diagnosis, biomarkers, and immunopathogenesis," *Clinical and Experimental Immunology*, Vol. 219, no. 1, p. uxae123, 2025.

23. Akman-Demir, G., E. Tüzün, S. İçöz, N. Yeşilot, and et al., "Interleukin-6 in Neuro-Behcet's disease: association with disease subsets and long-term outcome," *Cytokine*, Vol. 44, no. 3, pp. 373–376, 2008.

24. Tüzün, E., and M. Kürtüncü, *Neuro-Behcet's Disease: Pathogenesis, Clinical Aspects, Treatment*, Springer Nature, 2021.

25. Tramontini, P. L., A. Finkelsztejn, J. A. Duarte, G. T. Santos, and et al., "Neuro-Behcet's disease mimicking brain tumor: a case report," *Surgical Neurology International*, Vol. 8, 2017.

26. Zhan, H., L. Cheng, Y. Liu, H. Xu, and et al., "Significance of immunoglobulins synthesis with central nervous system involvement in Neuro-Behcet's disease," *Clinica Chimica Acta*, Vol. 559, p. 119681, 2024.

27. Le Bihan, D., "Diffusion, confusion and functional MRI," *NeuroImage*, Vol. 62, no. 2, pp. 1131–1136, 2012.

28. Mohammadi, S., H. E. Möller, H. Kugel, D. K. Müller, *et al.*, "Correcting eddy current and motion effects by affine whole-brain registrations: Evaluation of three-dimensional distortions and comparison with slicewise correction," *Magnetic Resonance in Medicine*, Vol. 64, no. 4, pp. 1047–1056, 2010.

29. Taylor, D. G., and M. C. Bushell, "The spatial mapping of translational diffusion coefficients by the NMR imaging technique," *Physics in Medicine and Biology*, Vol. 30, no. 4, p. 345, 1985.

30. Chenevert, T. L., J. A. Brunberg, and J. G. Pipe, "Anisotropic diffusion in human white matter: demonstration with MR techniques in vivo.," *Radiology*, Vol. 177, no. 2, pp. 401–405, 1990.

31. Moseley, M. E., Y. Cohen, J. Mintorovitch, L. Chileuitt, and et al., "Early detection of regional cerebral ischemia in cats: comparison of diffusion-and T2-weighted MRI and spectroscopy," *Magnetic Resonance in Medicine*, Vol. 14, no. 2, pp. 330–346, 1990.

32. Douek, P., R. Turner, J. Pekar, N. Patronas, and et al., "MR color mapping of myelin fiber orientation," *J Comput Assist Tomogr*, Vol. 15, no. 6, pp. 923–929, 1991.

33. Beaulieu, C., "The basis of anisotropic water diffusion in the nervous system–a technical review," *NMR in Biomedicine: An International Journal Devoted to the Development and Application of Magnetic Resonance In Vivo*, Vol. 15, no. 7-8, pp. 435–455, 2002.

34. Pierpaoli, C., P. Jezzard, P. J. Basser, A. Barnett, and et al., "Diffusion tensor MR imaging of the human brain.," *Radiology*, Vol. 201, no. 3, pp. 637–648, 1996.

35. Song, S.-K., S.-W. Sun, M. J. Ramsbottom, C. Chang, and et al., "Dysmyelination revealed through MRI as increased radial (but unchanged axial) diffusion of water," *NeuroImage*, Vol. 17, no. 3, pp. 1429–1436, 2002.

36. Hagmann, P., L. Jonasson, P. Maeder, J.-P. Thiran, and et al., "Understanding diffusion MR imaging techniques: from scalar diffusion-weighted imaging to diffusion tensor imaging and beyond," *Radiographics*, Vol. 26, no. suppl_1, pp. S205–S223, 2006.

37. Basser, P. J., J. Mattiello, and D. LeBihan, "Estimation of the effective self-diffusion tensor from the NMR spin echo," *Journal of Magnetic Resonance, Series B*, Vol. 103, no. 3, pp. 247–254, 1994.

38. Le Bihan, D., J.-F. Mangin, C. Poupon, C. A. Clark, and et al., "Diffusion tensor imaging: concepts and applications," *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, Vol. 13, no. 4, pp. 534–546, 2001.

39. Moseley, M., R. Bammer, and J. Illes, "Diffusion-tensor imaging of cognitive performance," *Brain and Cognition*, Vol. 50, no. 3, pp. 396–403, 2002.

40. Sundgren, P. C., Q. Dong, D. Gomez-Hassan, S. K. Mukherji, and et al., "Diffusion tensor imaging of the brain: review of clinical applications," *Neuroradiology*, Vol. 46, pp. 339–350, 2004.

