

Glio-SERS: Artificial Intelligence and Surface Enhanced Raman Spectroscopy Driven Liquid Biopsy Method for Brain Tumor Classification

by

Hulya Torun

A Dissertation Submitted to the
Graduate School of Sciences and Engineering
in Partial Fulfillment of the Requirements for
the Degree of

Doctor of Philosophy

in

Biomedical Sciences and Engineering



KOÇ ÜNİVERSİTESİ

June 25, 2024

**Glio-SERS: Artificial Intelligence and Surface Enhanced Raman
Spectroscopy Driven Liquid Biopsy Method for Brain Tumor
Classification**

Koç University

Graduate School of Sciences and Engineering

This is to certify that I have examined this copy of a doctoral dissertation by

Hulya Torun

and have found that it is complete and satisfactory in all respects,
and that any and all revisions required by the final
examining committee have been made.

Committee Members:

Prof. Ihsan Solaroglu (Advisor)

Prof. Utkan Demirci (Co-Advisor)

Prof. Tugba Bagci-Onder

Assoc. Prof. Atay Vural

Prof. M. Burcin Unlu

Prof. Nesrin Ozoren

Date: _____

This thesis is dedicated to the memory of my beloved grandparents, who always dreamed of seeing me become Dr. Torun. Although their lifetimes did not allow them to witness this moment, their love, faith, and encouragement continue to inspire me every day. I also dedicate this work to my parents, whose boundless support, sacrifices, and nurturing have made me who I am today. Their faith in me has been my greatest strength and motivation. I extend this dedication to all the women and scientists who tirelessly work to make the world a better place. Your dedication and passion are a beacon of hope and progress for future generations. Lastly, I dedicate this thesis to the patients and their families who are seeking diagnosis and treatment. Your resilience and courage are the driving force behind my research and the pursuit of better medical solutions.

ABSTRACT

Glio-SERS: Artificial Intelligence and Surface Enhanced Raman Spectroscopy Driven Liquid Biopsy Method for Brain Tumor

Classification

Hulya Torun

Doctor of Philosophy in Biomedical Sciences and Engineering

June 25, 2024

Glioblastoma (GB), the most aggressive adult brain tumor, requires invasive procedures such as biopsy or surgical intervention, along with sophisticated laboratory settings and prolonged, costly molecular testing for accurate diagnosis. Nearly all patients with GB experience tumor regrowth within two years after primary surgery. The current management of gliomas relies on diagnostic imaging, which lacks the high sensitivity and specificity needed to evaluate recurrence after primary treatment. Implementing liquid biopsy in gliomas is essential for early diagnosis, detecting residual disease after surgery, and assessing disease status post-treatment. With an average survival rate of just 14 months, there is an urgent need for rapid, accurate, cost-effective, and minimally invasive diagnostic strategies. This study introduces a transformative diagnostic paradigm—a liquid biopsy approach that merges Surface-Enhanced Raman Spectroscopy (SERS) analysis of exosomes with artificial intelligence (AI).

In a prospective study, we evaluated the efficacy of this SERS and AI-based liquid biopsy analysis for GB detection. We collected 20 glioblastoma, 23 meningioma (MNG) as the most frequent benign tumor control, and 30 healthy control (HC) plasma samples under informed consent and IRB approval. SERS was employed to analyze the isolated plasma exosomes, generating spectral data on molecular signatures indicative of each condition. Deep learning algorithms were integrated into the analysis pipeline to facilitate rapid and accurate differentiation between GB, MNG, and HC plasma exosome SERS signatures.

Our approach achieved a remarkable 87% prediction accuracy in distinguishing GB exosomal signatures from those of MNG and HC individuals. This result signifies

a substantial advancement in the precision and speed of GB diagnostics compared to traditional methods.

This pioneering liquid biopsy technique emerges as a front-line solution for GB detection and monitoring. The ongoing integration of an expansive library of tumor signatures signifies a significant leap forward in analytical technologies tailored for neuro-oncology. Our study demonstrates not only the technological superiority of our approach but also its potential to revolutionize GB detection and improve patient care once it is clinically validated with a larger cohort.



ÖZETÇE

**Glio-SERS: Beyin Tümörlerinin Sınıflandırılması için Yapay Zeka ve
Yüzey Güçlendirilmiş Raman Spektroskopisi Tabanlı Sıvı Biyopsi
Yöntemi
Hulya Torun
Biyomedikal Bilimler ve Mühendislik, Doktora
25 Haziran 2024**

Glioblastoma (GB), en agresif yetişkin beyin tümörü olarak, gelişmiş laboratuvar ortamları ve uzman personel gerektiren uzun ve maliyetli moleküler ve histopatoloji değerlendirmeleri nedeniyle önemli tanısal zorluklar sunmaktadır. Kesin tanı genellikle doku örneklerinin moleküler ve histolojik analizi için biyopsi ve cerrahi müdahaleler gibi invaziv prosedürleri gerektirir. Mevcut tanı ve tedavi yaklaşımları, 14 aylık bir ortalama yaşam süresi sunarak, daha etkin ve düşük maliyetli yöntemlere olan acil ihtiyacı vurgulamaktadır.

Bu çalışma, GB için Yüzey Artırılmış Raman Spektroskopisi (SERS) ve yapay zeka (YZ) destekli sıvı biyopsi kullanarak yeni bir tanı yaklaşımı sunmaktadır. Plazma eksozomları, sağlıklı olanlardan beyin tümörlerini tespit edip ayırt edebilen önemli ayırt edici imzalar taşır. SERS ile, GB-spesifik eksozomal imzaları tanımladık ve YZ ile hızlı ve doğruluğu yüksek analizler gerçekleştirdik, %87 tanı doğruluğu elde ettik. SERS, YZ ile birleştiğinde kan eksozomlarından beyin tümörlerinin tespiti için kullanılabilir. Metodolojimiz, GB imzalarını MNG ve sağlıklı plazma eksozom imzalarından başarıyla ayırarak, bu sıvı biyopsi yaklaşımı ile nöro-onkoloji tanılarını ilerletme potansiyeline sahiptir. Genişletilmiş bir tümör imza kütüphanesinin entegrasyonu, yöntemin potansiyel etkisini önemli ölçüde artırarak nöro-onkoloji için analitik teknolojilerde büyük bir ilerlemeyi işaret etmektedir. YZ sınıflandırması, eksozom SERS spektrumlarının beyin tümörlerinin eksozom kaynaklı moleküler imzalarını ayırt etmek için umut verici bir analiz tekniğidir. Bu yaklaşım, beyin tümörü tespiti ve sınıflandırması için hızlı, düşük maliyet ve yüksek doğrulukta bir yöntem sağlayarak doktorlara teşhis ve takip konusunda yardımcı olma vaadini taşımakta ve böylece beyin tümörü tanılarında yeni bir dönemi müjdelemektedir.

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my MSc and Ph.D. advisor, Prof. İhsan Solaroğlu, for his unwavering support, trust, and encouragement throughout my academic journey. I could not have imagined a better advisor, and I feel incredibly fortunate to have had the opportunity to work with such an outstanding scientist, physician, and mentor.

I am deeply grateful to my Co-Advisor Prof. Utkan Demirci, for all his support and guidance for my PhD research. I am incredibly thankful for the opportunity to carry out my PhD research at Stanford for two years, made possible through his support. Working at such a renowned and successful university, with esteemed collaborators, has given me unique insights and significantly impacted my career, helping me to become a better scientist.

I would also like to thank my mentor and colleague, Ugur Parlattan, for his invaluable guidance, support, and collaboration throughout my research. His insights and expertise have been instrumental in shaping my work and career, and I am truly grateful for his mentorship.

My heartfelt thanks go to both Koc University and Stanford University for providing funding and infrastructure for my research and supporting me in numerous ways. I am also grateful to the Koc University Research Center for Translational Medicine and the Canary Center at Stanford for Cancer Early Detection for their infrastructure and support, which were crucial for my research.

Tim Valencony is greatly acknowledged for his efforts in SERS spectra pre-processing and AI-classification, upon continuous discussions and under the guidance of Dr. Ugur Parlattan.

I am thankful to Dr. Abel Bermudez, Dr. Fernando Jose Garcia Marques,

and Sharon Pitteri for their collaboration and guidance in experimental and data analysis for proteomics analysis of exosomes. I appreciate the support of Dr. Ozgur Albayrak for his endless support in flow cytometry analyses. I would also like to extend my gratitude to Dr. Bharti Singal for her immense support in cryo-EM imaging of exosomes.

I am deeply appreciative of Dr. Goktug Albayrak, Dr. Oguz Baran, and Furkan Kaysin for their dedication and assistance in sample collection, regardless of the day and time. I am also very thankful to Merve Cigdem Atsu and Rana Cansu Egitimci for their eternal support throughout my graduate studies.

I would also like to acknowledge the assistance of ChatGPT, provided by OpenAI, for its help in refining the language and structure of my thesis. The guidance and support received through this platform have contributed to the clarity and expressiveness of my writing.

TABLE OF CONTENTS

List of Tables	xv
List of Figures	xvi
Abbreviations	xxviii
Chapter 1: Introduction	1
1.1 Glioblastoma (GB)	1
1.2 Assessment of Glioma Progression and Pseudoprogession	3
1.3 Motivation for Liquid Biopsy in Gliomas	5
1.4 Importance of the Study	7
1.5 Hypothesis and Specific Aims	9
1.5.1 Hypothesis	9
1.5.2 Specific Aims	9
1.6 Scope of the Thesis	12
Chapter 2: Literature Review	14
2.1 Glioblastoma Diagnostics	14
2.1.1 Histopathological Features	14
2.1.2 Molecular Characteristics	14
2.1.3 Diagnostic Imaging and Techniques	15
2.1.4 Prognosis and Treatment Challenges	16
2.1.5 Future Directions	17
2.2 Liquid Biopsy	17
2.2.1 Introduction to Liquid Biopsy Techniques and Their Advantages	17
2.2.2 Challenges in Liquid Biopsy Techniques	18

2.2.3	Importance and Potential of Liquid Biopsy in Brain Tumors	18
2.2.4	Current Limitations and Future Directions	19
2.3	Exosomes in Cancer Diagnostics	20
2.3.1	Role of Exosomes in Cancer Detection	20
2.3.2	Exosomes in Brain-Related Diseases	22
2.4	Surface-Enhanced Raman Spectroscopy (SERS)	25
2.4.1	Principles of SERS and Its Applications in Biomedical Research	26
2.5	Artificial Intelligence in Diagnostics	28
2.5.1	Overview of AI and Deep Learning Algorithms in Medical Diagnostics	28
2.6	Emerging Techniques for Liquid Biopsy-Based Detection of Glioblastoma	29
2.6.1	Optical Sensing	29
2.6.2	Electrochemical Sensing	29
2.6.3	Other Sensing Techniques	30
2.6.4	Fluorescence-Based Techniques	31
2.7	Conclusion	33
Chapter 3:	Materials and Methods	35
3.1	Sample Collection	35
3.1.1	Sample Size Calculation	35
3.1.2	Pathological Evaluation and Diagnosis	36
3.1.3	IRB Approval	36
3.1.4	Material Transfer Agreement (MTA)	36
3.1.5	Study Design	36
3.2	Exosome Isolation from Plasma Samples	36
3.2.1	Plasma Processing	36
3.2.2	Exosome Isolation	37
3.3	Exosome Characterization	38

3.3.1	TEM Analysis of Exosomes	38
3.3.2	Cryo-TEM Analysis of Exosomes	38
3.3.3	NTA Analysis of Exosomes	38
3.3.4	Western Blot Analysis of Exosomes	38
3.3.5	iSCAT Analysis of Exosomes	39
3.3.6	Flow Cytometry Analysis of Exosomes	39
3.4	Surface-Enhanced Raman Spectroscopy Measurements	39
3.4.1	Experimental Setup	39
3.4.2	Sample Preparation	40
3.4.3	Data Collection	40
3.4.4	Data Analysis	41
3.5	Statistical Analysis	41
3.6	Limit of Detection (LOD) and Limit of Quantification (LOQ) Assessment for SERS Biosensor	41
3.7	Regression Analysis	42
3.8	Proteomics Analysis	42
3.8.1	Exosome Isolation and Protein Extraction	42
3.8.2	LC/MS analysis	44
Chapter 4:	Exosome Isolation and Characterization	45
4.1	Nanoparticle Tracking Analysis (NTA)	45
4.2	Room Temperature Transmission Electron Microscopy (TEM)	45
4.3	Cryo-Transmission Electron Microscopy (Cryo-TEM)	46
4.4	Interferometric Scattering Microscopy (iSCAT)	48
4.5	Bead-Assisted Flow Cytometry	49
4.6	Western Blotting	50
Chapter 5:	Proteomics Analysis of Exosomes	61
5.1	Exosome Isolation and Protein Extraction	61
5.2	Protein Sample Preparation and Mass Spectrometry	61

5.3	Protein Library Search	61
5.4	Data Analysis and Interpretation	62
5.4.1	Proteomic Datasets and Differential Expression Analysis . . .	62
5.4.2	Dimensionality Reduction and Clustering	62
5.4.3	Validation of Tumor-Related Proteins	62
5.5	Statistical Analysis and Power Calculation	62
5.6	Proteomics Data Analysis	63
5.6.1	Heatmap Analysis	63
5.7	Interpretation of Results	65
5.7.1	Gene Enrichment Analysis	66
5.7.2	Comprehensive Cluster Analysis of Proteins	72
5.8	Proteins Identified in the Proteomics Analysis	74
Chapter 6: Surface Enhanced Raman Spectroscopy Analysis of Plasma Exosomes		104
6.1	Motivation for Raman Spectroscopy in Gliomas	104
6.2	Spectra Pre-Processing	105
6.2.1	Raw Spectra Spectra Cropping	105
6.2.2	Cosmic Ray Removal	105
6.2.3	Spectra Smoothing-Denoising	106
6.2.4	Baseline Correction	107
6.2.5	Spectra Normalization	107
6.3	Spectral Analysis	107
6.3.1	Distribution Analysis using t-SNE	107
6.3.2	Visualization of Spectral Data	107
6.3.3	Average Spectra of Plasma Exosomes	108
6.3.4	Methodology	109
6.4	Regression Analysis	111
6.4.1	Feature Selection	112

6.4.2	Principal Component Analysis (PCA)	113
6.4.3	Recursive Feature Elimination (RFE)	113
6.4.4	Linear Regression	113
6.4.5	Lasso Regression	115
6.4.6	Ridge Regression	115
6.5	Proteomics and SERS Correlation	116
 Chapter 7: Classification of Brain Tumor Exosome SERS Spectra		
	Using Artificial Intelligence Models	118
7.1	Prediction Models	118
7.1.1	Quadratic Discriminant Analysis (QDA)	118
7.1.2	Random Forest (RF) Classifier	118
7.1.3	XG Boost (XGB)	123
7.1.4	Support Vector Machine (SVM)	130
7.1.5	Neural Networks (NN)	131
7.1.6	Convolutional Neural Networks (CNN)	131
7.1.7	Downsides of Neural Networks (NNs) and Convolutional Neu- ral Networks (CNNs)	138
7.2	Model Training and Evaluation	144
7.3	Training Set Composition	144
7.4	Testing and Performance Evaluation	146
7.5	Confusion Matrix Analysis	146
7.6	Conclusion and Future Work	148
 Chapter 8: Conclusion		
8.1	Results and Discussion	149
8.1.1	Interpretation of Results	149
8.1.2	Implications for Glioblastoma Diagnostics	151
8.1.3	Strengths and Limitations	151
8.1.4	Summary of Key Findings	152

8.1.5	Impact on Patient Care	152
8.1.6	Final Remarks	152
8.2	Discussion	152
8.3	Future Directions and Perspectives	154
Bibliography		155



LIST OF TABLES

5.1	Functions and Pathologies for Proteins with Increasing Trend in GB .	92
5.2	Raman/SERS Peaks for Proteins with Increasing Trend in GB	95
5.3	Functions and Pathologies for Proteins with Decreasing Trend in GB	96
5.4	Raman/SERS Peaks for Proteins with Decreasing Trend in GB	99
5.5	Functions and Pathologies for Proteins with Increasing Trend in Menin- gioma	100
5.6	Raman/SERS Peaks for Proteins with Increasing Trend in Meningioma	101
5.7	Functions and Pathologies for Proteins with Decreasing Trend in Meningioma	101
5.8	Raman/SERS Peaks for Proteins with Decreasing Trend in Meningioma	102
7.1	Confusion Matrix for CNN Patient-wise Classification	146
7.2	Sensitivity and Specificity for Glioblastoma, Meningioma, and Healthy Control	148

LIST OF FIGURES

1.1	Example of biopsy of a brain lesion in a 71-year-old female patient. Initial gadolinium-enhanced coronal (A) T1-weighted (B) T2-weighted (C) FLAIR MRI prior to any treatment showing hyperintense area with indistinct boundaries at the right frontal horn (white arrow). (D) In post-contrast DSC perfusion imaging, no increased perfusion is detected within the described infiltrative signal anomalies. The choline/NAA ratio is less than 1 in all voxels. Early postoperative (E) coronal and (F) axial CT scan after stereotactic biopsy showing the trajectory of the needle and sampling areas (green arrows). Histopathological examination showed atypical glial cells. (G) At 3 months, a follow-up postcontrast coronal T1-weighted MRI showed the development of the tumor (red arrow). She proceeded to further surgery; the histopathology was consistent with glioblastoma. From the archive of Prof. Ihsan Solaroglu, MD.	3
1.2	Figure. Key MRI images of a 7-year-old patient with a diffusely infiltrating pontine tumor. (A) axial T2-weighted MRI at initial diagnosis shows the abnormal signal intensity of pons and compression of the 4th ventricle (white arrow), (B) 3D color-coded diffusion tensor imaging showing the relationship between corticospinal tracts and the tumor (white arrow) (C) postcontrast axial T1-weighted MRI following radiotherapy treatment showing necrosis (red arrow) with (D) reduced cerebral blood volume (red arrow). From the archive of Prof. Ihsan Solaroglu, MD.	6

1.3	Schematic illustration of Specific Aim 1a. Sample collection from glioblastoma (GBM), meningioma (MNG), and healthy control (HC) participants, plasma separation, exosome isolation from these clinical samples. Created with BioRender.com	10
1.4	Schematic illustration of Specific Aim 1b: Characterization of exosomes isolated from glioblastoma (GBM), meningioma (MNG), and healthy control (HC) samples via Room Temperature Transmission Electron Microscopy (RT-TEM), Cryo-TEM, Interferometric Scattering Microscopy (iSCAT), Nanoparticle Tracking Analysis (NTA), Bead-Assisted Flow Cytometry (BA-FC), Western Blotting (WB) techniques. Created with BioRender.com	10
1.5	Schematic Illustration of Specific Aim 1c. Surface-Enhanced Raman Spectroscopy (SERS) measurements of exosomes derived from GB, MNG, and HC samples and spectral library construction. Created with BioRender.com	11
2.1	Clinical workflow for histopathological evaluation of the brain tumors. Adopted from [Mohamed et al., 2024])	15
2.2	Contrast-enhanced T1-weighted MRI images of a patient (A) with a left parietal lobe GB. (B) Early post-op image showing complete GB removal. (C) Radiation necrosis developed around surgical cavity, 6 months post radiation therapy (From the archive of Prof. Ihsan Solaroglu, MD)	16

2.3	Illustration of extracellular vesicle structure, formation, and content. Exosomes, ranging from 30 to 200 nm in diameter, carry proteomic and genetic information from their cell of origin. They contain a diverse array of biomolecules, including transmembrane proteins, tetraspanins, growth factors, signaling receptors, lipid rafts, cholesterol, cytosolic proteins, sphingomyelins, adhesion molecules, heat shock proteins, and various cargo molecules such as RNA and DNA [Théry et al., 2002, Simons and Raposo, 2011, Kowal et al., 2014]. Exosome formation involves the endosomal pathway, where early endosomes mature into late endosomes or multivesicular bodies (MVBs). The intraluminal vesicles (ILVs) within these MVBs are released into the extracellular space as exosomes upon fusion with the plasma membrane [Raposo and Stoorvogel, 2013, van Niel et al., 2018]. In contrast, microvesicles, another class of extracellular vesicles, form directly through the outward budding and fission of the plasma membrane [Cocucci et al., 2009, Yáñez-Mó et al., 2015].	21
3.1	Schematic illustration of study design: Sample collection, plasma separation and exosome isolation, exosome characterization, Surface-Enhanced Raman Spectra (SERS) analysis and library construction, training and testing SERS data using Artificial Intelligence (AI) models, first step AI classification for brain tumor (BT) and brain tumor free (BTF) individuals, second step AI classification for predicting brain tumor type (GB or MNG). Created with openAI and BioRender.com	37
3.2	Limit of detection (LOD) and limit of quantification (LOQ) of SERS sensor used for exosome and protein (EGFR) detection analysis . . .	43
4.1	Nanosight Tracking analysis (NTA) of plasma exosomes derived from glioblastoma (GB) patients.	46

4.2	Nanosight Tracking analysis (NTA) of plasma exosomes derived from meningioma (MNG) patients.	47
4.3	Nanosight Tracking analysis (NTA) of plasma exosomes derived from healthy control (HC) patients.	48
4.4	Transmission Electron Microscopy (TEM) imaging of plasma exosomes derived from glioblastoma (GB) patients.Exosomes were observed in between 30-200 nm in size.	51
4.5	Transmission Electron Microscopy (TEM) imaging of plasma exosomes derived from meningioma (MNG) patients.Exosomes were observed in between 30-200 nm in size.	52
4.6	Transmission Electron Microscopy (TEM) imaging of plasma exosomes derived from healthy control (HC) patients.Exosomes were observed in between 30-200 nm in size.	53
4.7	Cryo-Electron Microscopy (Cryo-EM) images of plasma exosomes derived from glioblastoma patients-1. Single, multi-laminar, and spherical vesicles.	54
4.8	Cryo-Electron Microscopy (Cryo-EM) images of plasma exosomes derived from meningioma patients-1. Single, multi-laminar, spherical and cylindrical vesicles.	55
4.9	Cryo-Electron Microscopy (Cryo-EM) images of plasma exosomes derived from healthy control participants. Single, multi-laminar, spherical and cylindrical vesicles.	56
4.10	Interferometric Scattering Microscopy (iSCAT) images of glioblastoma, meningioma, and healthy control plasma exosome samples and their size distributions.	57
4.11	Histograms of glioblastoma patient plasma exosome characterization via Bead-Assisted Flow Cytometry (BA-FC). Exosomes are validated using Calnexin, CD81, and HSPA8 antibodies.	58

4.12	Glioblastoma patient plasma xosome characterization via Bead-Assisted Flow Cytometry (BA-FC) showing the control and experimental groups used. Exosomes are validated using Calnexin, CD81, and HSPA8 antibodies. The first figure shows all particles and the Calnexin-negative (Cellular content contamination-free) population using APC (Red-Red) Secondary antibody. In the second figure, 12 analysis performed: Only PBS, only exosomes without bead (E), only beads (B), only plasma (P), exosome (E) + bead (B) + Primary Antibody (Pab), Exosome + Bead + Secondary Antibody (E+B+Sab), E+B+CD81+Sab (Alexa 488, Green-Blue), E+B+HSPA8+Sab (PE, Yellow-Blue), E+B+CD81+HSPA8+Sab(Alexa 488 and PE), P+B+CD81+HSPA8+Sab(Alexa 488 and PE), and Exosome standards (STND: Novus Biologicals, NBP3-11686) +B+CD81+HSPA8+Sab(Alexa 488 and PE) groups were analyzed.	59
4.13	Western blot analysis of glioblastoma, meningioma, and healthy control exosomes using Calnexin, HSPA8, CD63, TSG101, and Flot1 antibodies.	60
5.1	Heatmap Analysis of proteomics data acquired from glioblastoma (GB), meningioma (MNG), and healthy control (HC) exosomes ($n_{GB} = 17, n_{MNG} = 20, n_{HC} = 30$)	64
5.2	Gene enrichment analysis of glioblastoma (GB) exosome proteomics. ($n_{GB} = 17$)	67
5.3	Gene enrichment analysis of meningioma (MNG) exosome proteomics. ($n_{MNG} = 20$)	68
5.4	Gene enrichment analysis of healthy control (HC) exosome proteomics. ($n_{HC} = 30$)	69

5.5	Comparison of gene enrichment analysis for three groups: Glioblastoma (GB), meningioma (MNG), and healthy control (HC) exosome proteomics. ($n_{GB} = 17, n_{MNG} = 20, n_{HC} = 30$)	70
5.6	Proteomics Comparison for the Most Predictive 20 Proteins	103
6.1	Schematic illustration of pre-processing steps for Surface-Enhanced Raman Spectra analysis: Acquiring raw spectra from the plasma exosomes of all study participants, cropping the spectra to narrow it down to the fingerprint region, denoising the spectra using Savitzky-Golay filtering, baseline correction and vector normalization. Created with BioRender.com	106
6.2	t-SNE analysis of glioblastoma (GB) exosome SERS spectra ($n=20$). Each number refers to a different patient	108
6.3	t-SNE analysis of meningioma (MNG) exosome SERS spectra ($n=24$). Each number refers to a different patient	109
6.4	t-SNE analysis of healthy control (HC) exosome SERS spectra ($n=30$). Each number refers to a different study participant	110
6.5	t-SNE analysis of all groups including glioblastoma (GB), meningioma (MNG), healthy control (HC) exosome averaged SERS spectra ($n_{GB} = 20, n_{MNG} = 24, n_{HC} = 30$)	111
6.6	Average SERS spectra of glioblastoma (GB), meningioma (MNG), and healthy control (HC) exosomes ($n_{GB} = 20, n_{MNG} = 24, n_{HC} = 30$)	112
6.7	Violin plots of significant peaks differing in between three groups: glioblastoma (GB), meningioma (MNG), and healthy control (HC) exosomes ($n_{GB} = 20, n_{MNG} = 24, n_{HC} = 30$)	113
6.8	Different type of train-test splits: (a) Spectra-wise classification and (b) Patient-wise classification, which is more suitable for clinical setting	114

6.9	Principal Component Analysis (PCA) of glioblastoma (GB), meningioma (MNG), healthy control (HC) exosome averaged SERS spectra ($n_{GB} = 20, n_{MNG} = 24, n_{HC} = 30$)	115
6.10	Principal Component (PC) coefficients of glioblastoma (GB), meningioma (MNG), healthy control (HC) exosome averaged SERS spectra ($n_{GB} = 20, n_{MNG} = 24, n_{HC} = 30$)	116
6.11	Average Intensities with Error Bars for Glioblastoma (GB), Meningioma (MNG), and Healthy Control (HC) Exosome SERS Signatures	117
6.12	Log2Fold Change with Error Bars for Meningioma (MNG) and Healthy Control (HC) Exosome Raman Signatures Compared to Glioblastoma (GB)	117
7.1	Confusion matrix for the first step classification to identify brain tumor free (BTF) individuals and patients with brain tumor (BT) using quadratic discriminant analysis (QDA) for spectra-wise classification .	119
7.2	Confusion matrix for the second step classification to identify brain tumor free (BTF) individuals and patients with glioblastoma (GB), meningioma (MNG) using quadratic discriminant analysis (QDA) model for spectra-wise classification	119
7.3	ROC Curve for the classification to identify brain tumor-free (BTF) individuals and patients with glioblastoma (GB) or meningioma (MNG) using quadratic discriminant analysis (QDA) model for spectra-wise classification	120
7.4	Confusion matrix for the first step classification to identify brain tumor free (BTF) individuals and patients with brain tumor (BT) using quadratic discriminant analysis (QDA) for patient-wise classification .	121

7.5	Confusion matrix for the second step classification to identify brain tumor free (BTF) individuals and patients with glioblastoma (GB), meningioma (MNG) using quadratic discriminant analysis (QDA) model for patient-wise classification	121
7.6	ROC Curve for the classification to identify brain tumor-free (BTF) individuals and patients with glioblastoma (GB) or meningioma (MNG) using quadratic discriminant analysis (QDA) model for patient-wise classification	122
7.7	Confusion matrix for the first step classification to identify brain tumor free (BTF) individuals and patients with brain tumor (BT) using Random Forest (RF) for spectra-wise classification	123
7.8	Confusion matrix for the second step classification to identify brain tumor free (BTF) individuals and patients with glioblastoma (GB), meningioma (MNG) using Random Forest (RF) model for spectra-wise classification	124
7.9	ROC Curve for the classification to identify brain tumor-free (BTF) individuals and patients with glioblastoma (GB) or meningioma (MNG) using Random Forest (RF) model for spectra-wise classification . . .	124
7.10	Confusion matrix for the first step classification to identify brain tumor free (BTF) individuals and patients with brain tumor (BT) using Random Forest (RF) for patient-wise classification	125
7.11	Confusion matrix for the second step classification to identify brain tumor free (BTF) individuals and patients with glioblastoma (GB), meningioma (MNG) using Random Forest (RF) model for spectra-wise classification	125
7.12	ROC Curve for the classification to identify brain tumor-free (BTF) individuals and patients with glioblastoma (GB) or meningioma (MNG) using Random Forest (RF) model for patient-wise classification . . .	126

7.13	Confusion matrix for the first step classification to identify brain tumor free (BTF) individuals and patients with brain tumor (BT) using XG-Boost (XGB) model for spectra-wise classification	126
7.14	Confusion matrix for the second step classification to identify brain tumor free (BTF) individuals and patients with glioblastoma (GB), meningioma (MNG) using XG-Boost (XGB) model for spectra-wise classification	127
7.15	ROC Curve for the classification to identify brain tumor-free (BTF) individuals and patients with glioblastoma (GB) or meningioma (MNG) using XG-Boost (XGB) model for spectra-wise classification	127
7.16	Confusion matrix for the first step classification to identify brain tumor free (BTF) individuals and patients with brain tumor (BT) using XG-Boost (XGB) model for patient-wise classification	128
7.17	Confusion matrix for the second step classification to identify brain tumor free (BTF) individuals and patients with glioblastoma (GB), meningioma (MNG) using XG-Boost (XGB) model for spectra-wise classification	128
7.18	ROC Curve for the classification to identify brain tumor-free (BTF) individuals and patients with glioblastoma (GB) or meningioma (MNG) using XG-Boost (XGB) model for patient-wise classification	129
7.19	Confusion matrix for the first step classification to identify brain tumor free (BTF) individuals and patients with brain tumor (BT) using Support Vector Machine (SVM) model for spectra-wise classification .	130
7.20	Confusion matrix for the second step classification to identify brain tumor free (BTF) individuals and patients with glioblastoma (GB), meningioma (MNG) using Support Vector Machine (SVM) model for spectra-wise classification	131

7.21	ROC Curve for the classification to identify brain tumor-free (BTF) individuals and patients with glioblastoma (GB) or meningioma (MNG) using Support Vector Machine (SVM) model for spectra-wise classification	132
7.22	Confusion matrix for the first step classification to identify brain tumor free (BTF) individuals and patients with brain tumor (BT) using Support Vector Machine (SVM) model for patient-wise classification .	133
7.23	Confusion matrix for the second step classification to identify brain tumor free (BTF) individuals and patients with glioblastoma (GB), meningioma (MNG) using Support Vector Machine (SVM) model for spectra-wise classification	133
7.24	ROC Curve for the classification to identify brain tumor-free (BTF) individuals and patients with glioblastoma (GB) or meningioma (MNG) using Support Vector Machine (SVM) model for patient-wise classification	134
7.25	Confusion matrix for the first step classification to identify brain tumor free (BTF) individuals and patients with brain tumor (BT) using Neural Networks (NN) for spectra-wise classification	135
7.26	Confusion matrix for the second step classification to identify brain tumor free (BTF) individuals and patients with glioblastoma (GB), meningioma (MNG) using Neural Networks (NN) for spectra-wise classification	135
7.27	ROC Curve for the classification to identify brain tumor-free (BTF) individuals and patients with glioblastoma (GB) or meningioma (MNG) using Neural Networks (NN) for spectra-wise classification	136
7.28	Confusion matrix for the first step classification to identify brain tumor free (BTF) individuals and patients with brain tumor (BT) using Neural Networks (NN) for patient-wise classification	136

7.29	Confusion matrix for the second step classification to identify brain tumor free (BTF) individuals and patients with glioblastoma (GB), meningioma (MNG) using Neural Networks (NN) for spectra-wise classification	137
7.30	ROC Curve for the classification to identify brain tumor-free (BTF) individuals and patients with glioblastoma (GB) or meningioma (MNG) using Neural Networks (NN) for patient-wise classification	137
7.31	Confusion matrix for the first step classification to identify brain tumor free (BTF) individuals and patients with brain tumor (BT) using Convolutional Neural Networks (CNN) for spectra-wise classification .	138
7.32	Confusion matrix for the second step classification to identify brain tumor free (BTF) individuals and patients with glioblastoma (GB), meningioma (MNG) using Convolutional Neural Networks (CNN) for spectra-wise classification	139
7.33	ROC Curve for the classification to identify brain tumor-free (BTF) individuals and patients with glioblastoma (GB) or meningioma (MNG) using Convolutional Neural Networks (CNN) for spectra-wise classification	139
7.34	Confusion matrix for the first step classification to identify brain tumor free (BTF) individuals and patients with brain tumor (BT) using Convolutional Neural Networks (CNN) for patient-wise classification .	140
7.35	Confusion matrix for the second step classification to identify brain tumor free (BTF) individuals and patients with glioblastoma (GB), meningioma (MNG) using Convolutional Neural Networks (CNN) for patient-wise classification	140
7.36	ROC Curve for the classification to identify brain tumor-free (BTF) individuals and patients with glioblastoma (GB) or meningioma (MNG) using Convolutional Neural Networks (CNN) for patient-wise classification	141

7.37	Comparison of different artificial intelligence algorithms with spectral-wise classification: We obtained 99.3% prediction accuracy with Random Forest (RF), 99.7% with XG-Boost (XGB), 67.4% with Quadratic Discriminant Analysis (QDA), 80.3% with Support Vector Machine (SVM), 89.3% with Neural Networks (NN), 87% with Convolutional Neural Networks (CNN) classifiers	144
7.38	Comparison of different artificial intelligence algorithms with spectral-wise classification: We obtained 78.9% prediction accuracy with Random Forest (RF), 75.9% with XG-Boost (XGB), 63.5% with Quadratic Discriminant Analysis (QDA), 79.3% with Support Vector Machine (SVM), 50.8% with Neural Networks (NN), 84.2% with Convolutional Neural Networks (CNN) classifiers	145

ABBREVIATIONS

5-ALA	5-Aminolevulinic Acid
AD	Alzheimer's Disease
AI	Artificial Intelligence
BA-FC	Bead-Assisted Flow Cytometry
BBB	Blood-Brain Barrier
CNN	Convolutional Neural Network
CNS	Central Nervous System
CSF	Cerebrospinal Fluid
CTCs	Circulating Tumor Cells
ctDNA	Circulating Tumor DNA
Cryo-TEM	Cryogenic Transmission Electron Microscopy
DFA	Discriminant Function Analysis
EIS	Electrochemical Impedance Spectroscopy
EV	Extracellular Vesicle
FACS	Fluorescence-Activated Cell Sorting
FDTD	Finite-Difference Time-Domain
FISH	Fluorescence In Situ Hybridization
GB	Glioblastoma
HC	Healthy Control
iMRI	Intraoperative Magnetic Resonance Imaging
iSCAT	Interferometric Scattering Microscopy
K-NN	K-Nearest Neighbors
LDA	Linear Discriminant Analysis
LSPR	Localized Surface Plasmon Resonance

miRNA	MicroRNA
MNG	Meningioma
MS	Mass Spectrometry
MRI	Magnetic Resonance Imaging
NMR	Nuclear Magnetic Resonance
NN	Neural Network
NSC	Neural Stem Cell
NTA	Nanoparticle Tracking Analysis
OCT	Optical Coherence Tomography
PCA	Principal Component Analysis
PC	Photonic Crystal
PD	Parkinson's Disease
QDA	Quadratic Discriminant Analysis
QD	Quantum Dot
RF	Random Forest
RNN	Recurrent Neural Network
RT-TEM	Room Temperature Transmission Electron Microscopy
SAW	Surface Acoustic Wave
SERS	Surface-Enhanced Raman Spectroscopy
SOD1	Superoxide Dismutase 1
SPR	Surface Plasmon Resonance
SVM	Support Vector Machine
WB	Western Blot
WHO	World Health Organization
XGB	XG Boost

Chapter 1

INTRODUCTION

1.1 Glioblastoma (GB)

Glioblastoma (GB) is the most common and malignant primary brain tumor in adults, classified as a WHO Grade IV tumor based on histological and molecular features [Louis et al., 2021c, Louis et al., 2021d]. Characterized by rapid proliferation, diffuse infiltration, resistance to apoptosis, and significant cellular and molecular heterogeneity, GB presents a dire healthcare challenge with a median survival of just over a year [Ostrom et al., 2021, Dhermain and Marazano, 2022, Grochans et al., 2022]. The 2021 WHO Central Nervous System (CNS) tumor classification emphasizes the importance of molecular markers, with IDH-wild type status being a defining feature of GBs, while IDH-mutant tumors are categorized separately [Louis et al., 2021c, Weller et al., 2021, Ostrom et al., 2021]. GBs originate from neural stem cells (NSCs), NSC-derived astrocytes, and oligodendrocyte precursor cells, exhibiting aggressive traits such as mitotic activity, cellular pleomorphism, astrocytic differentiation, nuclear atypia, microvascular proliferation, and necrosis [Stoyanov et al., 2022, Louis et al., 2021d, Louis et al., 2021c].

Despite advancements in treatment options, including surgery, chemotherapy, and radiation therapy, the prognosis remains grim due to the tumor's intrinsic heterogeneity and resistance to treatment [Ali et al., 2020, Jackson et al., 2019]. This underscores the urgency to develop novel diagnostic strategies that allow for early cancer detection, thereby enhancing survival rates and improving patients' quality of life. Accurate diagnosis requires detailed neuropathological and molecular evaluations due to the significant heterogeneity in GBs [Louis et al., 2021d, Weller et al.,

2021]. Standard diagnostic methods such as hematoxylin and eosin (H&E) staining provide initial assessment but fail to identify specific molecular alterations necessary for targeted therapies [Louis et al., 2021c, Stoyanov et al., 2022]. The 2021 updates in CNS tumor classification underscore the need for comprehensive molecular testing, including immunohistochemistry and genetic analyses, to improve diagnostic precision and patient outcomes [Louis et al., 2021c, Ostrom et al., 2021]. However, these tests are time-consuming, and the subjective nature of histopathological evaluation poses challenges in diagnosing high-grade gliomas [Louis et al., 2021d, Dhermain and Marazano, 2022].

There is an unmet clinical need to detect GB rapidly, via easily accessible and accurate diagnostic tools. The current diagnosis of GB relies on the histopathological analysis of biopsy/surgically removed samples, in addition to costly and time-consuming molecular genetic tests, which are not available in every hospital, particularly in developing regions. Having a minimally invasive GB diagnostic tool, such as the one proposed in this study, would be a game-changer in eliminating unnecessary invasive procedures, including biopsies and surgeries, when not necessary.

GB remains a significant challenge in the field of neuro-oncology due to its rapid progression, infiltrative nature, and poor prognosis [Louis et al., 2021c, Gonzalez Castro and Wesseling, 2021, Hanif et al., 2017]. The current diagnostic techniques often fall short in providing early and accurate identification, leading to delayed interventions and compromised patient outcomes [Jain, 2018, Silantyev et al., 2019]. Traditional methods such as biopsy and surgical resection are time-consuming, costly, and invasive, so innovative approaches to revolutionize GB diagnosis are required [Horbinski et al., 2019]. This study aims to explore the potential of a liquid biopsy technique in revolutionizing GB diagnosis and therapy management, addressing the urgent need for more effective and less invasive diagnostic tools in the field of neuro-oncology. Herein, we utilized exosomes and artificial intelligence (AI) to enable the rapid, minimally invasive, and high-accuracy detection of GB through surface-enhanced Raman spectroscopy (SERS) analysis of blood exosomes.

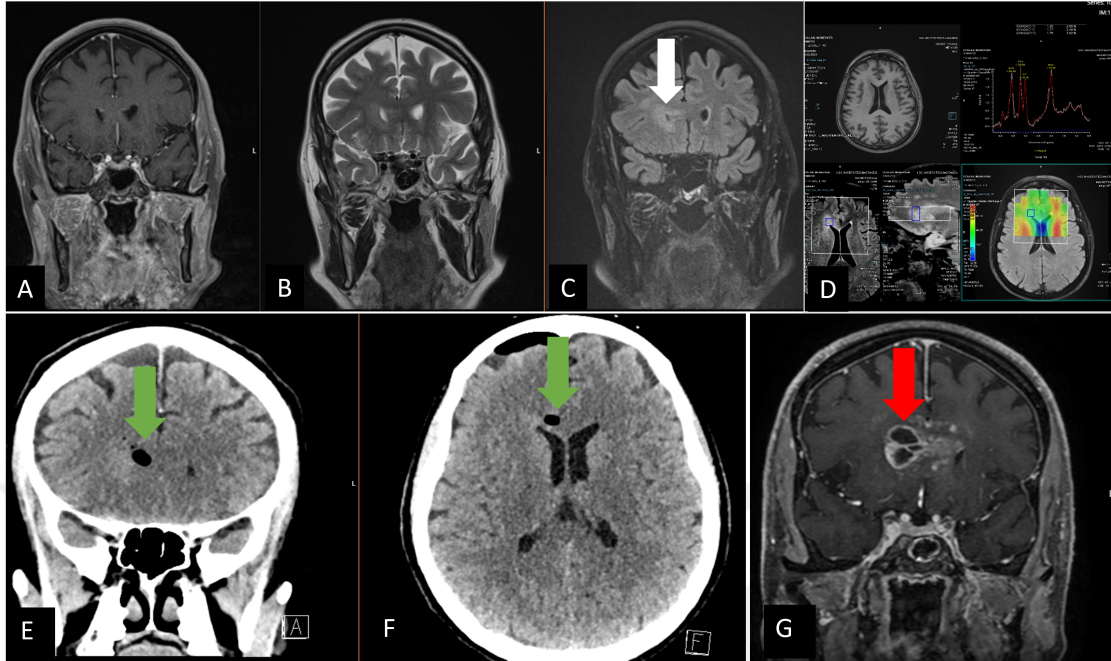


Figure 1.1: Example of biopsy of a brain lesion in a 71-year-old female patient. Initial gadolinium-enhanced coronal (A) T1-weighted (B) T2-weighted (C) FLAIR MRI prior to any treatment showing hyperintense area with indistinct boundaries at the right frontal horn (white arrow). (D) In post-contrast DSC perfusion imaging, no increased perfusion is detected within the described infiltrative signal anomalies. The choline/NAA ratio is less than 1 in all voxels. Early postoperative (E) coronal and (F) axial CT scan after stereotactic biopsy showing the trajectory of the needle and sampling areas (green arrows). Histopathological examination showed atypical glial cells. (G) At 3 months, a follow-up postcontrast coronal T1-weighted MRI showed the development of the tumor (red arrow). She proceeded to further surgery; the histopathology was consistent with glioblastoma. From the archive of Prof. Ihsan Solaroglu, MD.

1.2 Assessment of Glioma Progression and Pseudoprogression

In clinical practice, managing gliomas primarily relies on diagnostic imaging, which often lacks the sensitivity and specificity needed to accurately evaluate recurrence after primary treatment [Brandsma and van den Bent, 2009]. Distinguishing between pseudoprogression and true progression is vital for clinicians to avoid unnecessary surgeries and ineffective treatments. Traditionally, glioblastoma response assessment is conducted using contrast-enhanced (gadolinium) T1-weighted anatomical MRI [Wen et al., 2020]. However, this method mainly reflects blood-brain bar-

rier disruption and does not specifically measure tumor activity. True progression indicates actual tumor regrowth and spread despite treatment, whereas pseudoreponse, which can present as an apparent reduction in contrast enhancement in patients treated with bevacizumab, and necrosis following radiation therapy, may show similar MRI results [van Dijken et al., 2017].

Pseudoprogression, a temporary condition observed after radiotherapy and chemotherapy, can occur in approximately 10-30% of patients and may cause clinical symptoms [Muller Bark et al., 2020]. Differentiating these conditions using only contrast-enhanced MRI is often challenging, as both can exhibit similar imaging findings. Thus, additional imaging methods may be required for accurate diagnosis of pseudoprogression. Advanced imaging techniques such as Dynamic Susceptibility Contrast MRI (DSC-MRI) and amino acid PET imaging can aid in this differentiation. DSC-MRI assesses blood flow and volume in brain tissue, providing a more accurate evaluation of tumor activity. Although advanced MRI techniques offer higher diagnostic accuracy compared to anatomical MRI techniques in glioblastoma patients [van Dijken et al., 2017], recommendations for post-treatment imaging differ in terms of frequency and timing.

Currently, no biomarkers or clinical features are available to distinguish true glioma progression from pseudoprogression [Muller Bark et al., 2020]. Effective liquid biopsy methods may address these challenges in glioma management. Furthermore, advancements in understanding the tumor microenvironment of gliomas and progress in immunotherapy underscore the importance of assessing bodily fluid biomarkers for therapeutic planning and clinical management [Otsuji et al., 2024, Wen et al., 2020].

On the other hand, resection or biopsy of a brain tumor can pose risks to patients, including potential brain swelling, infection, hemorrhage, and possible impacts on neurological functions. Additionally, some tumors may be inaccessible due to their location [Nieder et al., 2005]. For instance, diffuse intrinsic pontine tumors (DIPG), comprising 10 to 15% of all childhood brain tumors, present one of the most significant therapeutic challenges in pediatric neuro-oncology. The role of diagnostic

biopsy in these tumors has been controversial due to the high eloquence of the brain stem and the lack of direct benefit for the patient [Puget et al., 2015]. If feasible, implementing liquid biopsy in DIPGs could aid in the molecular profiling of these tumors, help identify new therapeutic targets, and assist in assessing the response to current treatments.

Furthermore, tissue biopsies can sometimes fail to capture the heterogeneity of the entire tumor mass and may not accurately represent real-time tumor activity [Sottoriva et al., 2013]. (Figure 1.2)

Therefore, implementing liquid biopsy in gliomas is essential for early diagnosis, detecting minimal residual disease after surgery, assessing disease status post-treatment, identifying treatment resistance mechanisms, and predicting outcomes.

1.3 Motivation for Liquid Biopsy in Gliomas

Implementing liquid biopsy in gliomas is essential for several reasons. Firstly, it enables early diagnosis, which is critical for improving patient outcomes. By detecting circulating tumor DNA (ctDNA) and other biomarkers in bodily fluids, clinicians can identify the presence of gliomas before they become detectable through conventional imaging techniques [Wan et al., 2017]. This early detection can facilitate timely interventions, potentially slowing disease progression and enhancing the efficacy of treatments [Alix-Panabières and Pantel, 2020]. Secondly, liquid biopsy allows for the detection of minimal residual disease after surgery, helping to identify any remaining cancerous cells that could lead to recurrence [Siravegna et al., 2017]. This is particularly important in gliomas, where complete surgical resection is often challenging due to the tumor’s location and infiltrative nature [Miller and Perry, 2017]. Additionally, liquid biopsy provides a non-invasive means to assess disease status post-treatment, offering real-time insights into tumor dynamics and therapeutic responses [Pantel and Alix-Panabières, 2019]. This can guide adjustments in treatment plans and improve patient management. Moreover, by identifying treatment resistance mechanisms, liquid biopsy can help in tailoring personalized therapeutic strategies, ensuring that patients receive the most effective treatments

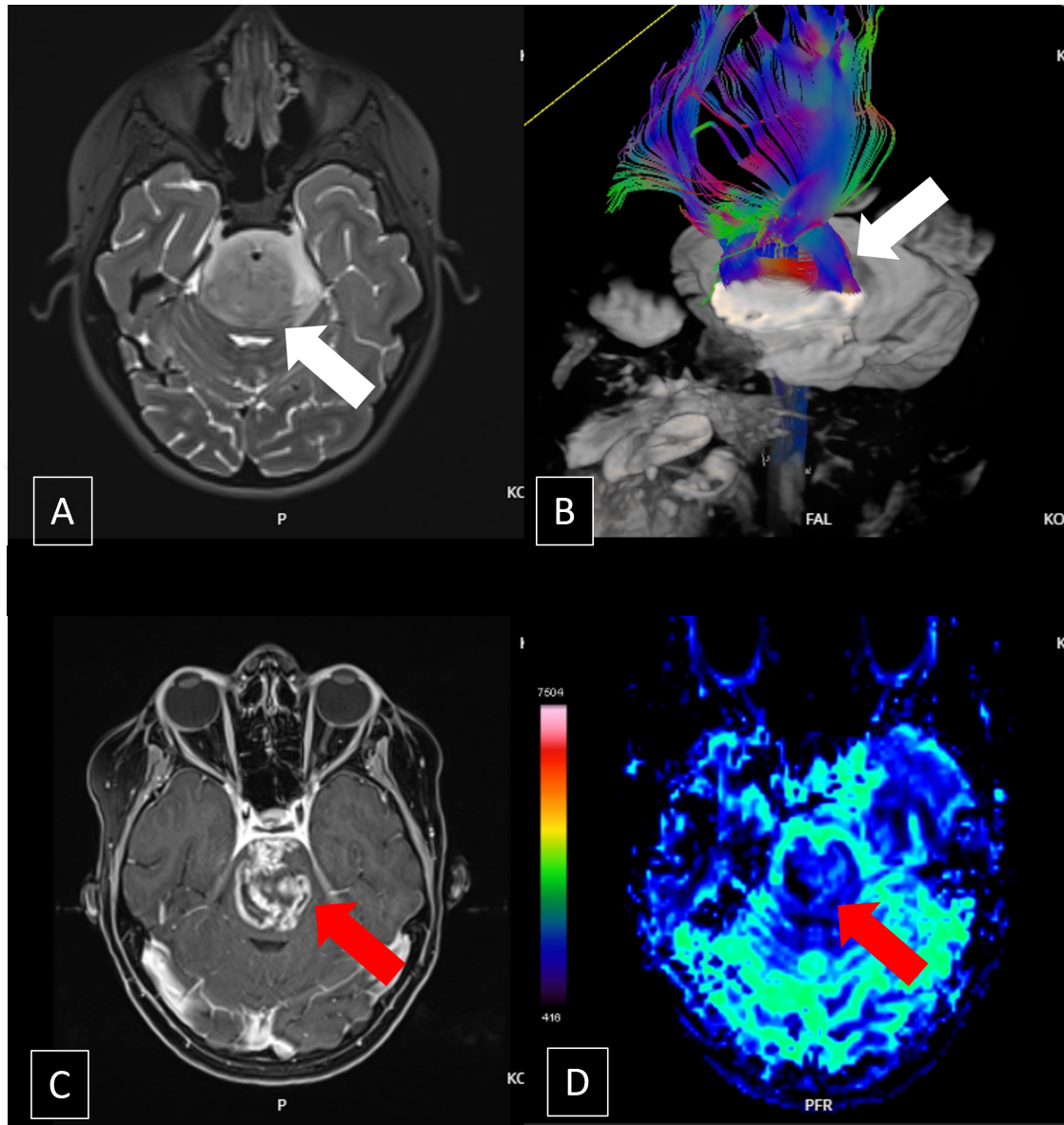


Figure 1.2: Figure. Key MRI images of a 7-year-old patient with a diffusely infiltrating pontine tumor. (A) axial T2-weighted MRI at initial diagnosis shows the abnormal signal intensity of pons and compression of the 4th ventricle (white arrow), (B) 3D color-coded diffusion tensor imaging showing the relationship between corticospinal tracts and the tumor (white arrow) (C) postcontrast axial T1-weighted MRI following radiotherapy treatment showing necrosis (red arrow) with (D) reduced cerebral blood volume (red arrow). From the archive of Prof. Ihsan Solaroglu, MD.