41. Mori, S., and J. Zhang, "Principles of diffusion tensor imaging and its applications to basic neuroscience research," *Neuron*, Vol. 51, no. 5, pp. 527–539, 2006.

42. Vilanova, A., S. Zhang, G. Kindlmann, and D. Laidlaw, "An introduction to visualization of diffusion tensor imaging and its applications," *Visualization and Processing of Tensor Fields*, pp. 121–153, 2006.

43. Nucifora, P. G. P., R. Verma, S.-K. Lee, and E. R. Melhem, "Diffusion-tensor MR imaging and tractography: exploring brain microstructure and connectivity," *Radiology*, Vol. 245, no. 2, pp. 367–384, 2007.

44. Voineskos, A. N., T. K. Rajji, N. J. Lobaugh, D. Miranda, and et al., "Age-related decline in white matter tract integrity and cognitive performance: a DTI tractography and structural equation modeling study," *Neurobiology of Aging*, Vol. 33, no. 1, pp. 21–34, 2012.

45. Park, S. H., and K. Han, "Methodologic guide for evaluating clinical performance and effect of artificial intelligence technology for medical diagnosis and prediction," *Radiology*, Vol. 286, no. 3, pp. 800–809, 2018.

46. Li, M., Z. Shang, Z. Yang, Y. Zhang, and et al., "Machine learning methods for MRI biomarkers analysis of pediatric posterior fossa tumors," *Biocybernetics and Biomedical Engineering*, Vol. 39, no. 3, pp. 765–774, 2019.

47. Aerts, H. J. W. L., E. R. Velazquez, R. T. H. Leijenaar, C. Parmar, and et al., "Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach," *Nature Communications*, Vol. 5, no. 1, p. 4006, 2014.

48. Orphanidou-Vlachou, E., N. Vlachos, N. P. Davies, T. N. Arvanitis, and et al., "Texture analysis of T1-and T2-weighted MR images and use of probabilistic neural network to discriminate posterior fossa tumours in children," *NMR in Biomedicine*, Vol. 27, no. 6, pp. 632–639, 2014.

49. Sabin, N. D., T. E. Merchant, X. Li, Y. Li, and et al., "Quantitative imaging analysis of posterior fossa ependymoma location in children," *Child's Nervous System*, Vol. 32, pp. 1441–1447, 2016.

50. Gevaert, O., L. A. Mitchell, A. S. Achrol, J. Xu, and et al., "Glioblastoma multiforme: exploratory radiogenomic analysis by using quantitative image features," *Radiology*, Vol. 273, no. 1, pp. 168–174, 2014.

51. Yousef, H., B. Malagurski Tortei, and F. Castiglione, "Predicting multiple sclerosis disease progression and outcomes with machine learning and MRI-based biomarkers: a review," *Journal of Neurology*, Vol. 271, no. 10, pp. 6543–6572, 2024.

52. Koçer, N., C. Islak, A. Siva, S. Saip, and et al., "CNS involvement in Neuro-Behcet syndrome: an MR study," *American Journal of Neuroradiology*, Vol. 20, no. 6, pp. 1015–1024, 1999.

53. Anik, Y., Z. Kural, A. Demirci, G. Akansel, and et al., "Magnetization transfer ratio in Neuro-Behcet disease," *Neuroradiology*, Vol. 47, pp. 108–113, 2005.

54. Kunimatsu, A., O. Abe, S. Aoki, N. Hayashi, and et al., "Neuro-Behcet's disease: analysis of apparent diffusion coefficients," *Neuroradiology*, Vol. 45, pp. 524–527, 2003.

55. Baysal, T., M. Dogan, R. Karlidag, H. I. Ozisik, and et al., "Diffusion-weighted imaging in chronic Behcet patients with and without neurological findings," *Neuroradiology*, Vol. 47, pp. 431–437, 2005.

56. Yildiz, S., N. Boyaci, D. S. Dokumaci, and A. Boyaci, "Different diffusion-weighted MRI findings in brainstem Neuro-Behcet's disease," *Case Reports*, Vol. 2013, p. bcr2013200738, 2013.

57. Hardy, T., "New developments in Susac's syndrome," *ACNR, Advances in Clinical Neuroscience and Rehabilitation*, 2013.

58. Alis, D., C. Alis, M. Tutuncu, N. Kocer, and et al., "Apparent diffusion coefficient characteristics of parenchymal Neuro-Behcet's disease," *International Journal of Rheumatic Diseases*, Vol. 22, no. 8, pp. 1452–1458, 2019.

59. Atasoy, B., H. Toprak, O. Su Kucuk, R. Selvitop, and et al., "Relationship of diffusion tensor imaging parameters with neurocognitive dysfunction in patients with Behcet's disease," *Acta Neurologica Belgica*, pp. 1–10, 2022.