[DeMattos-Arruda et al., 2011]. Lastly, the ability to predict outcomes based on biomarker profiles can inform prognosis and support clinical decision-making, ul-

timately contributing to better patient care and improved survival rates [Heitzer et al., 2015].

1.4 Importance of the Study

Exosomes, or small extracellular vesicles secreted by cells, have emerged as promising candidates for non-invasive disease detection owing to their ability to carry a cargo of bioactive molecules reflective of their parent cells' physiological and pathological states [Shehzad et al., 2021, Doyle and Wang, 2019]. Exosomes are present in various bodily fluids, including blood, making them accessible for biomarker discovery. Recent research has shown their potential as carriers of information about the tumor microenvironment and molecular changes associated with diseases, including cancer [Jin et al., 2022].

Raman spectroscopy exploits the scattering of light to provide information about molecular vibrations and chemical compositions of materials [Hanna et al., 2022, Klammer et al., 2021]. SERS amplifies the Raman signal by orders of magnitude through the use of gold nanoparticles [Langer et al., 2019]. This technique has shown promise in the analysis of biological samples, including exosomes, due to its ability to provide specific molecular-level signatures [Zhang et al., 2019].

This thesis seeks to investigate the potential of exosomes as disease indicators, focusing on GB. By analyzing surface-enhanced Raman spectra of isolated exosomes, the hope is to uncover disease-specific molecular signatures that distinguish GB-derived exosomes from healthy and non-glial tumor-derived exosomes like those from meningioma (MNG). Leveraging the power of neural networks and machine learning, subsets of AI algorithms inspired by the human brain's functioning [Alzubaidi et al., 2021], we intend to develop a predictive model capable of accurately identifying GB based on Raman spectra patterns.

The integration of exosomes, SERS, and AI has the potential to improve GB diagnosis. By harnessing the unique characteristics of exosomes and the analytical power of SERS, this thesis aims to contribute to the development of a non-invasive, reliable, and early diagnostic tool for GB. Such an innovation has the potential to

significantly impact patient care for the better, enabling timely interventions and improved overall outcomes for individuals afflicted by this disease.

A clinically applicable liquid biopsy technique that can distinguish different brain tumors and their inflammatory effects would reduce:

1. Unnecessary imaging,
2. Unnecessary invasive and risky surgical procedures, including biopsy and/or craniotomy (surgical removal of the tumor),
3. Imaging and operation costs,
4. Physiological and functional burdens that these operations may cause,
5. Psychological effects of hospitalization on patients,
6. Number of malpractices and their sanctions to the hospital and the physicians, as well as damages to the patients.

Moreover, this technique may benefit in:

1. Identifying pseudo-progression, radiation necrosis, and tumor recurrence,
2. Monitoring the inflammatory effects of benign or malignant brain tumors in patient metabolism,
3. Supporting neuropathologists and pathologists at centers lacking expert neuropathologists to make these distinctions,
4. Supporting diagnostic predictions during frozen sessions in surgery to direct the operation more accurately.

Therefore, we propose to use the most frequent types of benign and malignant tumors to provide proof of concept for our proposed study to distinguish different types of brain tumors via their exosomal biomarkers.

1.5 Hypothesis and Specific Aims

1.5.1 Hypothesis

This study aims to develop a rapid, accurate, and cost-effective diagnostic strategy for GB using a liquid biopsy technique that detects exosomes via Surface-Enhanced Raman Spectroscopy (SERS) and employs Artificial Intelligence (AI) for tumor classification. The hypothesis is that SERS combined with AI can provide additional molecular-level information about disease pathology. This method aims to enhance our understanding of the origin and nature of exosomes, thereby improving tumor detection capabilities.

This hypothesis was tested through the comparative analysis of plasma exosomes derived from healthy individuals, meningioma (MNG) patients as benign tumor controls, and a molecularly uniform set of IDH-wild type GB patients (according to the World Health Organization 2021 Central Nervous System Tumor Classification) using SERS and AI. With this approach, we aim to obtain further information beyond that gained through routine pathological evaluation. This thesis project specifically focuses on developing a liquid biopsy technique that can identify the pathology of exosomes through SERS spectral signatures and AI.

1.5.2 Specific Aims

Specific Aim 1: Develop a standardized protocol for isolating and characterizing plasma-derived exosomes from healthy individuals, MNG patients, and GB patients using Surface-Enhanced Raman Spectroscopy (SERS).

Rationale: A robust and reproducible protocol for exosome isolation and characterization is critical for obtaining reliable and comparable SERS spectra from different sample groups.

- **Specific Aim 1a:** Isolate plasma-derived exosomes from healthy individuals, MNG patients, and GB patients using ExoTIC (Exosome Total Isolation Chip) [Liu et al., 2017].

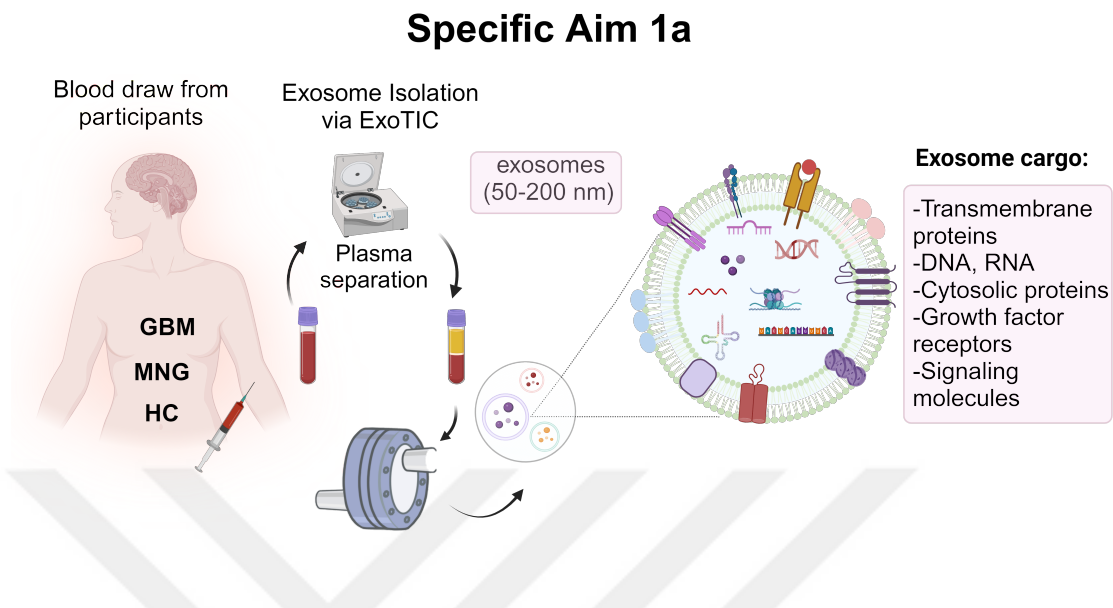


Figure 1.3: Schematic illustration of Specific Aim 1a. Sample collection from glioblastoma (GBM), meningioma (MNG), and healthy control (HC) participants, plasma separation, exosome isolation from these clinical samples. Created with BioRender.com

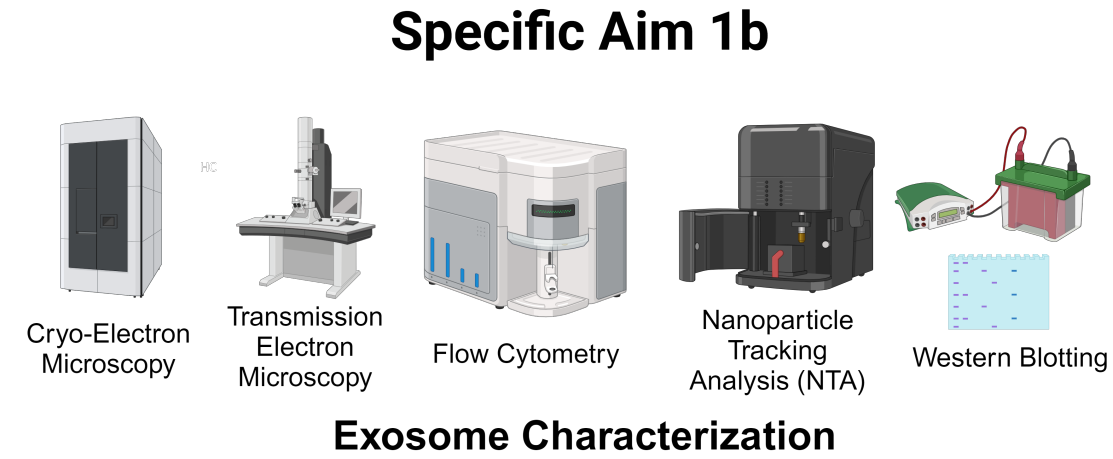


Figure 1.4: Schematic illustration of Specific Aim 1b: Characterization of exosomes isolated from glioblastoma (GBM), meningioma (MNG), and healthy control (HC) samples via Room Temperature Transmission Electron Microscopy (RT-TEM), Cryo-TEM, Interferometric Scattering Microscopy (iSCAT), Nanoparticle Tracking Analysis (NTA), Bead-Assisted Flow Cytometry (BA-FC), Western Blotting (WB) techniques. Created with BioRender.com

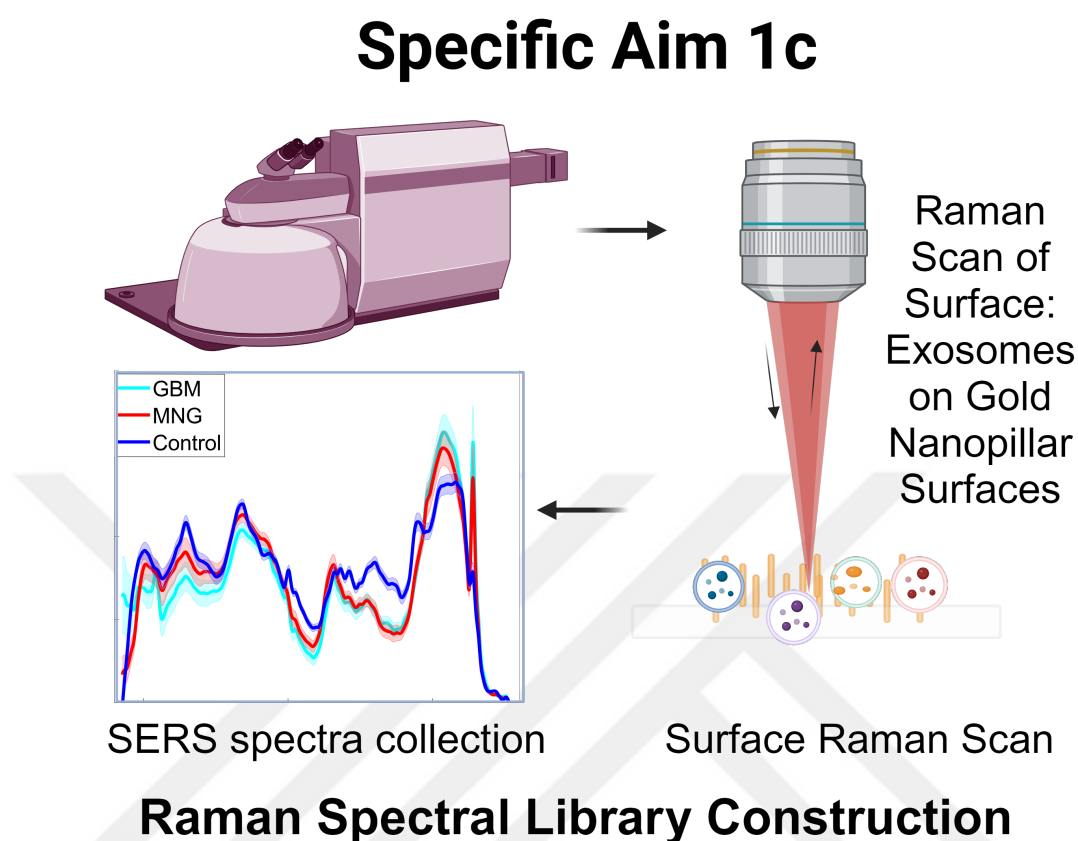


Figure 1.5: Schematic Illustration of Specific Aim 1c. Surface-Enhanced Raman Spectroscopy (SERS) measurements of exosomes derived from GB, MNG, and HC samples and spectral library construction. Created with BioRender.com

- **Specific Aim 1b:** Characterize isolated exosomes using nanoparticle tracking analysis (NTA), room temperature transmission electron microscopy (RT-TEM), cryo-TEM, iSCAT (Interferometric Scattering Microscopy), Western blotting, Bead-Assisted Flow Cytometry (BA-FC), for exosome markers.
- **Specific Aim 1c:** Acquire SERS spectra of isolated exosomes using optimized conditions.

Specific Aim 2: Perform proteomics analysis to investigate the protein signatures of exosomes from different patient groups.

Rationale: Proteomics analysis can provide a comprehensive molecular profile of exosomes, identifying protein signatures that are specific to different pathology

conditions. **Specific Aim 3: Apply artificial intelligence (AI) techniques to analyze SERS spectra for the identification and classification of exosomes from different patient groups.**

Rationale: AI can enhance the analysis of complex SERS spectra, enabling the identification of subtle molecular differences between exosomes from different sample groups.

- **Specific Aim 2a:** Pre-process SERS spectra to remove noise and normalize the data.
- **Specific Aim 2b:** Develop and train machine learning models (e.g., support vector machines, neural networks) using a subset of the SERS spectra.
- **Specific Aim 2c:** Validate the models with independent test sets and evaluate their performance in distinguishing exosomes from healthy individuals, MNG patients, and IDH-wt GB patients.
- **Specific Aim 3a:** Perform proteomics analysis on exosome samples to identify and quantify protein content.
- **Specific Aim 3b:** Compare the proteomics profiles of exosomes from healthy individuals, MNG patients, and GB patients.
- **Specific Aim 3c:** Correlate the SERS and proteomics results to identify key spectral features and molecular signatures associated with different patient groups.

1.6 Scope of the Thesis

This thesis covers the following:

1. An overview of glioblastoma and current diagnostic challenges.
2. A review of current diagnostic methods and their limitations.

3. An introduction to liquid biopsy techniques and their advantages.
4. The role of exosomes in cancer detection and their molecular characteristics.
5. Principles of SERS and its applications in biomedical research.
6. An overview of AI and deep learning algorithms in medical diagnostics.
7. The integration of proteomics analysis for comprehensive molecular profiling of exosomes.



Chapter 2

LITERATURE REVIEW

2.1 Glioblastoma Diagnostics

Glioblastoma (GB) is the most common and aggressive primary brain tumor in adults, characterized by a poor prognosis and high mortality rate [Delgado-López and Corrales-García, 2016, Louis et al., 2021a, O'Neill and Lee, 2021]. According to the 2021 World Health Organization (WHO) CNS tumor classification, GB is now strictly defined as an IDH-wildtype (IDH-wt) tumor, reflecting its highly malignant nature and specific molecular features [Louis et al., 2021a, O'Neill and Lee, 2021]. This updated classification emphasizes the integration of histopathological and molecular diagnostics for accurate GB identification and classification [Louis et al., 2021b].

2.1.1 Histopathological Features

Histopathological evaluation remains a cornerstone in the diagnosis of GB, typically involving hematoxylin and eosin (H&E) staining and light microscopy to identify key histological features such as cellular pleomorphism, nuclear atypia, mitotic activity, microvascular proliferation, and necrosis [et al., 2021c, Zhou et al., 2019]. Despite its importance, the subjective nature of histopathological assessments and the extensive time required for molecular testing pose significant challenges [et al., 2021b, T. Garzon-Muvdi and Chaichana, 2017].

2.1.2 Molecular Characteristics

GB is characterized by its molecular heterogeneity, which includes mutations in the TERT promoter, EGFR amplification, and alterations in PTEN and TP53 [et al.,

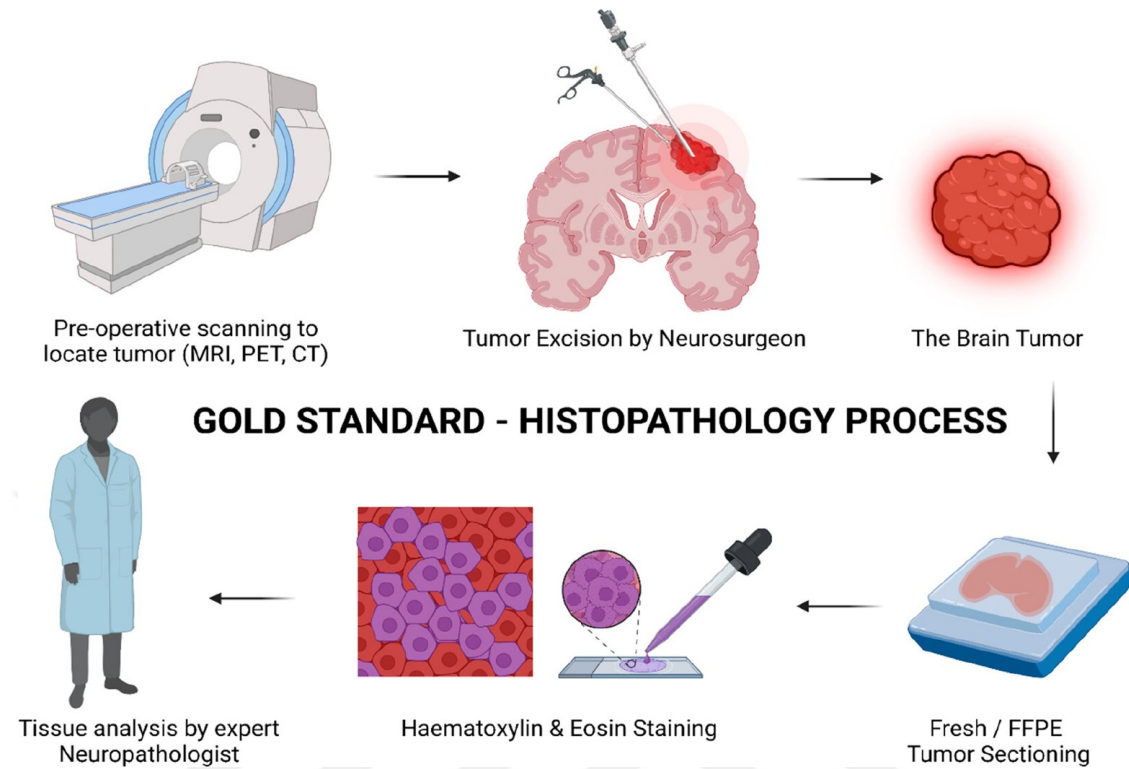


Figure 2.1: Clinical workflow for histopathological evaluation of the brain tumors. Adopted from [Mohamed et al., 2024])

2021e, Philipp Balermipas and Hegi, 2021, A. Radbruch and Wiestler, 2021]. These molecular features, combined with histological characteristics, necessitate advanced diagnostic techniques to ensure precise and timely diagnosis. Molecular diagnostics often involve immunohistochemistry, fluorescence in-situ hybridization (FISH), Sanger sequencing, and next-generation sequencing [et al., 2021a].

2.1.3 Diagnostic Imaging and Techniques

Recent advancements in diagnostic imaging have significantly improved the delineation of tumor boundaries during surgical interventions. Techniques such as functional neuro-navigation, intraoperative MRI (iMRI), ultrasound, fluorescence-guided surgery, and optical coherence tomography (OCT) have become integral to the surgical management of GB [et al., 2021d, Sven Schipmann and Stummer, 2022, Stummer et al., 2013]. Fluorescence-guided surgery using 5-aminolevulinic acid (5-ALA) has

shown increased accuracy in tumor resection by highlighting malignant tissues under a surgical microscope [Stummer et al., 2013]. However, there remains a critical need for high-resolution, intraoperative technologies that provide real-time differentiation between normal and tumor tissues to optimize surgical outcomes and improve patient prognosis [Sven Schipmann and Stummer, 2022, Michael et al., 2019].

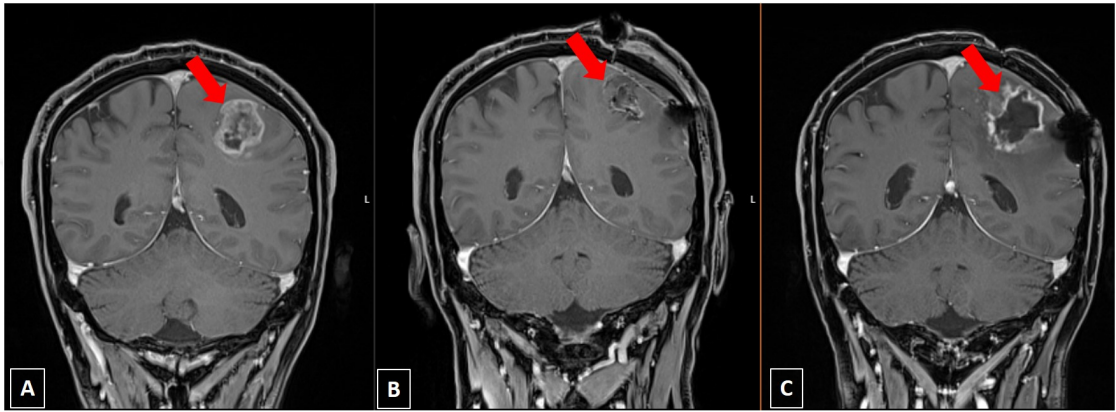


Figure 2.2: Contrast-enhanced T1-weighted MRI images of a patient (A) with a left parietal lobe GB. (B) Early post-op image showing complete GB removal. (C) Radiation necrosis developed around surgical cavity, 6 months post radiation therapy (From the archive of Prof. Ihsan Solaroglu, MD)

2.1.4 Prognosis and Treatment Challenges

Despite advancements in treatment modalities, the prognosis for patients with IDH-wt GB remains dismal, with a median survival of approximately 15 months post-diagnosis [et al., 2021f, Philipp Balermpas and Hegi, 2021]. The aggressive nature of GB, coupled with its diffuse infiltration into surrounding brain tissue, complicates surgical resection and subsequent treatments. Radiotherapy and chemotherapy, particularly with temozolomide, are standard post-surgical treatments, yet the recurrence rate of GB is high, and long-term survival rates remain low [Louis et al., 2021a, Philipp Balermpas and Hegi, 2021].

2.1.5 Future Directions

The development of novel diagnostic tools and therapeutic strategies is crucial for improving GB patient outcomes. Advances in molecular biology and imaging technologies hold promise for more accurate diagnoses and personalized treatment approaches. High-resolution imaging techniques, combined with real-time molecular diagnostics, could significantly enhance the precision of GB surgery and subsequent treatment plans. Moreover, ongoing research into the molecular pathways and genetic alterations of GB could lead to targeted therapies that improve survival rates and quality of life for patients with this devastating disease [A. Radbruch and Wiestler, 2021, Philipp Balermipas and Hegi, 2021].

2.2 Liquid Biopsy

Liquid biopsy represents a transformative approach in cancer diagnostics, enabling the detection and analysis of biomarkers from biofluids such as blood, urine, and cerebrospinal fluid. This technique offers a non-invasive alternative to traditional tissue biopsies, presenting several significant advantages.

2.2.1 Introduction to Liquid Biopsy Techniques and Their Advantages

Liquid biopsy encompasses various techniques to detect and analyze tumor-derived components circulating in biofluids. Key components of liquid biopsy include circulating tumor cells (CTCs), circulating tumor DNA (ctDNA), and extracellular vesicles (EVs) such as exosomes.

Minimally Invasive: Liquid biopsies involve the collection of a small volume of biofluid, reducing patient discomfort and the risk of complications associated with surgical biopsies [Pantel and Alix-Panabières, 2019, Cheung et al., 2021].

Real-Time Monitoring: They facilitate continuous monitoring of tumor dynamics, allowing for timely adjustments in treatment strategies based on real-time data [Cheung et al., 2021, Alix-Panabières and Pantel, 2020].

Early Detection: By detecting circulating tumor cells (CTCs), cell-free DNA

(cfDNA), and extracellular vesicles (EVs) at early stages, liquid biopsies can contribute to the early diagnosis of cancers, potentially improving patient outcomes [Mandel et al., 2024, Keller and Pantel, 2022].

Comprehensive Tumor Profiling: Liquid biopsies can capture the heterogeneity of tumors better than tissue biopsies, which may miss subclonal variations due to sampling limitations [Wan et al., 2017, Siravegna et al., 2017].

2.2.2 Challenges in Liquid Biopsy Techniques

Despite their potential, liquid biopsies face several challenges that hinder their widespread clinical adoption.

Sensitivity and Specificity: Ensuring high sensitivity and specificity in detecting rare biomarkers within a complex mixture of biofluid components is a significant challenge [Bettegowda et al., 2014, Crowley et al., 2020].

Standardization: Lack of standardized protocols for sample collection, processing, and analysis can lead to variability in results, impacting the reproducibility and reliability of liquid biopsies [Schwarzenbach et al., 2014, Heitzer et al., 2015].

Analytical Validity: Establishing the analytical validity of liquid biopsy assays is critical, requiring rigorous validation studies to confirm their accuracy and precision [Diamandis, 2015, Cheung et al., 2021].

Clinical Utility: Demonstrating the clinical utility of liquid biopsies in improving patient outcomes is essential for their integration into routine clinical practice [Merker et al., 2018, Haber and Velculescu, 2017].

2.2.3 Importance and Potential of Liquid Biopsy in Brain Tumors

The application of liquid biopsy in brain tumors, such as glioblastoma, holds significant promise due to the challenges associated with traditional biopsy methods.

Non-Invasiveness: Given the sensitive nature of brain tissue, liquid biopsies offer a safer alternative for obtaining tumor-related information without the risks associated with surgical interventions [Best et al., 2015, De Mattos-Arruda et al., 2015].

Monitoring Tumor Dynamics: Liquid biopsies enable the monitoring of tumor evolution and treatment response, providing critical information for adjusting therapeutic strategies [Ostrom et al., 2021, Mandel et al., 2024].

Overcoming Sampling Bias: Traditional biopsies may not capture the heterogeneity of brain tumors, while liquid biopsies can provide a more comprehensive tumor profile by analyzing biomarkers shed into the biofluid [Weller et al., 2021, Heitzer et al., 2015].

2.2.4 Current Limitations and Future Directions

While promising, current liquid biopsy techniques have limitations that need to be addressed to achieve clinical validation.

Low Biomarker Concentration: The low concentration of tumor-derived biomarkers in biofluids requires highly sensitive detection methods, which are still being optimized [Bettegowda et al., 2014, Crowley et al., 2020].

Technical Variability: Variability in pre-analytical and analytical procedures can affect the consistency of results, highlighting the need for standardized protocols [Schwarzenbach et al., 2014, Heitzer et al., 2015].

Cost and Accessibility: The cost of advanced liquid biopsy assays and the need for specialized equipment can limit their accessibility in clinical settings [Merker et al., 2018, Haber and Velculescu, 2017].

Future Research: Ongoing research focuses on improving the sensitivity, specificity, and standardization of liquid biopsy techniques and validating their clinical utility through large-scale studies [Cheung et al., 2021, Diamandis, 2015].

Integration with Other Technologies: Combining liquid biopsy with other diagnostic technologies, such as imaging and molecular profiling, may enhance its diagnostic accuracy and clinical relevance [Alix-Panabières and Pantel, 2020, Wan et al., 2017].

In summary, liquid biopsy represents a promising approach in the early detection, monitoring, and comprehensive profiling of brain tumors. Addressing the current challenges through continued research and technological advancements will be cru-

cial for its successful clinical implementation.

2.3 Exosomes in Cancer Diagnostics

Exosomes are small, membrane-bound vesicles ranging from 30 to 200 nm in diameter, secreted by various cell types, including cancer cells. They contain a complex cargo of proteins, lipids, and nucleic acids, which can provide valuable insights into the tumor's molecular characteristics. In glioblastoma, exosomes have been shown to carry specific molecular signatures that can detect and classify different tumor types [Hanson et al., 2016, Ferraro et al., 2003].

2.3.1 Role of Exosomes in Cancer Detection

Exosomes play a critical role in intercellular communication by transferring their cargo from donor to recipient cells, influencing various physiological and pathological processes [Tkach and Théry, 2016, Colombo et al., 2014]. In cancer, exosomes contribute to tumor progression by promoting angiogenesis, immune evasion, and metastasis [Zhang et al., 2015b, Peinado et al., 2012]. Their presence in biofluids makes them an attractive target for liquid biopsy.

Molecular Characteristics: Exosomes are generally 30-200 nm in size and are formed through the endosomal pathway. This involves the inward budding of early and late endosomes to form multivesicular bodies (MVBs), fusing with the plasma membrane to release exosomes into the extracellular space [Raposo and Stoorvogel, 2013, van Niel et al., 2018]. Exosomes carry a diverse range of biomolecules, including DNA, RNA, proteins, and lipids, reflecting the molecular composition of their cells of origin. This cargo can provide a comprehensive tumor molecular profile [Valadi et al., 2007, Kalluri and LeBleu, 2020].

Diagnostic Potential:

Exosomes are found in various bodily fluids, including blood, urine, saliva, and cerebrospinal fluid, making them accessible for non-invasive diagnostic tests [Armstrong and Wildman, 2018]. They are particularly useful in cancer diagnostics due to their stability and ability to reflect the dynamic state of the tumor. Recent ad-

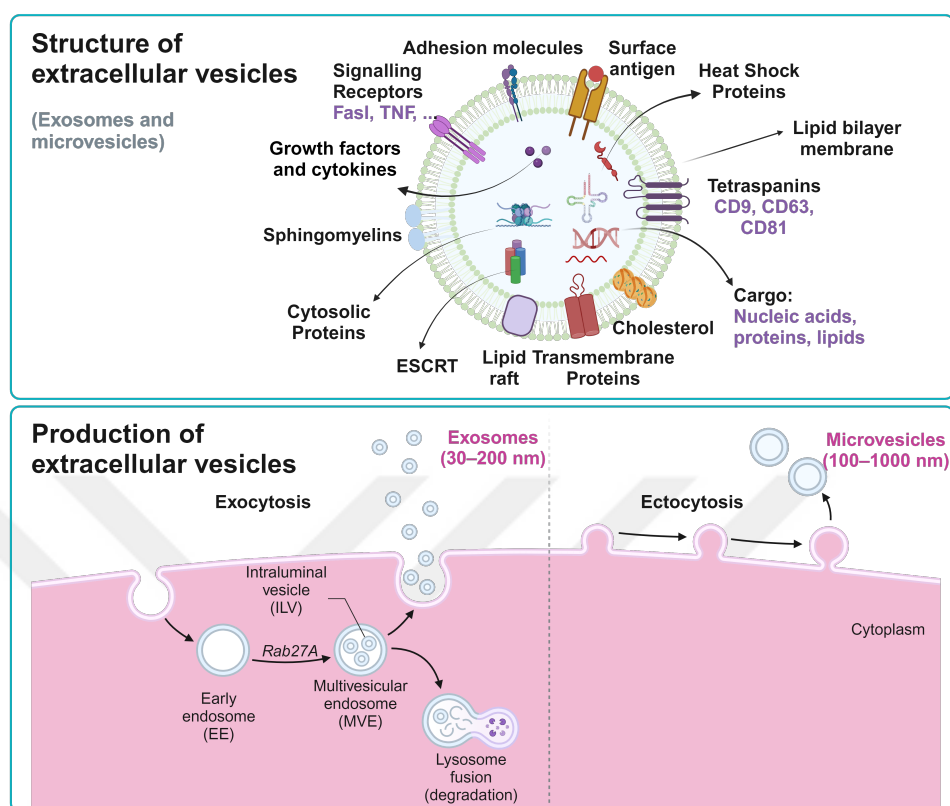


Figure 2.3: Illustration of extracellular vesicle structure, formation, and content. Exosomes, ranging from 30 to 200 nm in diameter, carry proteomic and genetic information from their cell of origin. They contain a diverse array of biomolecules, including transmembrane proteins, tetraspanins, growth factors, signaling receptors, lipid rafts, cholesterol, cytosolic proteins, sphingomyelins, adhesion molecules, heat shock proteins, and various cargo molecules such as RNA and DNA [Théry et al., 2002, Simons and Raposo, 2011, Kowal et al., 2014]. Exosome formation involves the endosomal pathway, where early endosomes mature into late endosomes or multivesicular bodies (MVBs). The intraluminal vesicles (ILVs) within these MVBs are released into the extracellular space as exosomes upon fusion with the plasma membrane [Raposo and Stoorvogel, 2013, van Niel et al., 2018]. In contrast, microvesicles, another class of extracellular vesicles, form directly through the outward budding and fission of the plasma membrane [Cocucci et al., 2009, Yáñez-Mó et al., 2015].

vances in high-throughput sequencing and proteomics have enabled the identification of specific exosomal biomarkers associated with different types of cancer [Melo et al., 2015, Thakur et al., 2014].

Exosome-based liquid biopsies have shown potential in diagnosing various can-

cers, including lung, breast, and prostate cancer, by analyzing specific exosomal biomarkers [Taylor and Gercel-Taylor, 2008, Skog et al., 2008]. Exosomes contain cancer-specific materials such as oncogenic DNA, RNA, and proteins, which facilitate tumor progression and metastasis through their roles in signaling pathways within the tumor microenvironment [Xu et al., 2018, Kalluri and LeBleu, 2020].

As exosomes circulating in extracellular fluid contain cancer-specific fingerprint materials (i.e., oncogenic genetic materials, oncoproteins), they facilitate tumor progression through their role in signaling pathways in the tumor microenvironment. Therefore, exosomes are critically important in diagnostics, as well as their role in the therapy of cancer and various diseases in clinics [Xu et al., 2018]. In recent decades, many techniques have been developed to detect exosomes in bodily fluids.

2.3.2 Exosomes in Brain-Related Diseases

Exosomes are emerging as crucial players in the pathology of brain-related diseases, participating in processes such as neuroinflammation, neurodegeneration, and tumorigenesis within the central nervous system (CNS). These extracellular vesicles can cross the blood-brain barrier (BBB), carrying a variety of biomolecules, including proteins, lipids, and nucleic acids, which reflect the state of brain health. This unique capability makes exosomes valuable tools for diagnosing and monitoring neurological conditions [Soria et al., 2017, Quek and Hill, 2017, Fiandaca et al., 2015].

Exosomes in Neuroinflammation

Neuroinflammation is a hallmark of various CNS diseases, including multiple sclerosis (MS), Alzheimer’s disease (AD), and Parkinson’s disease (PD). Exosomes derived from microglia and astrocytes can modulate inflammatory responses by transferring cytokines, chemokines, and other inflammatory mediators to surrounding cells. This intercellular communication can either exacerbate or ameliorate inflammation, depending on the context [Verderio et al., 2012, Frühbeis et al., 2013]. For instance, exosomes from activated microglia can propagate neuroinflammation by spreading pro-inflammatory cytokines, while those from astrocytes can carry anti-

inflammatory signals that help in resolving inflammation [Frühbeis et al., 2013].

Exosomes in Neurodegenerative Diseases

In neurodegenerative diseases such as Alzheimer's disease (AD), Parkinson's disease (PD), and amyotrophic lateral sclerosis (ALS), exosomes play significant roles in the propagation and spread of pathogenic proteins. For example, in AD, exosomes can transport amyloid-beta and tau proteins, which form amyloid plaques and neurofibrillary tangles, respectively [Soria et al., 2017, Quek and Hill, 2017]. In PD, exosomes can carry alpha-synuclein, facilitating its spread across different brain regions and contributing to the progression of the disease [Fiandaca et al., 2015, Goetzl et al., 2016]. Similarly, in ALS, exosomes can transport superoxide dismutase 1 (SOD1) and other misfolded proteins, promoting their dissemination and the associated neurotoxicity [Basso et al., 2013, Sproviero et al., 2018]. The presence of these pathogenic proteins in exosomes from patients' cerebrospinal fluid (CSF) and blood offers potential biomarkers for early diagnosis and monitoring of disease progression [Fiandaca et al., 2015, Goetzl et al., 2016].

Exosomes in Brain Tumors and Glioblastoma

Exosomes are particularly significant in the context of brain tumors, including glioblastoma (GB), the most aggressive primary brain tumor. GB-derived exosomes contain various oncogenic proteins, RNA, and DNA that can modulate the tumor microenvironment, promote angiogenesis, and facilitate tumor cell invasion and immune evasion [Skog et al., 2008, O'Brien et al., 2013]. These exosomes can carry molecular signatures specific to GB, which can be detected in biofluids, providing non-invasive diagnostic and prognostic biomarkers. Additionally, exosomes can influence the tumor microenvironment by transferring oncogenic signals to neighboring cells, thus promoting tumor growth and metastasis [Skog et al., 2008, O'Brien et al., 2013].

Diagnostic and Prognostic Biomarkers:

The diagnostic potential of exosomes lies in their ability to carry disease-specific biomarkers that can be detected in easily accessible biofluids such as blood, urine, and CSF. This non-invasive approach can significantly aid in the early detection and monitoring of brain-related diseases, offering a less risky alternative to traditional biopsy methods [Melo et al., 2015, Thakur et al., 2014]. Advances in high-throughput sequencing, proteomics, and lipidomics have enabled the identification of specific exosomal biomarkers associated with different neurological conditions, further enhancing their diagnostic utility [Armstrong and Wildman, 2018, Wan et al., 2017].

Glioblastoma-derived exosomes carry specific molecular signatures that can serve as diagnostic and prognostic biomarkers. For example, EGFRvIII, a common mutation in GB, has been detected in exosomes isolated from the plasma of GB patients [Skog et al., 2008]. Similarly, miRNAs such as miR-21, miR-221, and miR-222, up-regulated in GB, have been found in exosomes and can differentiate GB from other brain tumors [O'Brien et al., 2013, Akers et al., 2013].

Therapeutic Potential:

Exosomes are being actively explored as delivery vehicles for therapeutic agents in treating brain tumors and other brain-related diseases. Engineered exosomes can be loaded with small interfering RNA (siRNA), microRNA (miRNA), chemotherapeutic drugs, or other therapeutic molecules, allowing for targeted delivery to tumor cells. This targeted system can cross the BBB and achieve higher concentrations of therapeutic agents at the tumor site, minimizing systemic side effects [Alvarez-Erviti et al., 2011, Zhuang et al., 2011]. The natural ability of exosomes to traverse the BBB and deliver their cargo to specific target cells makes them promising candidates for revolutionizing the treatment of CNS disorders with minimal side effects [Alvarez-Erviti et al., 2011, Sun et al., 2010a].

Intercellular Communication and Tumor Microenvironment Modulation:

GB-derived exosomes can influence the tumor microenvironment by transferring oncogenic signals to surrounding cells, promoting a pro-tumorigenic environment. They can modulate the behavior of endothelial cells to enhance angiogenesis, recruit immune cells to create an immunosuppressive microenvironment, and alter the extracellular matrix to facilitate tumor invasion [O'Brien et al., 2013, Treps et al., 2017].

Given their critical role in brain tumor biology, exosomes offer a non-invasive means to monitor tumor dynamics, assess treatment response, and potentially deliver targeted therapies. Ongoing research is focused on improving the isolation and characterization of exosomes from patients with brain tumors, optimizing their use as biomarkers, and developing exosome-based therapeutic strategies.

In summary, exosomes play multifaceted roles in brain-related diseases, from modulating neuroinflammation and propagating neurodegenerative proteins to influencing tumor microenvironments and providing diagnostic and therapeutic opportunities. Addressing the challenges associated with their clinical application, such as standardization and sensitivity of detection methods, will be crucial for harnessing their full potential in neurological healthcare.

2.4 Surface-Enhanced Raman Spectroscopy (SERS)

Surface-Enhanced Raman Spectroscopy (SERS) is a powerful analytical technique that enhances the Raman scattering signals of molecules adsorbed on nanostructured metal surfaces. SERS offers high sensitivity and specificity, making it an ideal tool for detecting the molecular content of exosomes. By combining SERS with liquid biopsy, it is possible to develop a non-invasive method to detect and classify glioblastoma using the molecular signatures carried by exosomes [Hanson et al., 2016, Ferraro et al., 2003].

2.4.1 Principles of SERS and Its Applications in Biomedical Research

SERS relies on the enhancement of Raman signals by metal nanoparticles, typically gold or silver, which create localized surface plasmon resonances that amplify the Raman scattering of nearby molecules [Kneipp et al., 2002, Nie and Emery, 1997]. Raman spectroscopy is an analytical chemistry method that uses the principle of inelastic scattering of light. Raman spectroscopy provides molecular fingerprint information as a result of the wavelength difference between the incident and scattered light based on the chemical structure of the given molecule [Park et al., 2017, Carmicheal et al., 2019].

Raman spectroscopy is a relatively rare event, occurring approximately once in every 10^6 scattering events, leading to lower signal intensity compared to elastic Rayleigh scattering. To address this limitation, the Raman signal is enhanced by metal or nanostructures, resulting in SERS [Kneipp et al., 2002, Nie and Emery, 1997]. As a label-free and non-destructive technique that provides molecule-specific information, SERS is a highly powerful technique utilized in biomarker detection in biomedicine. This feature of Raman spectroscopy offers a high translational advantage for analyzing patient sample-derived exosomes without requiring lengthy processing times [Park et al., 2017, Carmicheal et al., 2019].

To detect either isolated exosomes from various body fluids or distinguish exosomes from other molecules in body fluids using SERS, many techniques have been employed. One such technique is using the nanoplasmonic effect on a SERS substrate as a signal enhancer to distinguish tiny differences in Raman spectra. Gold (Au) or silver (Ag) nanoparticles are commonly used to coat SERS substrates, facilitating the enhancement of the Raman signal and allowing for more sensitive Raman spectra for each given molecule. Besides signal enhancement, metal-coated SERS substrates require less sample for exosome detection and further analysis [Park et al., 2017, Carmicheal et al., 2019].

Various pattern geometries obtained through different algorithms have been utilized to enhance the plasmonic effect on SERS substrates. One such algorithm resulting in an optimal theoretical nanostructure geometry providing up to 10^{15-17} in-

crease in Raman signal enhancement is the genetic algorithm, generated using finite-difference time-domain (FDTD) modeling. This genetic-algorithm-driven SERS metallic substrate optimization technique provides functional plasmonic surface design through several competing design objectives [Bilgin et al., 2020].

The use of mathematical processing and machine learning algorithms can also help distinguish the exosome Raman spectra from other molecules. Principal component analysis (PCA) is used to reduce variations in the Raman spectral dataset, and the identified principal components are then coupled with discriminant functional analysis (DFA) or linear discriminant analysis (LDA) [Maguire et al., 2015, Shin et al., 2018]. Other machine learning algorithms, such as support vector machines, random forests, and k-nearest neighbors (k-NN), have also been utilized to classify the Raman dataset [Ullah et al., 2018, Sattlecker et al., 2010, Seifert et al., 2020, Dingari et al., 2013, Jiang et al., 2019].

SERS has been used to identify cancer-derived exosomes in various studies. For example, the Raman spectra of lung cancer-derived exosomes were distinguished from normal cell-derived exosomes using PCA analysis, achieving 95.3% sensitivity and 97.3% specificity [Park et al., 2017]. In another study, pancreatic cancer cell line-derived exosomes were investigated using gold-nanoparticle coated SERS substrates, and the Raman signal enhancement of gold nanoparticles allowed for the determination of exosome subpopulations when coupled with PCA-DFA processing of spectral data [Carmicheal et al., 2019]. Non-small cell lung cancer (NSCLC)-derived exosomal protein markers (CD9, CD81, EpCAM, and EGFR) were also analyzed using SERS, and the Raman spectra of the cancerous exosomal proteins were distinguished using PCA analysis [Shin et al., 2018]. Gold nanorods were used to enhance the Raman signal, allowing the detection of exosomes from MDA-MB-231, MDA-MB-468, and SKBR3 breast cancer cells with a limit of detection of 2×10^6 exosomes/mL. The Raman spectra of exosomes derived from these breast cancer cells were clearly distinguished from exosomes derived from MCF12A healthy breast cells.

Conclusion: Exosomes are promising molecules in cancer diagnostics as they carry parental cell-specific information and have garnered increased attention in

cancer research. When coupled with high specificity and sensitivity engineering techniques, exosomes can be used in clinical diagnostics for cancer and other diseases. The specificity of Raman spectroscopy, empowered with the high sensitivity of plasmonic surface signal enhancement and machine learning techniques, allows for distinguishing healthy and cancerous cell-derived exosomes. Although these techniques require further improvement, once the SERS-aided exosome detection technique is validated with clinical samples of reasonable size, it may be used as a diagnostic tool for cancer and various diseases in clinical settings in the near future.

Biomedical Applications: SERS has been employed in various biomedical applications, including pathogen detection, drug delivery monitoring, and cancer diagnostics. Its ability to provide molecular-level information makes it a valuable tool in these fields [Vo-Dinh et al., 1999, Zong et al., 2018].

Detection of Exosomes: Combining SERS with exosome isolation techniques allows for the sensitive detection of exosomal biomarkers, enabling the differentiation of tumor-derived exosomes from those of healthy cells [Lane et al., 2015, Shin et al., 2018].

2.5 Artificial Intelligence in Diagnostics

Artificial intelligence (AI) has revolutionized the field of medical diagnostics by enabling the analysis of complex datasets and the identification of patterns that may not be apparent to human observers. In the context of glioblastoma detection, AI can be used to analyze the SERS spectra of exosomes to classify different tumor types. Machine learning algorithms, such as support vector machines (SVM) and neural networks, can be trained on SERS data to develop predictive models for glioblastoma classification [Ellis and Goodacre, 2006, Butler et al., 2016].

2.5.1 Overview of AI and Deep Learning Algorithms in Medical Diagnostics

AI encompasses various computational techniques, including machine learning and deep learning, which can process large amounts of data to identify patterns and make predictions [LeCun et al., 2015, Jordan and Mitchell, 2015].

Deep Learning: Deep learning algorithms, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have demonstrated exceptional performance in image and signal analysis, making them suitable for medical diagnostics [Esteva et al., 2017, Litjens et al., 2017].

Application in SERS Data Analysis: AI algorithms can be trained on SERS spectra to distinguish between different molecular signatures, enabling the classification of exosomes based on their tumor origin. This approach has the potential to improve the accuracy and efficiency of liquid biopsy-based diagnostics [Butler et al., 2016, Ellis and Goodacre, 2006].

2.6 Emerging Techniques for Liquid Biopsy-Based Detection of Glioblastoma

2.6.1 Optical Sensing

Optical sensing techniques, including fluorescence spectroscopy, surface plasmon resonance (SPR), and optical coherence tomography (OCT), have shown promise in the detection of glioblastoma biomarkers in biofluids. These techniques offer high sensitivity and specificity, enabling the detection of low-abundance biomarkers such as ctDNA and exosomes. For instance, SPR-based biosensors can be functionalized with antibodies or aptamers to specifically capture glioblastoma-derived exosomes, allowing for real-time monitoring of tumor dynamics [Li et al., 2021, Ma et al., 2022].

2.6.2 Electrochemical Sensing

Electrochemical sensing involves the detection of biochemical interactions through electrical signals. Techniques such as electrochemical impedance spectroscopy (EIS) and voltammetry have been employed to detect glioblastoma biomarkers with high sensitivity. Electrochemical sensors can be integrated with microfluidic platforms to enable the rapid and cost-effective analysis of biofluids. For example, EIS sensors functionalized with glioblastoma-specific aptamers have demonstrated the ability to detect exosomes with high specificity and sensitivity, providing a promising approach

for liquid biopsy applications [Guo et al., 2019, Zhang et al., 2020].

2.6.3 Other Sensing Techniques

In addition to optical and electrochemical sensing, several other advanced techniques are being explored for the detection of glioblastoma and other brain tumors. These include mass spectrometry, magnetic resonance, and nanopore sequencing.

Mass Spectrometry

Mass spectrometry (MS) is a powerful analytical technique that measures the mass-to-charge ratio of ions to identify and quantify molecules in complex samples. In cancer diagnostics, MS is used for proteomic and metabolomic profiling of exosomes, providing detailed information about their molecular composition. This technique can detect specific protein markers and metabolic changes associated with glioblastoma, aiding in the early diagnosis and monitoring of tumor progression [Shao et al., 2021, Zhang et al., 2020].

Exosome-derived proteins and metabolites can be isolated from biofluids such as blood and cerebrospinal fluid (CSF) and analyzed using MS. For example, glioblastoma-specific protein markers such as EGFRvIII and IDH1 mutations have been identified in exosomal samples using this technique, demonstrating its potential for non-invasive cancer diagnostics [Shao et al., 2021].

Magnetic Resonance Techniques

Magnetic resonance techniques, including nuclear magnetic resonance (NMR) and magnetic resonance imaging (MRI), are widely used in medical diagnostics. These techniques exploit the magnetic properties of atomic nuclei to generate detailed images or spectra of tissues and biofluids.

Nuclear Magnetic Resonance (NMR): NMR spectroscopy provides information about the molecular structure and composition of exosomes by detecting the magnetic properties of atomic nuclei. This technique can be used to identify specific lipid and protein signatures in exosomes derived from glioblastoma patients, offering

insights into tumor metabolism and microenvironment [Belderbos et al., 2014, Sitter et al., 2009].

Magnetic Resonance Imaging (MRI): MRI is commonly used to visualize brain tumors in clinical settings. Advanced MRI techniques, such as diffusion-weighted imaging (DWI) and dynamic contrast-enhanced imaging (DCE-MRI), can be combined with exosome analysis to improve the specificity and sensitivity of glioblastoma detection. Exosomes labeled with magnetic nanoparticles can enhance MRI contrast, allowing for better tumor margins visualization and treatment response assessment [Xu et al., 2019].

Nanopore Sequencing

Nanopore sequencing is an emerging technique that allows for the direct analysis of nucleic acids by passing them through a nanopore and measuring changes in electrical conductivity. This technology enables real-time sequencing of DNA and RNA, providing rapid and accurate genetic information.

In glioblastoma diagnostics, nanopore sequencing can be used to detect specific genetic mutations and epigenetic modifications in exosomal DNA and RNA. This technique offers the advantage of long-read sequencing, which is particularly useful for identifying structural variations and complex genomic rearrangements associated with brain tumors [Wang et al., 2022, Brinkerhoff et al., 2019].

2.6.4 Fluorescence-Based Techniques

Fluorescence-based techniques utilize the emission of light by fluorophores to detect and quantify biomolecules in biological samples. These techniques are widely used for their high sensitivity and specificity.

Fluorescence-Activated Cell Sorting (FACS): FACS is a flow cytometry technique that sorts exosomes based on their fluorescence properties. By labeling exosomes with fluorescent antibodies targeting specific surface markers, FACS can isolate and quantify glioblastoma-derived exosomes from biofluids [Gyorgy et al., 2011].

Quantum Dots (QDs): Quantum dots are semiconductor nanoparticles that exhibit unique optical properties, such as size-tunable fluorescence emission. QDs can be conjugated to antibodies or aptamers specific to glioblastoma markers and used to label exosomes for fluorescence imaging and detection [Michalet et al., 2005].