60. Aslan, K., B. Genç, N. Bolat, and L. Incesu, "Diffusion tensor imaging in Behcet's disease with and without neurological involvement patients: evaluation of microstructural white matter abnormality with a tract-based spatial statistical analysis," *British Journal of Radiology*, Vol. 97, no. 1162, pp. 1645–1652, 2024.

61. Gündüz, T., S. Server, C. I. Küçükali, O. Özyurt, and et al., "Diffusion tensor imaging in parenchymal Neuro-Behcet's disease," *Archives of Neuropsychiatry*, Vol. 61, no. 1, p. 39, 2024.

62. Uygunoglu, U., N. Erener, A. Kargin, S. Arslan, and et al., "Brain white matter microstructural alterations in Behcet's syndrome correlate with cognitive impairment and disease severity: A diffusion tensor imaging study (p3-1.002)," in *Neurology*, Vol. 104(7), p. 4535, Lippincott Williams & Wilkins Hagerstown, MD, 2025.

63. Kim, J. M., J. G. Kang, S. Kim, and J. H. Cheon, "Deep-learning system for real-time differentiation between crohn's disease, intestinal Behcet's disease, and intestinal tuberculosis," *Journal of Gastroenterology and Hepatology*, Vol. 36, no. 8, pp. 2141–2148, 2021.

64. Hammam, N., A. Bakhiet, E. A. El-Latif, I. I. El-Gazzar, and et al., "Development of machine learning models for detection of vision threatening Behcet's disease BD using egyptian college of rheumatology (ecr)–bd cohort," *BMC Medical Informatics and Decision Making*, Vol. 23, no. 1, p. 37, 2023.

65. Karadeli, D., "Behcet üveiti hastalığının tespitinde farklı makine öğrenme algoritmaları ve yeniden örnekleme tekniklerinin incelenmesi," *EMO Bilimsel Dergi*, Vol. 15, no. 1, pp. 65–76, 2025.

66. Tustison, N. J., and S. G. Avants, "Bb advanced normalization tools: V1. 0," *Insight Journal*, Vol. 2, 2009.

67. Glantz, S. A., *Primer of Biostatistics*, The McGraw-Hill Medical, 1981.

68. Joachims, T., "Text categorization with support vector machines: Learning with many relevant features," in *European conference on machine learning*, pp. 137–142, Springer, 1998.

69. DeCoste, D., and B. Schölkopf, "Training invariant support vector machines," *Machine Learning*, Vol. 46, pp. 161–190, 2002.

70. Brown, M. P. S., W. N. Grundy, D. Lin, N. Cristianini, and et al., "Knowledge-based analysis of microarray gene expression data by using support vector machines," *Proceedings of the National Academy of Sciences*, Vol. 97, no. 1, pp. 262–267, 2000.

71. Lanckriet, G. R. G., T. De Bie, N. Cristianini, M. I. Jordan, and et al., "A statistical framework for genomic data fusion," *Bioinformatics*, Vol. 20, no. 16, pp. 2626–2635, 2004.

72. Bradford, J. R., and D. R. Westhead, "Improved prediction of protein–protein binding sites using a support vector machines approach," *Bioinformatics*, Vol. 21, no. 8, pp. 1487–1494, 2005.

73. Zhang, S.-W., Q. Pan, H.-C. Zhang, Y.-L. Zhang, and et al., "Classification of protein quaternary structure with support vector machine," *Bioinformatics*, Vol. 19, no. 18, pp. 2390–2396, 2003.

74. MathWorks, "Support vector machine SVM," 2024. Accessed: 2025-05-10.

75. Taunk, K., S. De, S. Verma, and A. Swetapadma, "A brief review of nearest neighbor algorithm for learning and classification," in *2019 international conference on intelligent computing and control systems (ICCS)*, pp. 1255–1260, IEEE, 2019.

76. Peretz, O., M. Koren, and O. Koren, "Naive bayes classifier–an ensemble procedure for recall and precision enrichment," *Engineering Applications of Artificial Intelligence*, Vol. 136, p. 108972, 2024.

77. Zhou, Z.-H., *Ensemble methods: foundations and algorithms*, CRC press, 2025.

78. Johnson, J. E., V. Laparra, A. Pérez-Suay, M. D. Mahecha, and et al., "Kernel methods and their derivatives: Concept and perspectives for the earth system sciences," *PLOS One*, Vol. 15, no. 10, p. e0235885, 2020.

79. White, L., R. Togneri, W. Liu, and M. Bennamoun, "Introduction to neural networks for machine learning," *Neural Representations of Natural Language*, pp. 1–21, 2019.

80. Schober, P., and T. R. Vetter, "Logistic regression in medical research," *Anesthesia & Analgesia*, Vol. 132, no. 2, pp. 365–366, 2021.