Surface Acoustic Wave (SAW) Sensors

Surface Acoustic Wave (SAW) sensors are devices that use acoustic waves to detect changes in the properties of a surface. These sensors are highly sensitive to mass changes and can be used to detect biomolecules at low concentrations.

SAW sensors can be functionalized with antibodies or aptamers specific to glioblastoma biomarkers. When exosomes bind to the sensor surface, they cause a shift in the acoustic wave properties, which can be measured to quantify exosome concentration. This technique offers a rapid and label-free method for exosome detection [Luchansky and Bailey, 2012].

Photonic Crystal (PC) Sensors

Photonic Crystal (PC) sensors use the principles of photonic bandgap materials to detect changes in the refractive index of a surface. These sensors can achieve high sensitivity and specificity for biomolecule detection.

PC sensors can be integrated with microfluidic platforms to detect glioblastoma-derived exosomes in biofluids. By immobilizing antibodies or aptamers specific to exosomal markers on the PC surface, these sensors can measure changes in the refractive index upon exosome binding, providing a sensitive and specific detection method [Zhu et al., 2017].

Plasmonic Sensors

Plasmonic sensors exploit the resonant oscillation of conduction electrons at the surface of metallic nanoparticles or nanostructures to enhance the detection of biomolecules.

Surface Plasmon Resonance (SPR): SPR sensors are widely used for real-time, label-free detection of biomolecular interactions. By functionalizing the sensor surface with antibodies or aptamers specific to glioblastoma markers, SPR can be used to detect and quantify exosomes in biofluids with high sensitivity [Homola, 2008].

Localized Surface Plasmon Resonance (LSPR): LSPR sensors utilize the localized resonance of metallic nanoparticles to enhance the detection of exosomes. These sensors can achieve high sensitivity due to the strong electromagnetic field enhancement at the nanoparticle surface, making them suitable for detecting low-abundance exosomal biomarkers [Jiang et al., 2020].

Microfluidic Devices

Microfluidic devices integrate multiple laboratory functions on a single chip, allowing for the precise manipulation of small volumes of fluids. These devices offer high-throughput and automated analysis, making them ideal for exosome detection and characterization.

Microfluidic devices can be designed to isolate and analyze exosomes from biofluids using various principles such as size exclusion, immunoaffinity capture, and acoustic trapping. These devices can be coupled with other sensing techniques, such as fluorescence detection, SERS, or electrochemical sensing, to provide a comprehensive platform for glioblastoma diagnostics [Zhang et al., 2020, He et al., 2018].

Integration with SERS: By integrating microfluidic platforms with SERS, it is possible to achieve rapid and sensitive detection of exosomal biomarkers. Microfluidic devices can precisely control the flow and mixing of samples, enhancing the reproducibility and sensitivity of SERS measurements [Zhu et al., 2019].

2.7 Conclusion

Liquid biopsy offers a promising non-invasive alternative for the detection and monitoring of brain tumors. Exosomes, with their rich molecular cargo, provide valuable insights into the tumor microenvironment. The integration of SERS and AI further

enhances the potential of liquid biopsy by enabling sensitive detection and accurate classification of tumor-derived biomarkers.

The development and integration of various sensing techniques, including mass spectrometry, magnetic resonance, nanopore sequencing, fluorescence-based methods, SAW sensors, PC sensors, plasmonic sensors, and microfluidic devices, offer promising approaches for the detection and monitoring of glioblastoma. These advanced techniques provide high sensitivity, specificity, and the potential for non-invasive diagnostics, contributing to improved patient outcomes and personalized medicine in the clinical management of brain tumors.

Addressing current challenges through ongoing research and technological advancements will be crucial for the successful clinical implementation of these innovative diagnostic techniques.

Chapter 3

MATERIALS AND METHODS

3.1 Sample Collection

3.1.1 Sample Size Calculation

Biostatistician from Canary Center for Cancer Early Detection carried out sample size calculation for this study as follows: If we aim to demonstrate predictive sensitivity and specificity of 90% using 20 glioma and 20 non-glioma cases, the non-glioma cases would typically include individuals who do not have gliomas but may have other conditions or be healthy controls. Specifically, the non-glioma group could include:

- Healthy Controls: Individuals with no brain tumors or neurological conditions.
- Patients with Other Brain Tumors: Individuals with benign brain tumors, such as meningiomas or pituitary adenomas.
- Patients with Other Neurological Conditions: Individuals with non-tumor-related neurological conditions, such as multiple sclerosis or stroke. By including these diverse non-glioma cases, the study can robustly demonstrate the predictive sensitivity and specificity of the diagnostic method for glioma detection.

Therefore, we used 23 meningioma samples as 'patients with other brain tumors' and 30 healthy control samples as control groups, besides the disease group of 20 glioblastoma samples.

3.1.2 Pathological Evaluation and Diagnosis

Pathological evaluation of tumor tissues is performed by Koc University Pathologists according to established and validated standard operating protocols.

3.1.3 IRB Approval

Blood samples were collected under the approval of (2019.328.IRB2.106) Koç University Institutional Review Board with oral and written consent of the patient and/or legal guardians.

3.1.4 Material Transfer Agreement (MTA)

This study was conducted under the ‘Material Transfer Agreement (MTA)’ between Koc University Research Center for Translational Medicine (KUTTAM) and Canary Center at Stanford for Cancer Early Diagnosis at the Department of Radiology, Stanford University.

3.1.5 Study Design

Blood samples from 20 glioblastoma (GB) and 23 meningioma (MNG) patients were collected in EDTA tubes prior to tumor resection. Recurrent cases and/or cases with radiotherapy were not included in this study. The diagnosis of the cases was performed according to standard operating procedures by Koç University Department of Pathology upon the evaluation of tumor samples obtained from patients. Blood samples from 30 healthy control (HC) individuals were collected from volunteers without any known tumors or other malignancies.

3.2 Exosome Isolation from Plasma Samples

3.2.1 Plasma Processing

Blood samples were centrifuged at 2000 rpm and 4°C for 20 minutes, and the upper part of the supernatant (plasma) was collected without contacting the buffy coat.

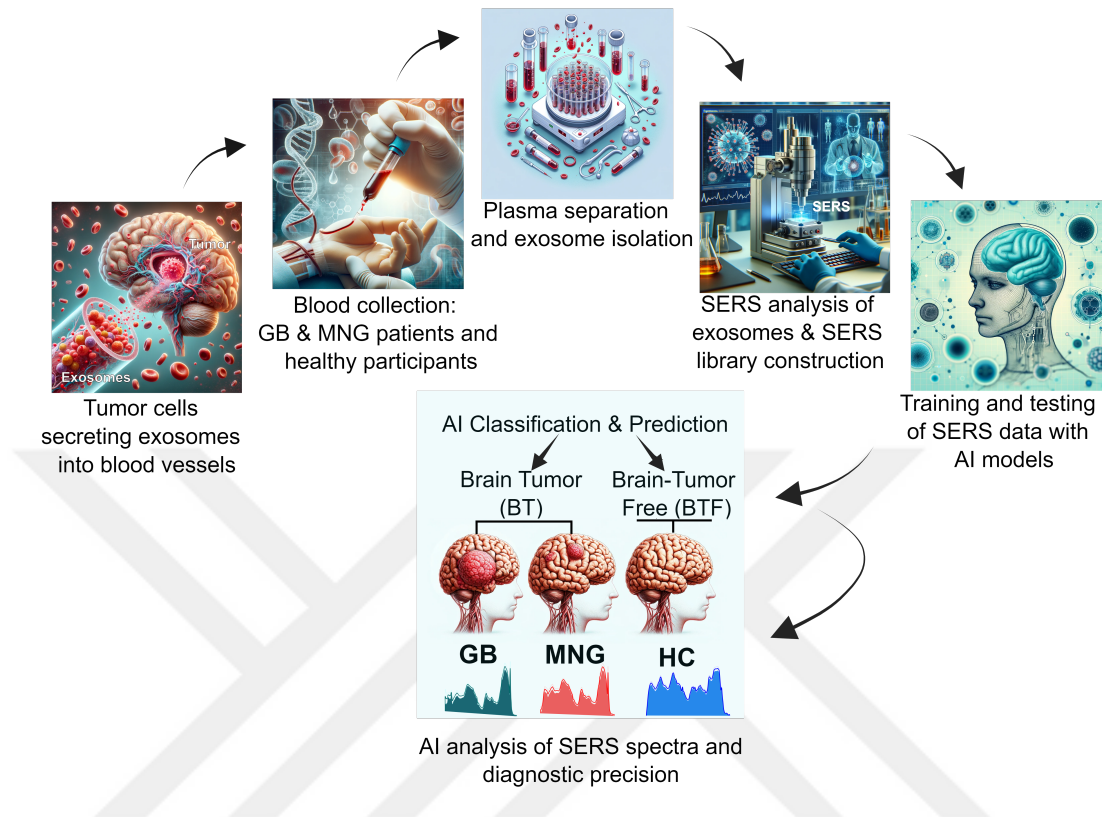


Figure 3.1: Schematic illustration of study design: Sample collection, plasma separation and exosome isolation, exosome characterization, Surface-Enhanced Raman Spectra (SERS) analysis and library construction, training and testing SERS data using Artificial Intelligence (AI) models, first step AI classification for brain tumor (BT) and brain tumor free (BTF) individuals, second step AI classification for predicting brain tumor type (GB or MNG). Created with openAI and BioRender.com

Plasma was aliquoted and kept at -80°C for further use. Plasma samples were thawed before exosome isolation.

3.2.2 Exosome Isolation

Exosomes were isolated using the Exosome Total Isolation Chip (EXOTIC).[Liu et al., 2017] The ExoTIC membrane was kept at 4°C overnight for release of exosomes from the collection membrane. Harvested exosomes in Phosphate Buffer Saline (PBS) and Ammonium Bicarbonate (AmBIC) for proteomics analysis were collected the next day from the ExoTIC membrane and stored at -80°C until further use.

3.3 Exosome Characterization

Exosomes were characterized using various methods including Transmission Electron Microscopy (TEM), Cryo-Electron Microscopy (Cryo-EM), Interferometric Scattering Microscopy (iSCAT), Nanoparticle Tracking Analysis (NTA), Bead-Captured Flow Cytometry (BC-FC), and Western Blotting (WB).

3.3.1 TEM Analysis of Exosomes

Exosomes were diluted to the optimal concentration (10x-100x) and underwent negative staining. 10 μ L of exosomes were cast onto previously glow-discharged grids and allowed to settle for 3 minutes. Staining was performed using uranyl acetate, and grids were left to dry before imaging with a JEOL JEM 1400 120 kV TEM microscope.

3.3.2 Cryo-TEM Analysis of Exosomes

High-resolution imaging of exosomes was performed using Cryo-EM. Samples were applied onto glow-discharged grids, blotted, plunge-frozen in liquid ethane, and examined with a ThermoFisher Scientific GlaciosTM Cryo-TEM.

3.3.3 NTA Analysis of Exosomes

Exosome size and concentration distribution were evaluated using a Nanosight NS300 (Malvern Panalytical) with NTA software 4.2. Samples were diluted (1000x) and analyzed with specific laser settings to ensure optimal concentration.

3.3.4 Western Blot Analysis of Exosomes

Western Blot analysis was performed using standard protocols with antibodies specific to Calnexin, HSPA8, CD63, TSG101, and Flot1, following MISEV 2018 and 2024 guidelines.[Théry et al., 2018, Welsh et al., 2024]

3.3.5 *iSCAT Analysis of Exosomes*

A custom interferometric scattering microscope was developed for exosome quality control. Exosome samples were dropped onto a SiO₂/Si thin film substrate, and interferometric images were captured to analyze the size distribution of the particles.

3.3.6 *Flow Cytometry Analysis of Exosomes*

Exosomes were diluted for optimal attachment with antibody and exosome-bead complex, and incubated with Aldehyde Sulfate Latex beads (ThermoFisher) for 15 minutes at room temperature. The exosome-bead complex was blocked with 2% BSA for 2 hours, followed by glycine incubation (100 mM) for 30 minutes. Samples were washed with cold PBS and incubated with primary antibodies (Calnexin, CD81, HSPA8) overnight. The next day, samples were washed and incubated with secondary antibodies (A28175, A-10931, 31860 from Invitrogen) for 2 hours. Flow cytometry analysis was performed using a Guava easyCyte flow cytometer and FlowJo v10 software. Controls included only exosomes, only beads, and beads with exosomes to identify the true exosome population. Multi-gating strategies identified CD81+ and HSPA8+ populations, with calnexin-negative exosomes representing pure exosome populations.

3.4 *Surface-Enhanced Raman Spectroscopy Measurements*

3.4.1 *Experimental Setup*

A customized Raman spectroscopy system was designed with a 785 nm diode laser, a motorized XYZ stage, and a Raman spectrometer equipped with a CCD camera to achieve enhanced flexibility for multi-modality and multi-functionality while maintaining sensitivity comparable to state-of-the-art instruments. The laser beam was tuned and directed to the sample plane of the microscope, and scattered photons were collected and analyzed. The diode laser output (CrystaLaser, 785 nm, maximum output power of 130 mW) was tuned using a half-wave plate and a linear polarizer to deliver the required power without causing damage to the highly sensi-

tive gold (Au) surface. The adjusted laser beam was precisely directed to the sample plane of the microscope, passing through a long pass dichroic mirror (Thorlabs, DMLP 805) and a microscope objective (10x, NA 0.22). To enable lateral scanning of the sample surface and achieve fine focal plane adjustment, we utilized a XYZ motorized stage (Standa). Scattered photons were collected using the same microscope objective with a 180-degree geometry, and elastically scattered photons (Rayleigh beams) were effectively filtered by the dichroic mirror. Inelastically-scattered photons (Stokes beams) were directed towards a multimode fiber through an ultra-steep long pass filter (Semrock) and a fiber collimator (Thorlabs, F220SMA-780). The fiber was connected to a Raman spectrometer (StellarNet, HyperNova) equipped with a CCD camera (Andor iVAC 316) cooled to -60°C.

3.4.2 Sample Preparation

Hydrophobic Au surfaces were used for SERS measurements. 2 μ L of exosome sample was dropped on the surface and allowed to dry for 15 minutes.

3.4.3 Data Collection

We employed the Andor MATLAB SDK to establish a connection between the PC and CCD of the spectrometer. Subsequently, a custom MATLAB code was developed to automate the surface scanning process with predefined parameters. The laser output power was fine-tuned to 30 mW (approximately 20 mW on the sample plane). Real-time spectral feedback was utilized to identify the optimal focus. Following the focus optimization, we set the exposure time to one second, the scanning step to 2.5 μ m, and the number of steps to 21 along each axis (X-Y). This configuration resulted in the acquisition of 441 spectra for each sample. A custom MATLAB code was used to automate the surface scanning process. Laser output power was set to 30 mW, with exposure time and scanning steps optimized for data acquisition. 441 spectra (21x21 map area) were collected for each sample.

3.4.4 Data Analysis

Pre-processing of the spectra included cosmic ray removal, smoothing, baseline correction, and normalization. Dimensionality reduction was performed using t-SNE, and various mathematical, machine learning, and deep learning models (CNN, NN, RF, SVC, QDA, XG-Boost) were used for classification. We used RamanSPY Python library for AI classification. [Georgiev et al., 2024]

3.5 Statistical Analysis

Statistical analysis was performed using the Kolmogorov-Smirnov test and Kruskal-Wallis test to assess significant differences in Raman intensity distributions among different sample groups. We conducted a Kolmogorov-Smirnov test on selected intense Raman peaks around 456, 587, 715, 761, 919, 1069, 1219, 1254, 1304, and 1422 cm^{-1} , revealing a non-normal distribution. Consequently, we opted for a non-parametric approach in our statistical analysis. Given the presence of three distinct sample groups in our study (healthy control, meningioma, and glioblastoma), we employed a Kruskal-Wallis test to assess the selected peaks. Our findings indicate significant differences ($p < 0.001$) in the Raman intensity distributions among these groups compared to others. We show the distribution of the spectral data using violin plots for the selected peaks in violin figure.

3.6 Limit of Detection (LOD) and Limit of Quantification (LOQ) Assessment for SERS Biosensor

The limit of detection (LOD) and limit of quantification (LOQ) for a Raman spectrometer were established using EGFR solutions with concentrations ranging from 10^{-9} to 10^{-18} mg/mL . The spectra were arranged in an $m \times n$ matrix, where m denotes the wavenumbers and n represents the samples. The intensity at 1550 cm^{-1} was identified for all samples. Both concentration and intensity values were log-transformed, followed by a linear regression analysis with log-transformed intensity on the x-axis and log-transformed concentration on the y-axis. Residuals were com-

puted as the differences between the observed and predicted log concentrations, and their standard deviation was calculated. LOD and LOQ were determined by multiplying the standard deviation of the residuals by factors of 3 and 10, respectively, and adding these values to the regression intercept. The results were visualized by plotting log-transformed intensity against log-transformed concentration, with vertical lines indicating the LOD and LOQ. This approach ensured precise determination and visualization of the LOD and LOQ for the Raman spectrometer in EGFR (Recombinant Human EGFR Protein (ECD, His Tag), HPLC-verified, SinoBiological, 10001-H08H) quantification.

To express the LOD in terms of molarity, we start with a molecular mass of 69.8 kDa (69,800 g/mol) and a starting solution concentration of 1 mg/mL (0.001 g/mL). The concentration in molarity is calculated as follows:

$$\text{Concentration in molarity} = \frac{0.001 \text{ g/mL}}{69800 \text{ g/mol}} = 1.43 \times 10^{-8} \text{ mol/L}$$

For a LOD at 10^{-18} mg/mL, the LOD in molarity is:

$$\text{LOD in molarity} = \frac{10^{-18} \times 10^{-3}}{69800} = 1.43 \times 10^{-26} \text{ mol/L}$$

3.7 Regression Analysis

A dataset of EGFR preparations was created, and linear models were validated using k-fold cross-validation. The presence of selected amino acids in the SERS spectra of EVs was confirmed, and RMSE values were calculated.

3.8 Proteomics Analysis

3.8.1 Exosome Isolation and Protein Extraction

Exosomes were isolated as mentioned in the section "Exosome Isolation from Plasma samples", and exosomal proteins were exposed by breaking down the lipid bilayer

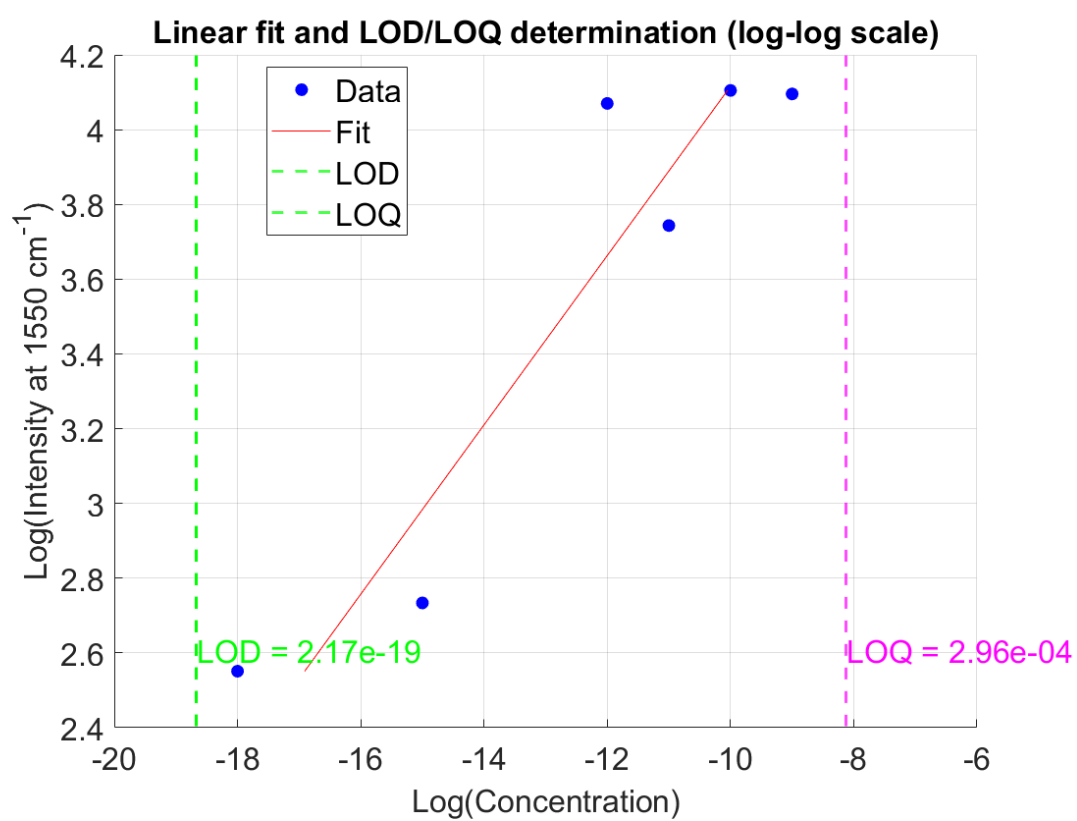


Figure 3.2: Limit of detection (LOD) and limit of quantification (LOQ) of SERS sensor used for exosome and protein (EGFR) detection analysis

using 2% Sodium Dodecyl Sulfate (SDS) before proteomic analysis. Protein extraction and mass spectrometry analysis steps, which are well-established in our lab and the Pitteri lab, were employed. The protein samples were reduced using 10 mM tris-carboxyethyl phosphine (TCEP) and alkylated with 15 mM iodoacetamide, followed by overnight acetone precipitation. Subsequently, the samples were digested using trypsin, and the resulting samples were kept at -20°C until LC-MS/MS analysis.

3.8.2 LC/MS analysis

Four micrograms of tryptic peptides were loaded on an Acclaim PepMap C18 trap column (Thermo Fisher Scientific) coupled to a Dionex Ultimate Rapid Separation Liquid Chromatography HPLC system (Thermo Fisher Scientific) at a rate of 5 μ L/min for 10 minutes. Tryptic peptides were separated by reversed-phase chromatography on a 25 cm long C18 analytical column (New Objective) packed in house with ReproSil-Pur 120 C18 AQ resin (Dr. Maisch GmbH). Eluted peptides were ionized using a Nanospray flex ion source (Thermo Fisher Scientific) with a 1.8 kV voltage coupled to a LTQ-Orbitrap Elite mass spectrometer (Thermo Fisher Scientific). The chromatography gradient program was as follows: A flow rate of 0.4 μ L/min was used throughout the gradient using mobile phase A (consisting of 0.1% formic acid in water) and mobile phase B (0.1% formic acid in acetonitrile) starting at 98% and 2% B respectively for the first 10 minutes, slowly ramped up to 35% B over 100 minutes, followed by an increase to 85% B over 2 minutes with a 7 minute hold. The analytical column was re-equilibrated prior to the next sample injection. Each sample was analyzed in triplicates. The top 10 most abundant ions per MS1 scan were selected for higher energy collision-induced dissociation (35 eV) in a data-dependent fashion. MS1 resolution was set at 60,000, FT AGC target was set at 1e6, and the m/z scan range was set from m/z 400-1800. MS2 AGC target at 3e4, and dynamic exclusion was enabled for 30 seconds.

Chapter 4

EXOSOME ISOLATION AND CHARACTERIZATION

4.1 Nanoparticle Tracking Analysis (NTA)

Exosome size and concentration distribution were evaluated using Nanoparticle Tracking Analysis (NTA) with a Nanosight NS300 instrument (Malvern Panalytical). The NTA software 4.2 was employed for analysis. We used 405 nm (55mW) violet and 532 nm green lasers (50 mW) with 8-10 camera levels and 5 detection thresholds. Exosome samples were diluted to an optimal concentration (100x-1000x) to ensure accurate measurements. NTA allows for the visualization and analysis of particles in liquid suspension, providing detailed information on the size distribution and concentration of exosomes. This technique is crucial for confirming the presence of exosomes and assessing their size and concentration before further characterization and functional studies.

4.2 Room Temperature Transmission Electron Microscopy (TEM)

Exosomes were visualized using Room Temperature Transmission Electron Microscopy (TEM). Samples were diluted to the optimal concentration (10x-100x) and subjected to negative staining. 10 μ L of exosome suspension was placed on previously glow-discharged grids (Electron Microscopy Sciences, FCF-300-CU) and allowed to settle for 3 minutes. Grids were then stained with 1% uranyl acetate for 1 minute. Excess stain was removed with filter paper, followed by three washes with distilled water to remove residual stain. Grids were dried for approximately 10 minutes before imaging. Imaging was performed using a JEOL JEM 1400 120 kV TEM microscope equipped with a Gatan Ultra scan digital high-resolution camera. This technique provides detailed morphological information and confirms the presence of exosomes.

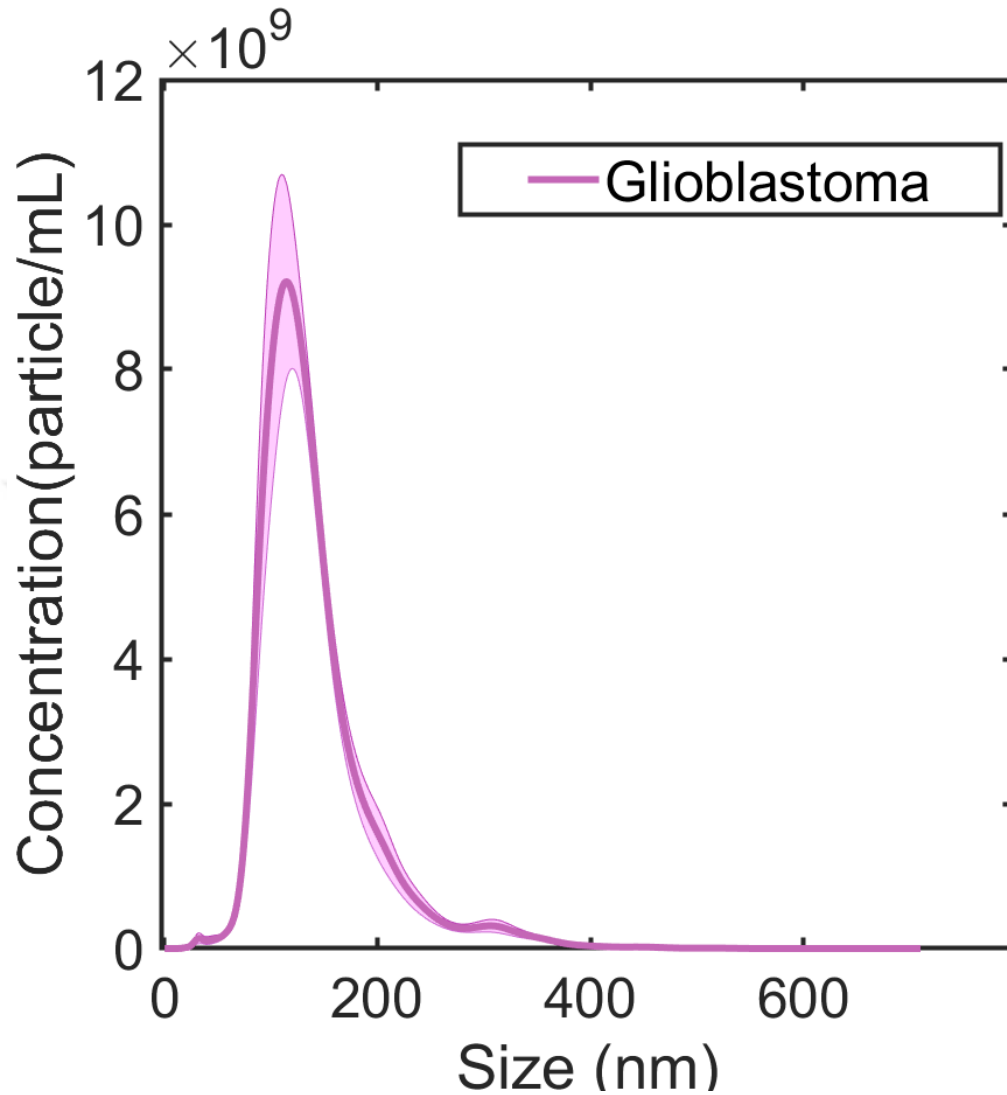


Figure 4.1: Nanosight Tracking analysis (NTA) of plasma exosomes derived from glioblastoma (GB) patients.

4.3 Cryo-Transmission Electron Microscopy (Cryo-TEM)

For high-resolution imaging, Cryo-Transmission Electron Microscopy (Cryo-TEM) was employed. Quantifoil Holey Carbon 1.2/1.3 Cu 200 mesh grids were glow discharged using PELCO easiGlow™ (TED PELLA Inc.) for 45 seconds at 15 mA. 3 μ L of the sample was applied to the grids, blotted with Vitrobot Mark IV (FEI) for 3 seconds at blot force 3, and quickly plunge-frozen in liquid ethane. Grids were then loaded into a ThermoFisher Scientific Glacios™ Cryo-TEM operating at 200 kV.

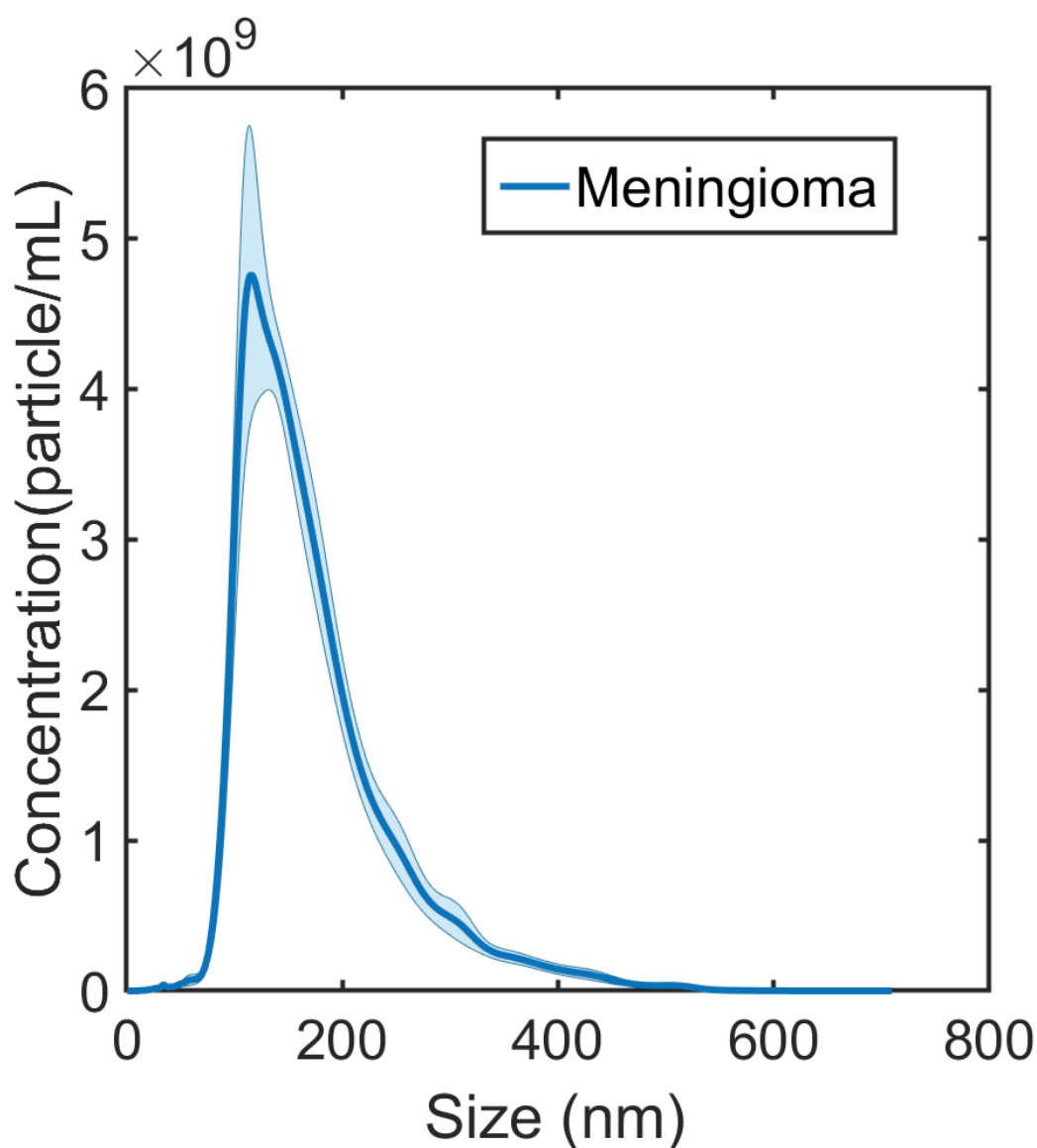


Figure 4.2: Nanosight Tracking analysis (NTA) of plasma exosomes derived from meningioma (MNG) patients.

Specimens were examined at 36,000x magnification with a pixel size of 1.15 Å on a K3 direct electron camera (Gatan) using SerialEM software. The total accumulated dose did not exceed 60 e-/Å² with a defocus range of -1.5 to -3 microns. Cryo-TEM offers high-resolution structural information and preserves the native state of the exosomes.

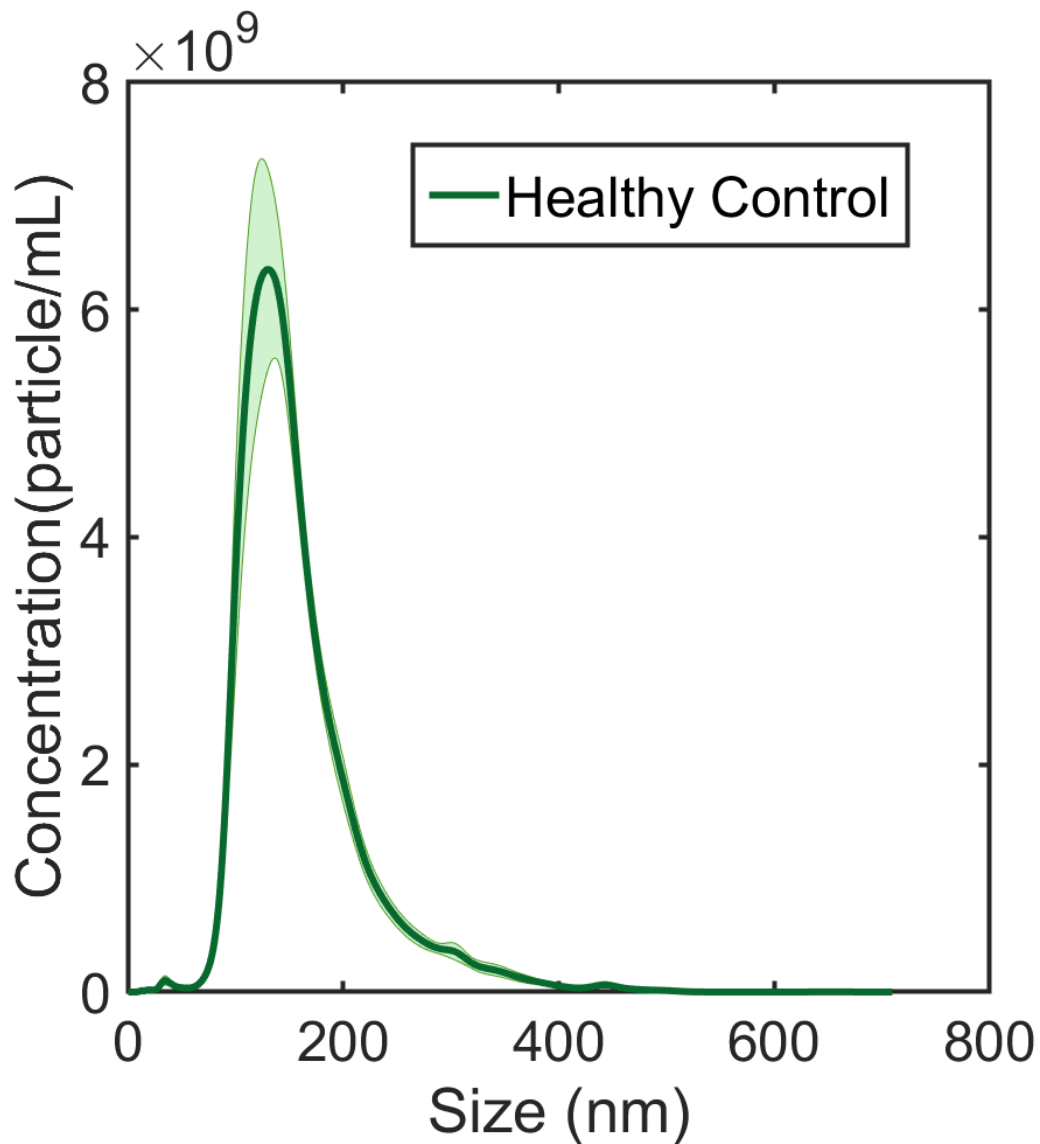


Figure 4.3: Nanosight Tracking analysis (NTA) of plasma exosomes derived from healthy control (HC) patients.

4.4 Interferometric Scattering Microscopy (iSCAT)

Interferometric Scattering Microscopy (iSCAT) was utilized for exosome quality control. A custom iSCAT microscope was developed, using a thin film substrate composed of SiO₂/Si. An LED (Thorlabs M53OL4 530 nm) provided homogeneous illumination, focused on the back focal plane of the objective lens. Scattered light from exosomes and reflected light from the sample were captured and imaged. Exosome

samples were placed on a motorized stage (Standa) to adjust the focus. Samples were dropped onto the substrate and allowed to dry for 15 minutes. iSCAT images were collected within a 200 μm x 200 μm field of view, and particle size distribution was analyzed using custom MATLAB scripts. This technique provides sensitive detection and sizing of individual exosomes.

4.5 Bead-Assisted Flow Cytometry

Exosomes were incubated with Aldehyde Sulfate Latex beads and primary antibodies (Calnexin, CD81, HSPA8), followed by secondary antibodies. Flow cytometry analysis was performed using a Guava easyCyte flow cytometer, with data analyzed using FlowJo v10 software.

Exosomes were diluted to obtain the optimal attachment with antibody and exosome-bead complex and incubated with Aldehyde Sulfate Latex beads (ThermoFisher) for 15 minutes at RT on a roller. Exosome-bead complex was blocked using %2 BSA incubation for 2 hours, followed by a glycine incubation (100mM) for half an hour. The exosome-bead complex was washed with cold PBS to wash the exosomes and remove the unbound beads and exosomes from the solution. The samples were incubated with primary antibodies Calnexin (PA1-30197, Invitrogen), CD81 (SAB4700232, Sigma Aldrich), and HSPA8 (NB100-41377, Novus Biologicals) overnight. The next day, samples were centrifuged and washed with cold PBS to remove the unbound antibodies. Then, the samples were incubated with secondary antibodies A28175, A-10931, and 31860 (Invitrogen) for two hours. The samples were centrifuged and washed with cold PBS before flow cytometry analysis. Flow Cytometry analysis was performed with Guava easyCyte flow cytometer using FlowJo v10 software for the data analysis. We used controls including only exosomes, only beads, and beads with exosomes to identify the truest exosome population as described in MISEV 2018. We first gated the calnexin negative (Calnexin-) exosomes to get the pure exosome population, then applied multi-gating strategy for identifying CD81+ and HSPA+, double positive and double negative populations for these antibodies. Calnexin- and CD81+ population is stained 95.4%, Calnexin- and

HSPA8+ population stained 65.3%, triple negative (Calnexin-, CD81-, and HSPA8-) population stained 2.5%, Calnexin- and double positive CD81+ and HSPA8+ population stained 68.9%.

4.6 Western Blotting

Western Blotting was used to analyze exosomal protein markers. Exosome samples were lysed and proteins were separated by SDS-PAGE. Proteins were transferred to PVDF membranes and blocked with 5% non-fat dry milk. Membranes were incubated with primary antibodies specific to Calnexin (Invitrogen, PA1-30197), HSPA8 (Novus Bio, 41377), CD63 (R&D Systems, MAB5048), TSG101 (Novus Bio, NB200-112), and Flot1 (Abclonal, A6220) following MISEV 2018 guidelines [Théry et al., 2018]. After washing, membranes were incubated with HRP-conjugated secondary antibodies (Invitrogen, 31460, 31430, 81-1620) and visualized using ECL substrate. Bands were detected using a chemiluminescence imaging system, confirming the presence of exosomal markers and validating the isolation procedure.

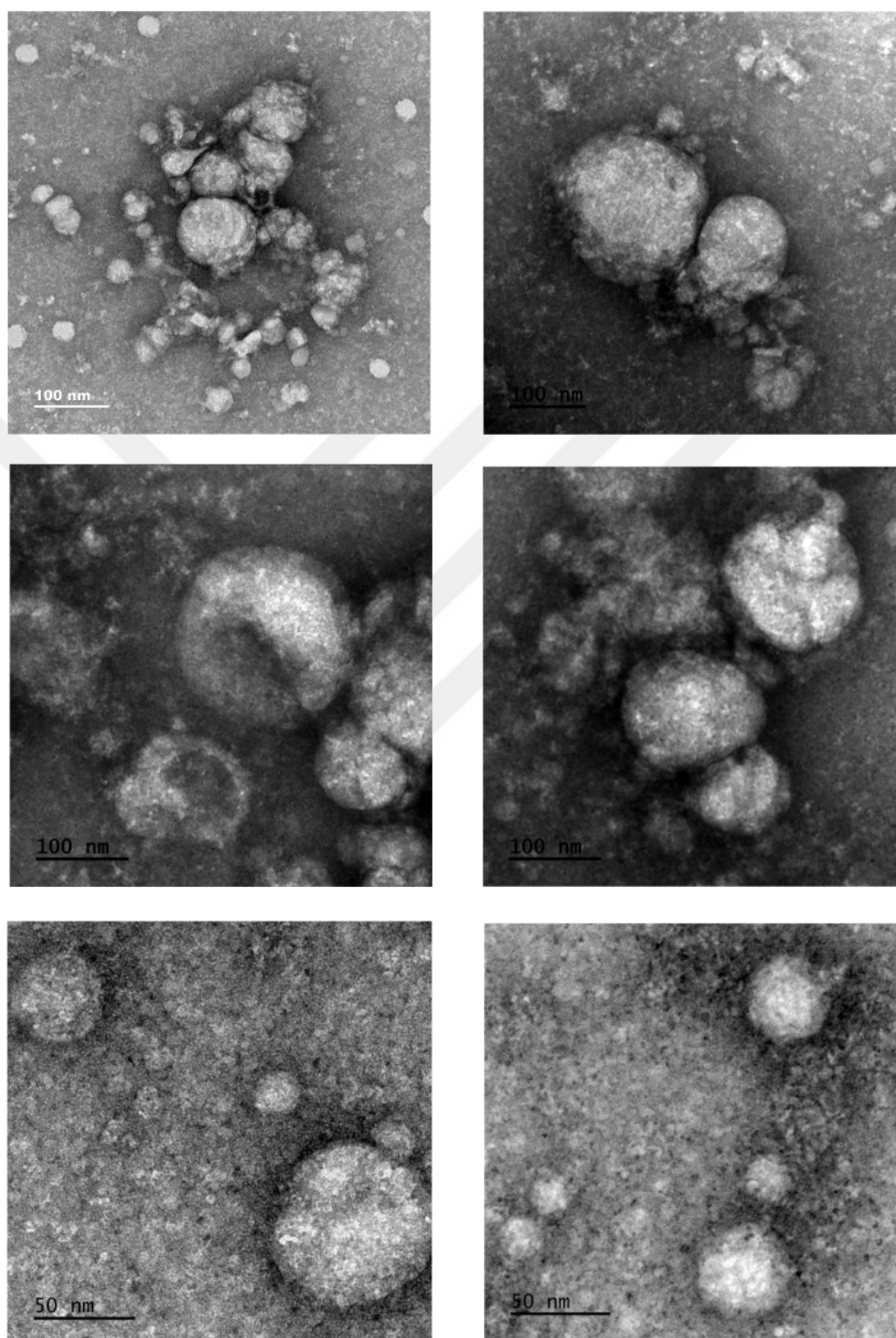


Figure 4.4: Transmission Electron Microscopy (TEM) imaging of plasma exosomes derived from glioblastoma (GB) patients. Exosomes were observed in between 30-200 nm in size.

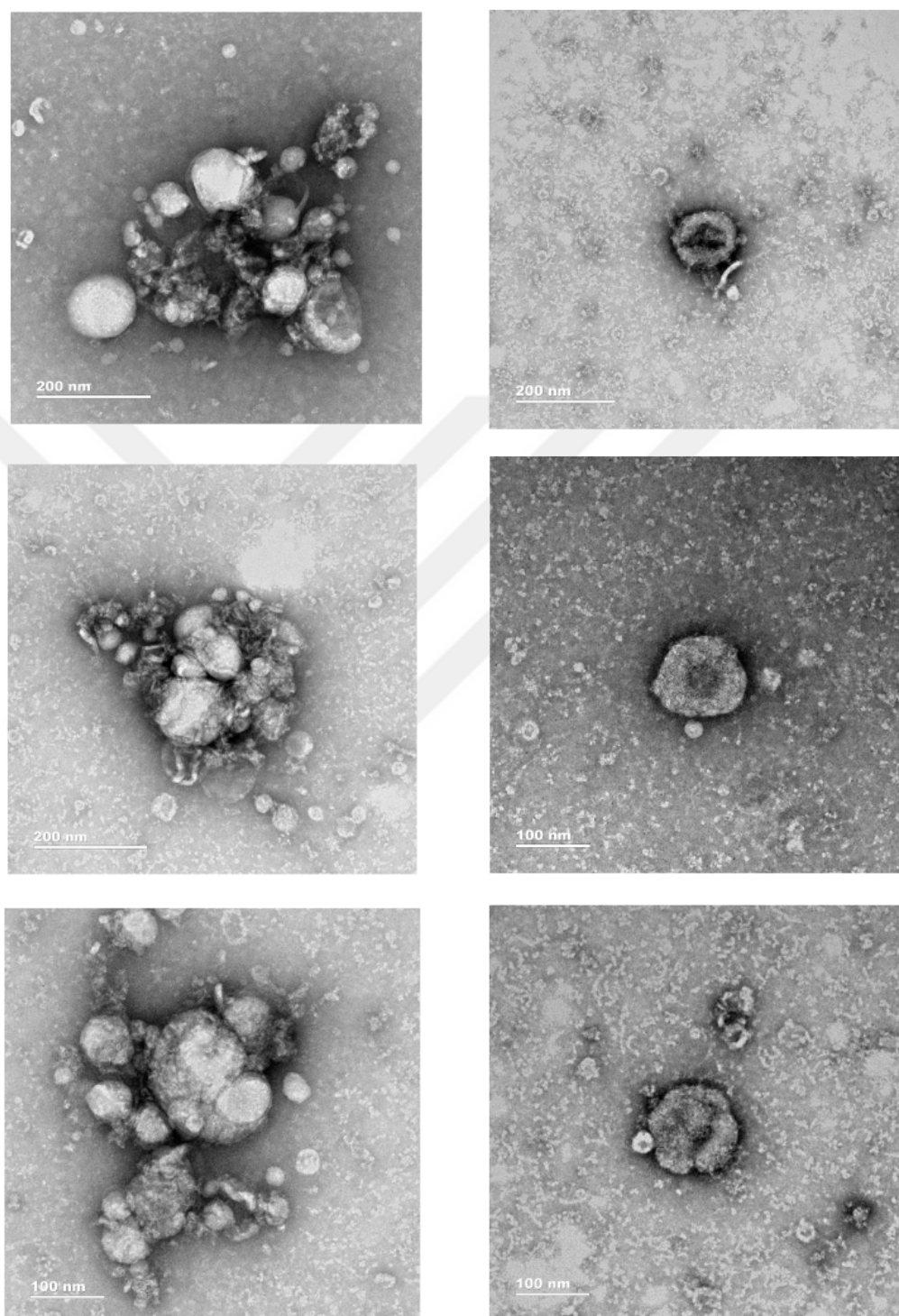


Figure 4.5: Transmission Electron Microscopy (TEM) imaging of plasma exosomes derived from meningioma (MNG) patients. Exosomes were observed in between 30-200 nm in size.

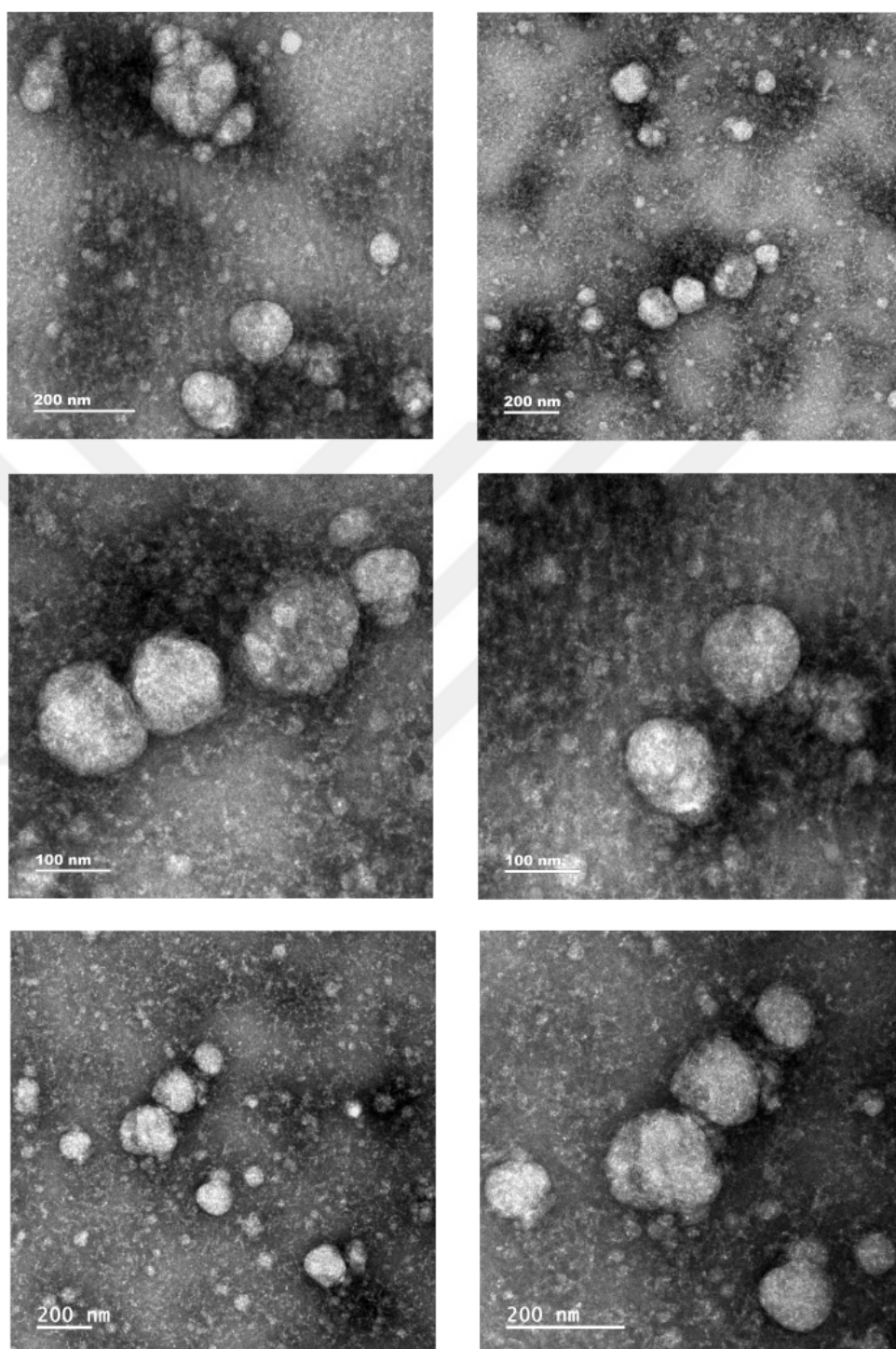


Figure 4.6: Transmission Electron Microscopy (TEM) imaging of plasma exosomes derived from healthy control (HC) patients. Exosomes were observed in between 30-200 nm in size.

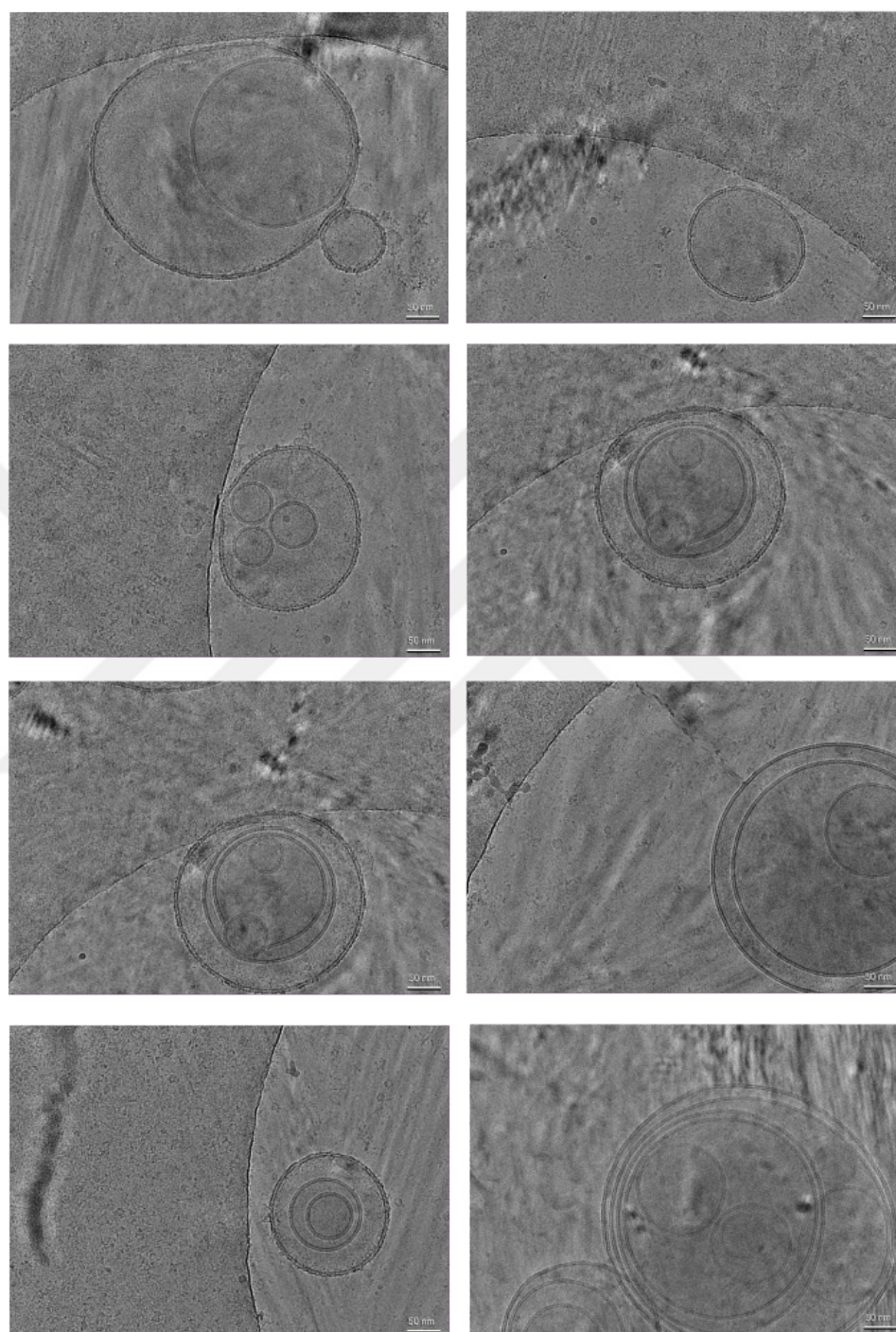


Figure 4.7: Cryo-Electron Microscopy (Cryo-EM) images of plasma exosomes derived from glioblastoma patients-1. Single, multi-laminar, and spherical vesicles.

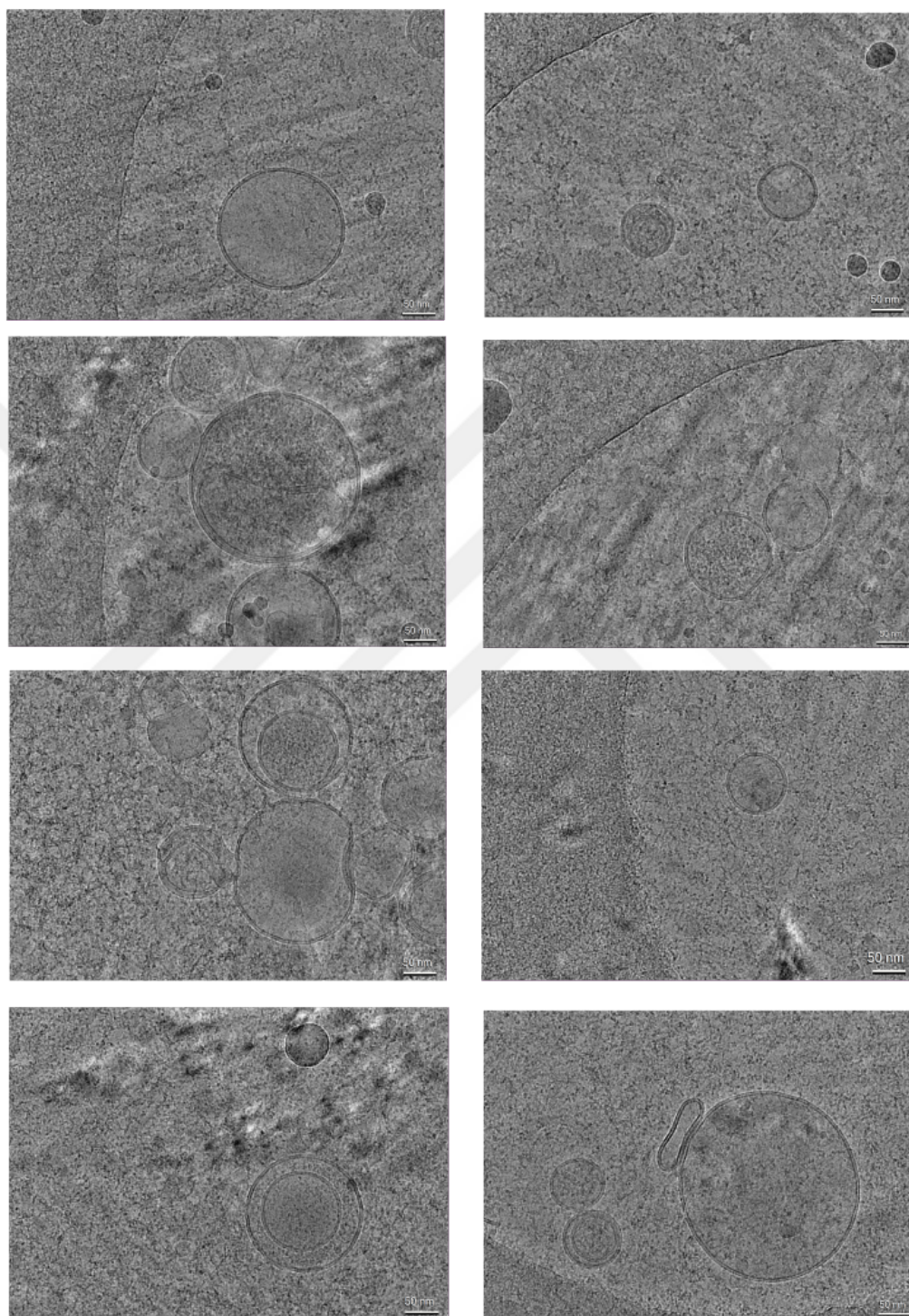


Figure 4.8: Cryo-Electron Microscopy (Cryo-EM) images of plasma exosomes derived from meningioma patients-1. Single, multi-lamellar, spherical and cylindrical vesicles.

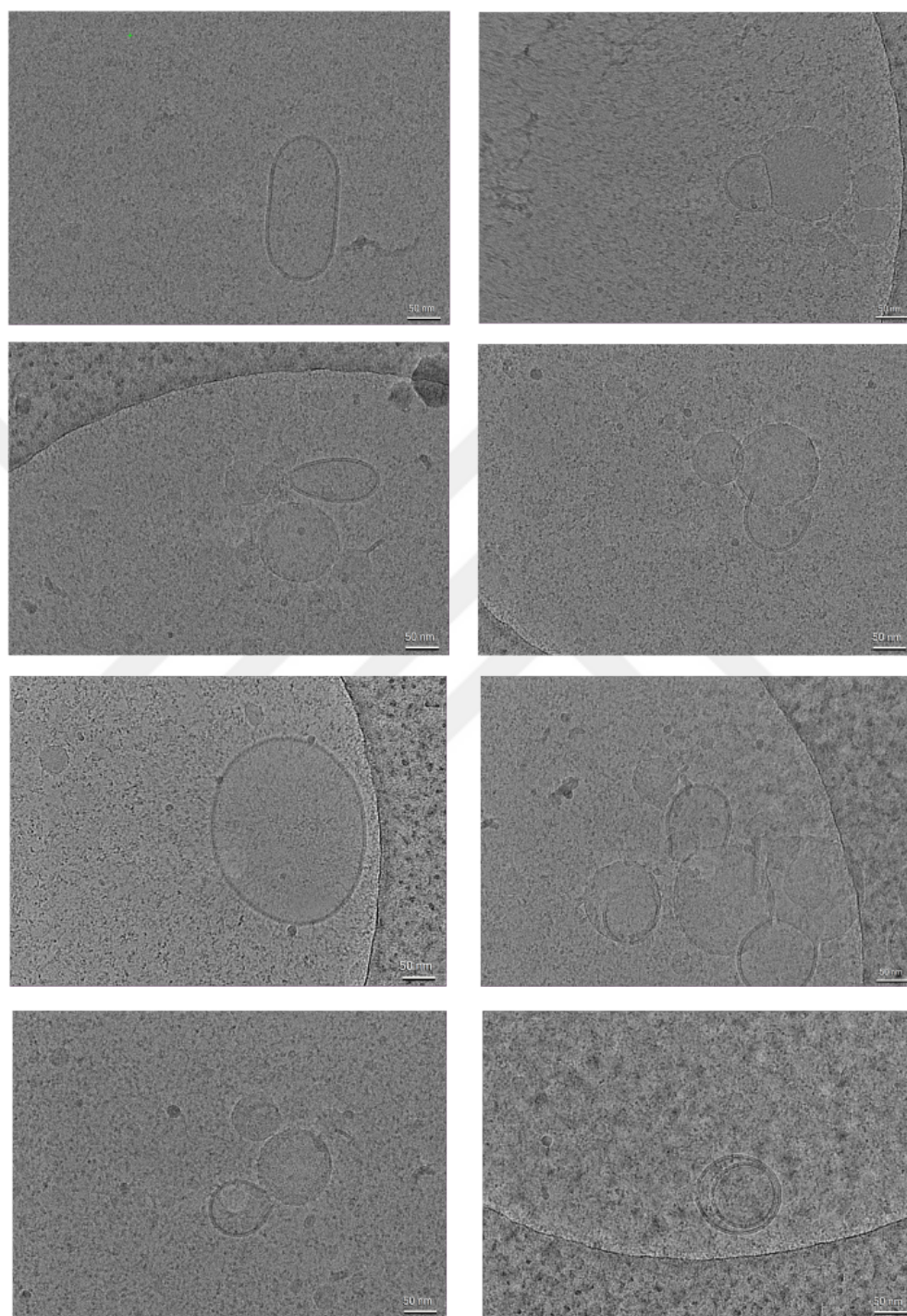


Figure 4.9: Cryo-Electron Microscopy (Cryo-EM) images of plasma exosomes derived from healthy control participants. Single, multi-lamellar, spherical and cylindrical vesicles.

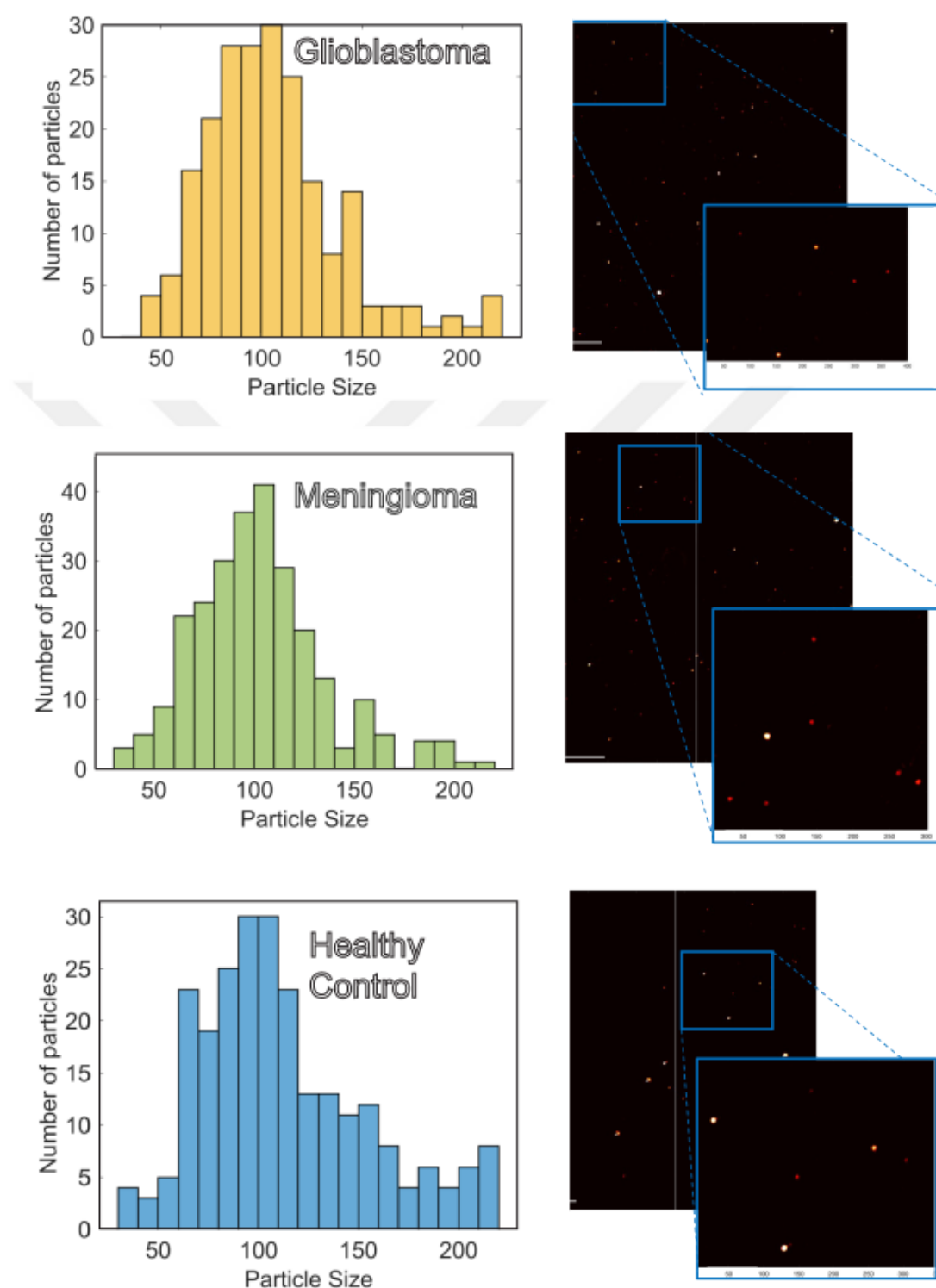


Figure 4.10: Interferometric Scattering Microscopy (iSCAT) images of glioblastoma, meningioma, and healthy control plasma exosome samples and their size distributions.

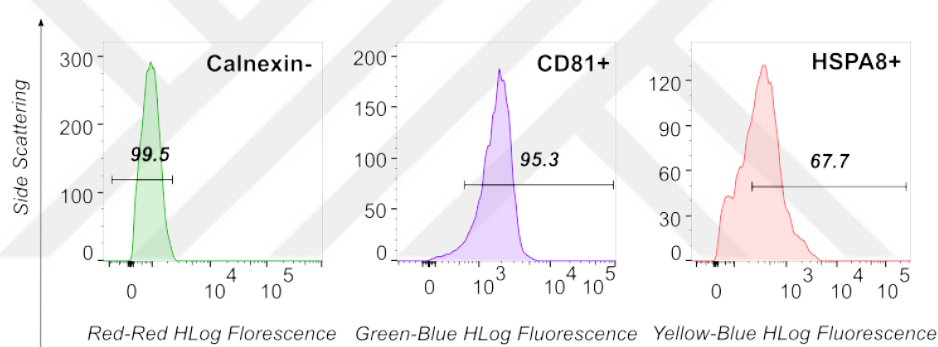


Figure 4.11: Histograms of glioblastoma patient plasma exosome characterization via Bead-Assisted Flow Cytometry (BA-FC). Exosomes are validated using Calnexin, CD81, and HSPA8 antibodies.

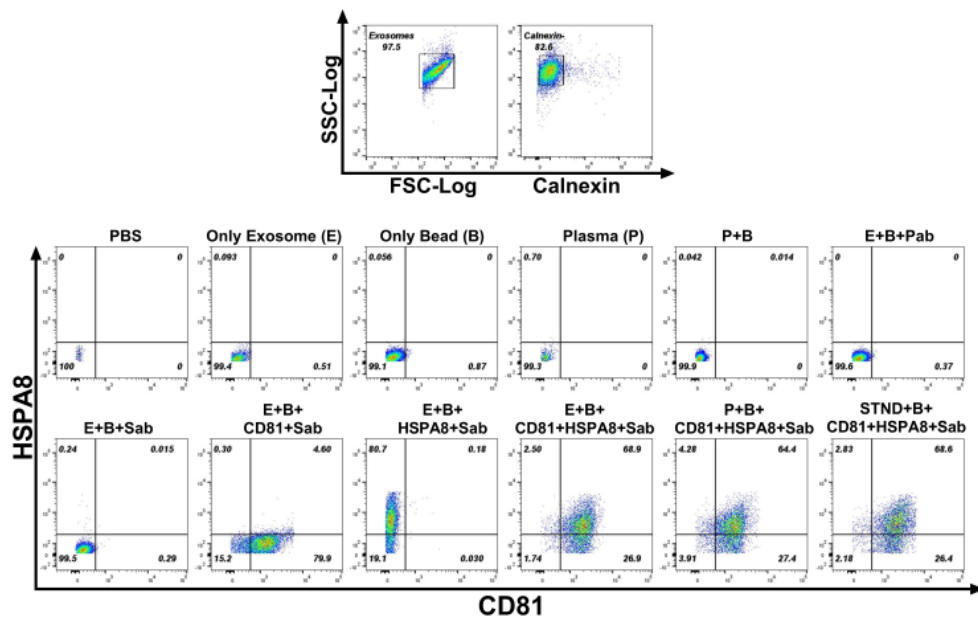


Figure 4.12: Glioblastoma patient plasma xosome characterization via Bead-Assisted Flow Cytometry (BA-FC) showing the control and experimental groups used. Exosomes are validated using Calnexin, CD81, and HSPA8 antibodies. The first figure shows all particles and the Calnexin-negative (Cellular content contamination-free) population using APC (Red-Red) Secondary antibody. In the second figure, 12 analysis performed: Only PBS, only exosomes without bead (E), only beads (B), only plasma (P), exosome (E) + bead (B) + Primary Antibody (Pab), Exosome + Bead + Secondary Antibody (E+B+Sab), E+B+CD81+Sab (Alexa 488, Green-Blue), E+B+HSPA8+Sab (PE, Yellow-Blue), E+B+CD81+HSPA8+Sab (Alexa 488 and PE), P+B+CD81+HSPA8+Sab (Alexa 488 and PE), and Exosome standards (STND: Novus Biologicals, NBP3-11686) +B+CD81+HSPA8+Sab (Alexa 488 and PE) groups were analyzed.

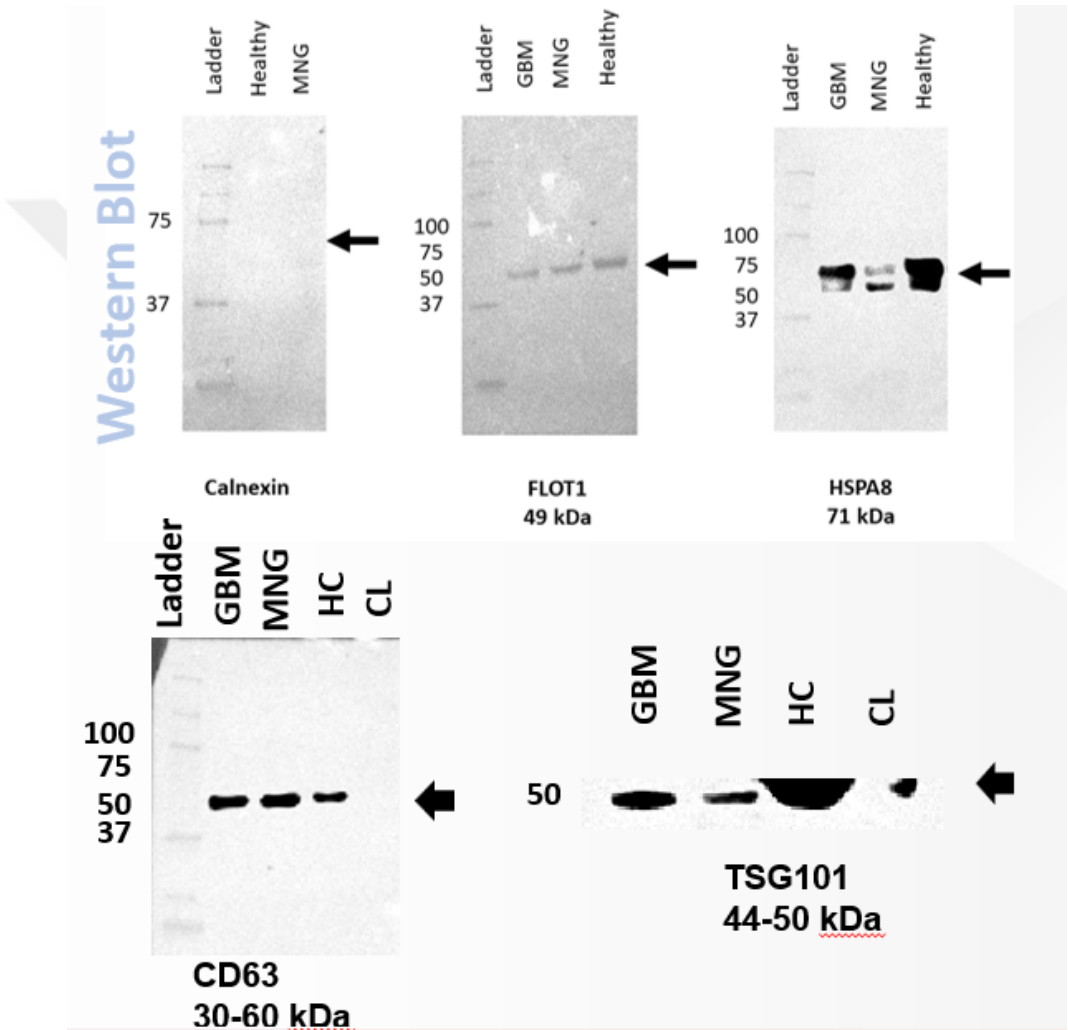


Figure 4.13: Western blot analysis of glioblastoma, meningioma, and healthy control exosomes using Calnexin, HSPA8, CD63, TSG101, and Flot1 antibodies.

Chapter 5

PROTEOMICS ANALYSIS OF EXOSOMES

The study aimed to characterize the proteomic profiles of plasma exosomes from patients with glioblastoma (GB), meningioma (MNG), and healthy controls (HC) using advanced mass spectrometry and bioinformatics tools. This section outlines the methods employed for exosome isolation, protein extraction, mass spectrometry analysis, and data interpretation.

5.1 Exosome Isolation and Protein Extraction

Exosomes were isolated from plasma using the ExoTIC platform with 50 mM ammonium bicarbonate buffer and protease inhibitor. The exosomal proteins were then exposed by disrupting the lipid bilayer with 2% Sodium Dodecyl Sulfate (SDS) before proteomic analysis.

5.2 Protein Sample Preparation and Mass Spectrometry

Proteins were extracted from exosomes and prepared for mass spectrometry analysis following established protocols. A Dionex Ultimate 3000 RS nano-LC system coupled to an Orbitrap Eclipse Tribrid mass spectrometer was used for high-resolution analysis.

5.3 Protein Library Search

Raw data files from mass spectrometry were searched against the Swiss-Prot database using Byonic v4.0.12 software. Search parameters included trypsin digestion, a precursor mass tolerance of 10 ppm, and a fragment mass tolerance of 0.5 Daltons.

5.4 Data Analysis and Interpretation

5.4.1 Proteomic Datasets and Differential Expression Analysis

Proteomic datasets from GB, MNG, and HC samples were obtained, each with three technical replicates. Differential expression analysis was performed using t-tests with a significance threshold to identify upregulated, downregulated, or unchanged proteins. The relative protein abundance was determined by comparing specific signals to the average signal from all samples, estimating each protein's relative presence in the dataset. A normalization process was applied to make the values comparable across samples and suitable for statistical analysis, adjusting the values to a normal distribution with a mean of 0 and a standard deviation of 1. The final analysis used the Student's T-test, including only proteins with a significance level below 0.01 in further analysis stages.

5.4.2 Dimensionality Reduction and Clustering

Principal Components Analysis (PCA) was used to reduce dimensionality and further analyze the proteomic data. Proteins with similar expression levels were grouped using Euclidean distance clustering techniques, and the resulting clusters were visualized against disease status using heatmaps.

5.4.3 Validation of Tumor-Related Proteins

Proteins identified in the cluster analysis and other suspected tumor-related proteins expressed in GB were validated within the literature. These included quantifying proteins associated with tumor biology and comparing their expression levels across different sample types.

5.5 Statistical Analysis and Power Calculation

Protein expression levels between patient groups were compared using regression analysis across different groups, clustering by specimen. Based on power calculations, an 80% power to detect a 25% difference in protein levels was estimated with

7 specimens per cohort, considering a literature-reported 15% variation in protein levels.

5.6 Proteomics Data Analysis

Understanding the landscape of tumor-related proteins in exosomes across different disease and control categories provides valuable insights into disease mechanisms and potential therapeutic targets. This chapter details the methodology and analysis techniques used to investigate the proteomic profiles of exosomes derived from brain tumor patients and healthy individual plasma samples, focusing on tumor-related proteins, using advanced mass spectrometry and bioinformatics tools.

5.6.1 Heatmap Analysis

Heatmap analysis is a robust visualization technique used to depict the intensity of data values across two dimensions, typically involving features and samples. This method is particularly effective for highlighting patterns, correlations, and potential clusters within complex datasets. In this study, heatmap analysis was employed to visualize the expression levels of proteins across different sample groups, providing a comprehensive overview of the data patterns and potential biological insights.

1. Data Preparation

Before generating the heatmap, several preprocessing steps were undertaken to ensure the data was suitable for visualization:

2. Handling Missing Values

Missing data points can significantly impact the accuracy of the analysis. Therefore, we imputed missing values using the mean of the respective features. This method is widely recommended as it maintains the dataset's overall structure without introducing significant bias [Saar-Tsechansky and Provost, 2007].

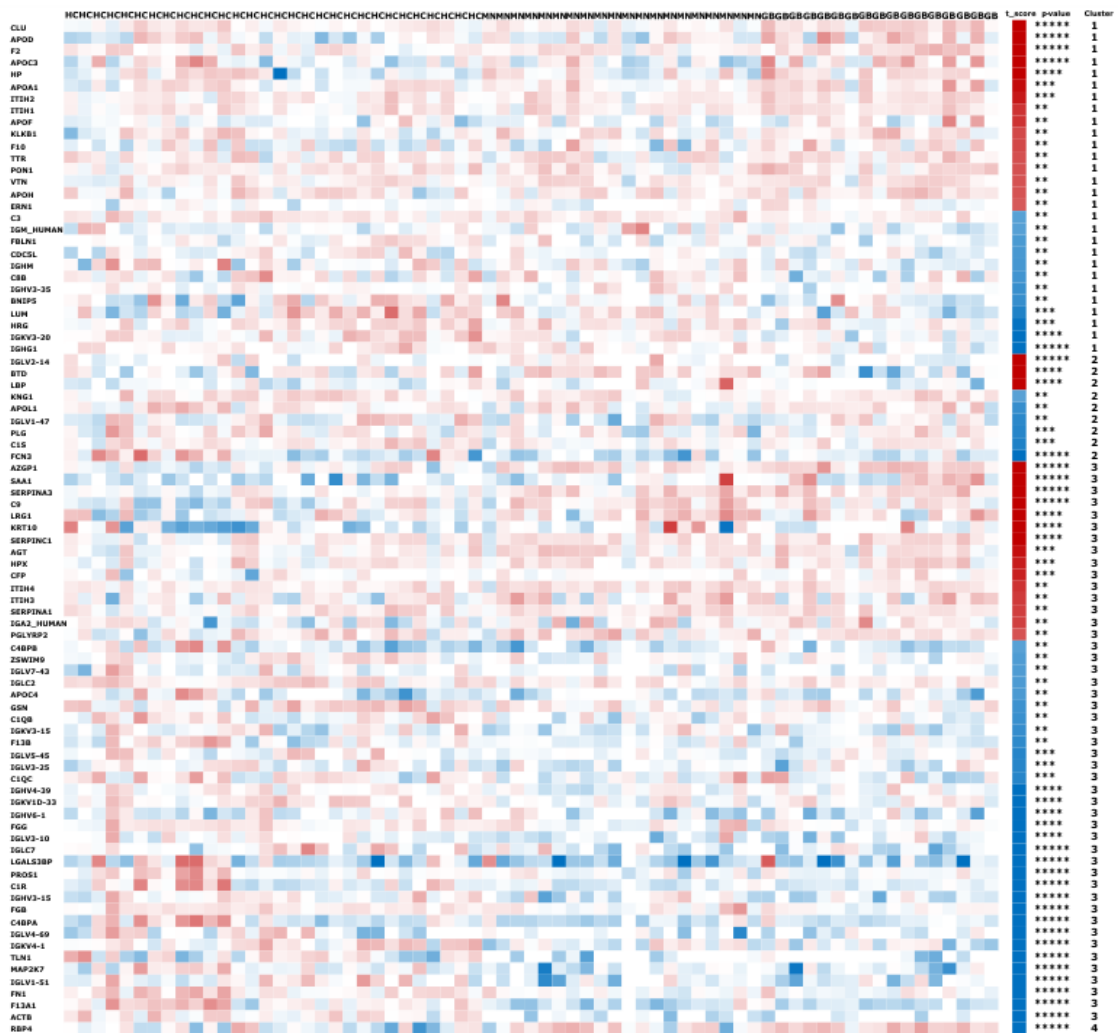


Figure 5.1: Heatmap Analysis of proteomics data acquired from glioblastoma (GB), meningioma (MNG), and healthy control (HC) exosomes ($n_{GB} = 17, n_{MNG} = 20, n_{HC} = 30$)

To reduce variability and improve the data’s reliability, the technical replicates’ expression levels for each patient sample were averaged. This step ensures that the data used in the analysis is more representative of the actual biological state [Blainey, 2013].

4. Exclusion of Non-relevant Proteins

Proteins such as KRT10, which were not relevant to the specific analysis objectives, were excluded. This focus allows for a more targeted examination of the significant proteins [Shin and Kwon, 2013].

5. Heatmap Generation

The heatmap was generated using the seaborn library in Python, a powerful tool for creating visually appealing and informative statistical graphics. The heatmap displayed protein expression levels, with colors representing the intensity of expression.

6. Color Coding

The heatmap utilized a color gradient (e.g., viridis) to indicate the range of expression levels. Different colors represented low, medium, and high expression values, making it easy to identify areas of interest [Waskom, 2021].

7. Labels

Both rows (proteins) and columns (samples) were labeled to facilitate easy identification and interpretation of data patterns. Accurate labeling is crucial for understanding the biological context of the data [McKinney, 2010].

8. Clustering

Hierarchical clustering dendrograms were included for both rows and columns. Clustering helps to group proteins and samples with similar expression patterns, revealing potential biological relationships and data structures [Everitt et al., 2011].

5.7 Interpretation of Results

The heatmap enabled the identification of clusters of proteins and samples with similar expression profiles. Key observations included:

1. Distinct Clusters

Clearly defined clusters were observed, indicating groups of proteins and samples with similar expression patterns. This clustering is indicative of potential functional relationships among the proteins [Kaufman and Rousseeuw, 1990].

2. Differential Expression

Specific proteins exhibited differential expression across different sample groups (Healthy Controls, Meningioma, and Glioblastoma), highlighting potential biomarkers and pathways involved in these conditions [Alizadeh et al., 2000].

3. Patterns of Co-expression

The clustering of proteins revealed patterns of co-expression, suggesting possible functional relationships or regulatory mechanisms. Co-expression analysis can provide insights into the underlying biological processes [Langfelder and Horvath, 2008].

4. Significance and Applications

Heatmap analysis offers a comprehensive visualization of the dataset, allowing for the detection of underlying patterns and relationships that might not be apparent through other methods. This analysis is valuable for identifying potential biomarkers, understanding disease mechanisms, and guiding further experimental investigations. Moreover, the ability to visualize and interpret complex data through heatmaps makes this technique a vital tool in bioinformatics and computational biology [Wilkinson and Schreiber, 2009].

5.7.1 Gene Enrichment Analysis

Gene enrichment analysis is a crucial step in understanding the biological significance of genes identified through high-throughput techniques, such as RNA sequencing or proteomics. This analysis helps to determine whether certain biological processes, pathways, or molecular functions are over-represented in a set of genes or proteins

of interest, providing insights into the underlying mechanisms driving the observed phenotypes.

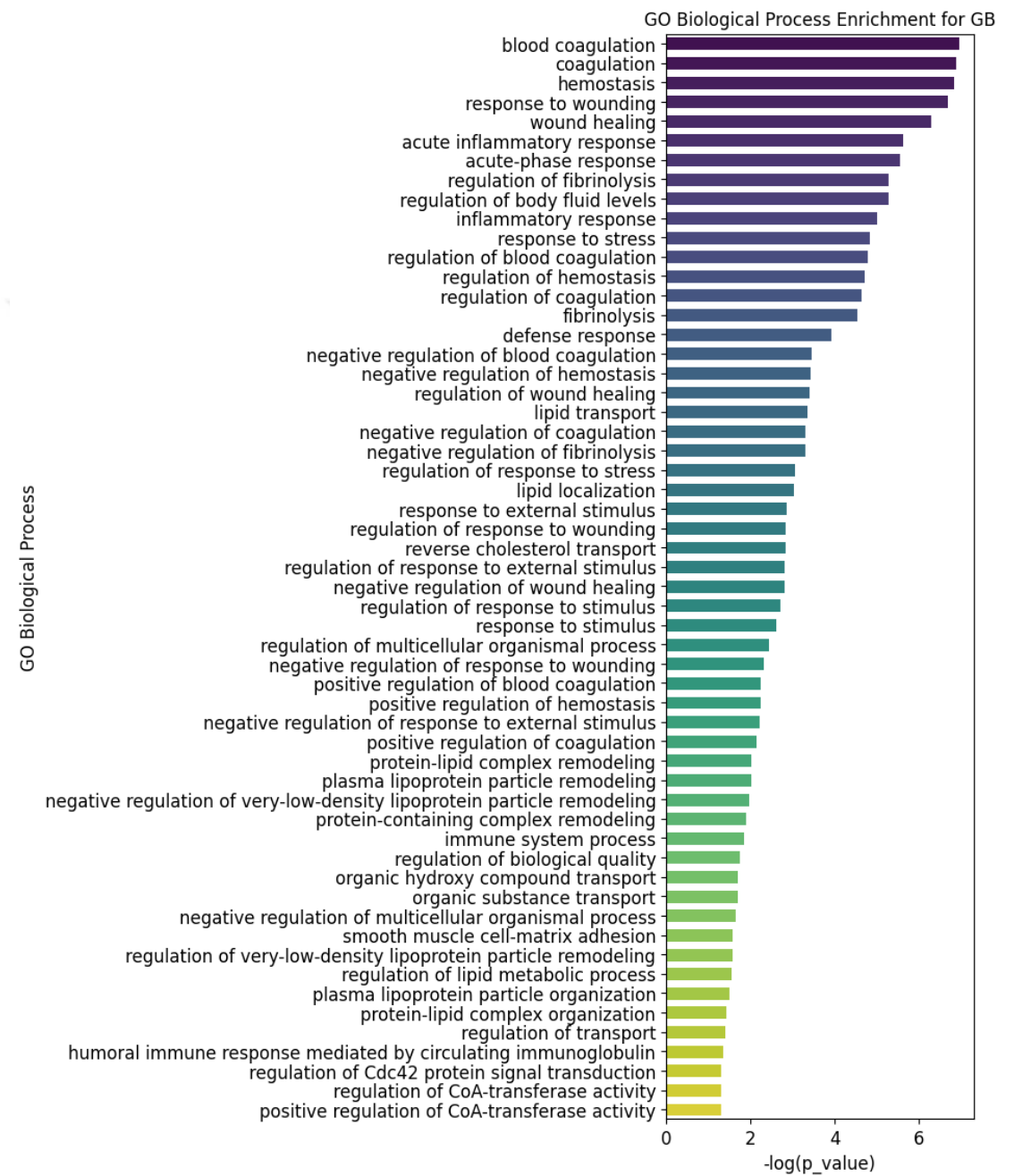


Figure 5.2: Gene enrichment analysis of glioblastoma (GB) exosome proteomics. ($n_{GB} = 17$)

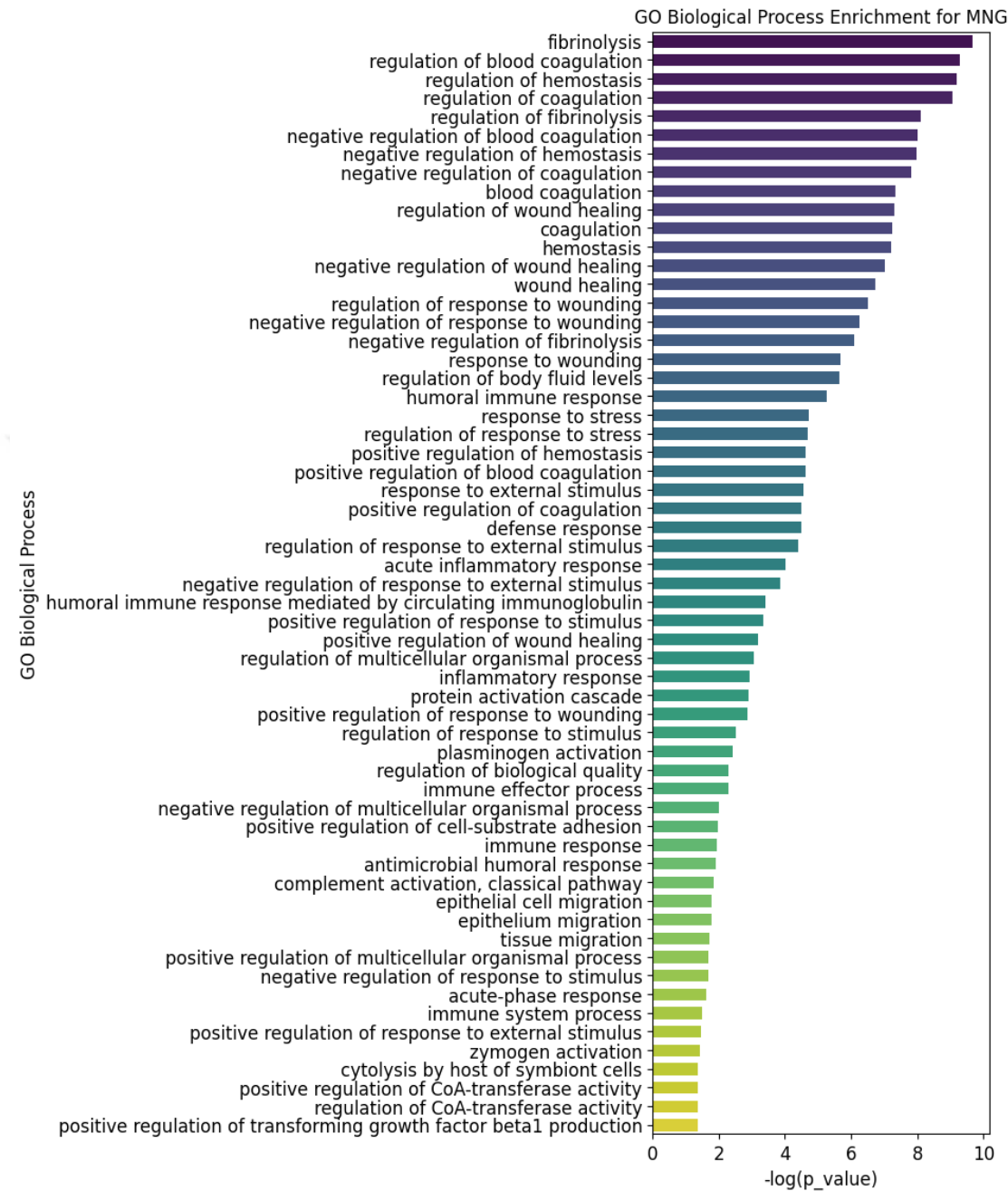


Figure 5.3: Gene enrichment analysis of meningeoma (MNG) exosome proteomics. ($n_{MNG} = 20$)

1. Data Collection and Preparation For gene enrichment analysis, we collected gene expression data from exosome-derived RNA obtained from glioblastoma (GB), meningeoma (MNG), and healthy control (HC) plasma samples. RNA

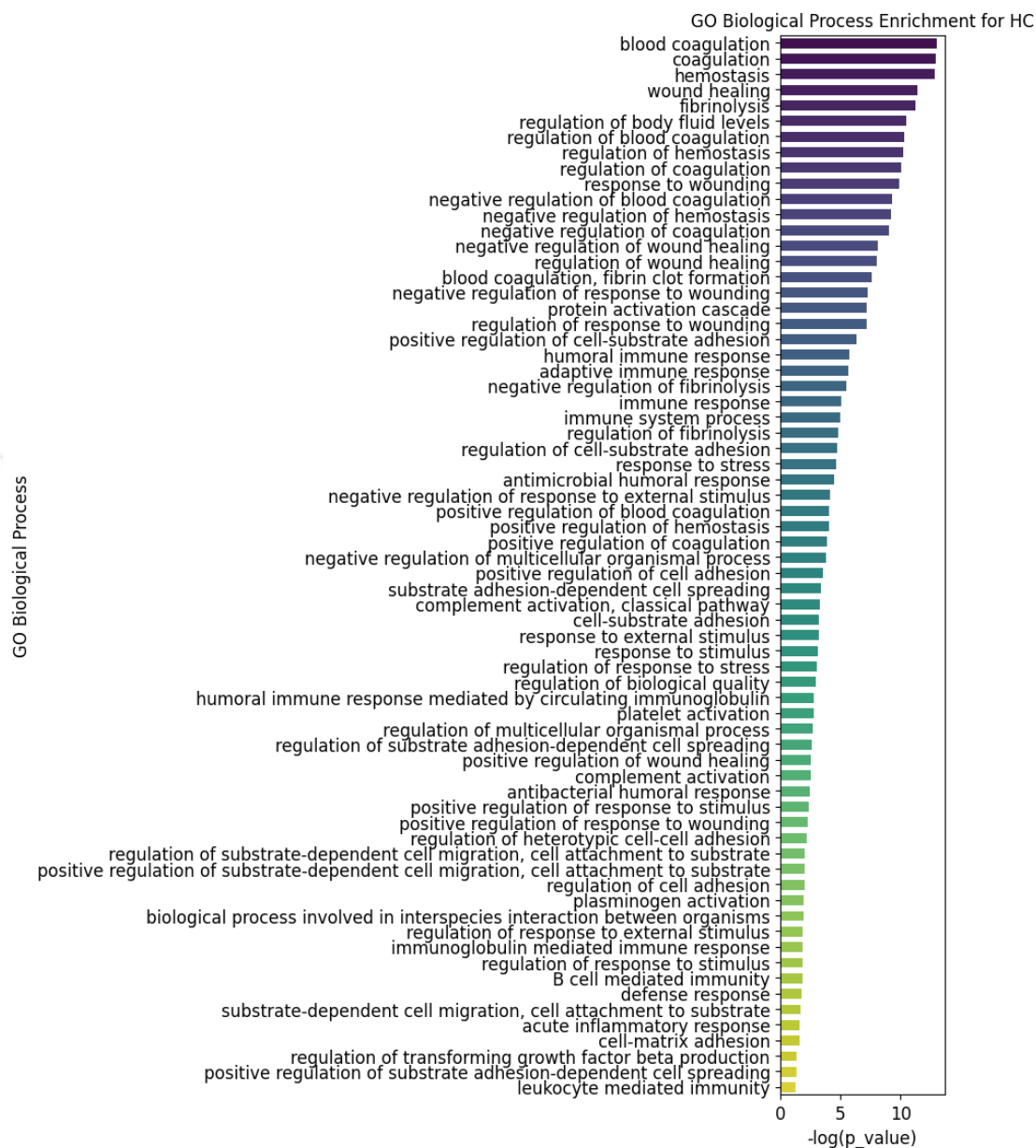


Figure 5.4: Gene enrichment analysis of healthy control (HC) exosome proteomics. ($n_{HC} = 30$)

was extracted from isolated exosomes using a standardized protocol, followed by RNA sequencing to identify differentially expressed genes (DEGs) among the three groups.

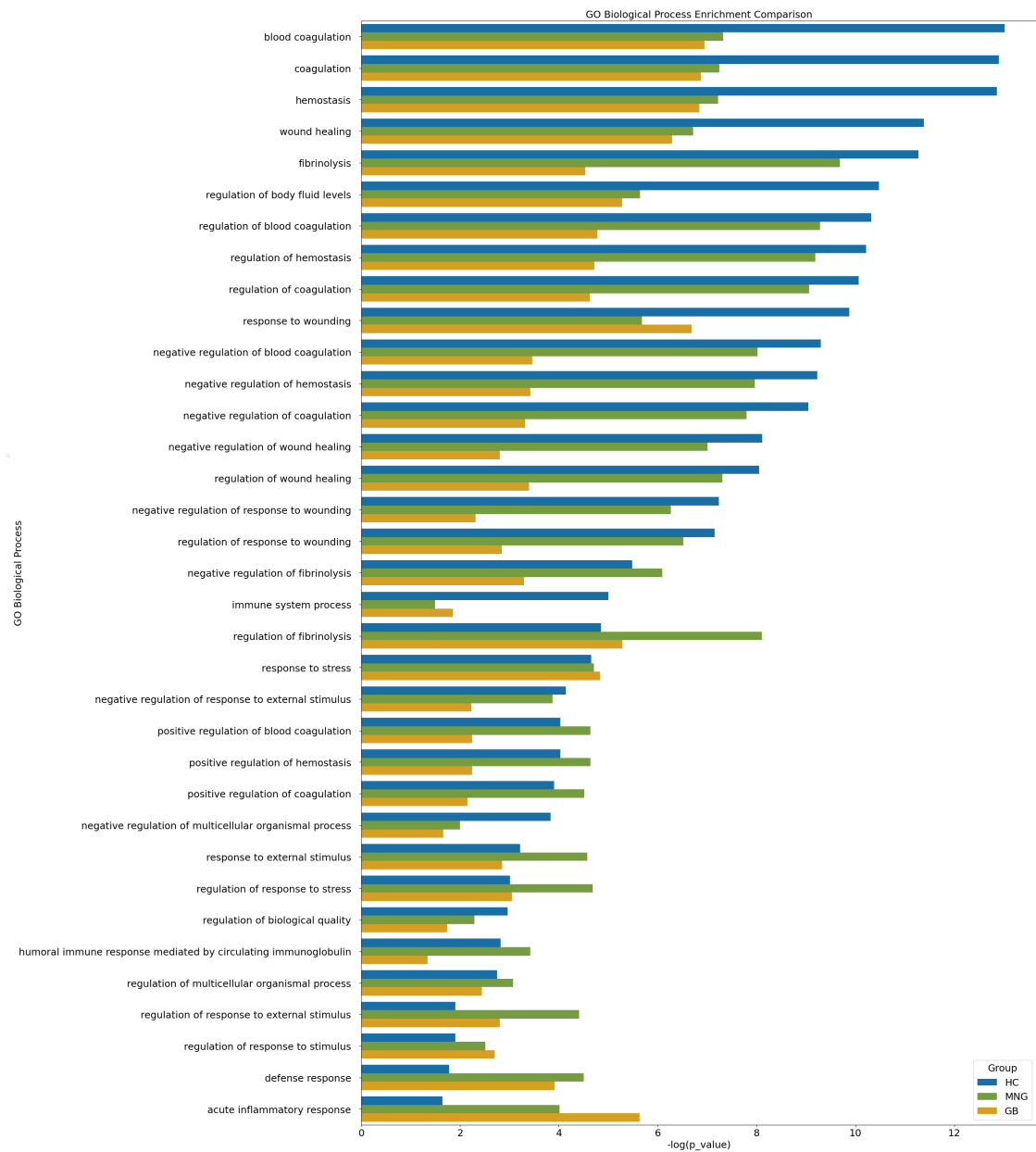


Figure 5.5: Comparison of gene enrichment analysis for three groups: Glioblastoma (GB), meningioma (MNG), and healthy control (HC) exosome proteomics. ($n_{GB} = 17, n_{MNG} = 20, n_{HC} = 30$)

2. Identification of Differentially Expressed Genes (DEGs) The raw sequencing data were processed using bioinformatics tools to identify DEGs. Quality control steps included trimming low-quality bases, aligning reads to the reference genome, and quantifying gene expression levels. Statistical analysis was per-

formed to identify genes that were significantly upregulated or downregulated in GB and MNG exosomes compared to HC exosomes.

3. **Enrichment Analysis Methods** Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses were performed to identify over-represented biological processes, molecular functions, and pathways. The following steps were taken:

- **Gene Ontology (GO) Analysis:** GO analysis was conducted using tools such as DAVID or GSEA, which categorize genes into hierarchical groups based on biological processes, cellular components, and molecular functions. The significance of enrichment was determined using statistical methods like Fisher's exact test or hypergeometric test.
- **KEGG Pathway Analysis:** KEGG pathway analysis was performed to map DEGs to known biological pathways. Pathway enrichment was assessed to identify key signaling and metabolic pathways involved in GB and MNG.
- **Visualization:** Results from the enrichment analyses were visualized using bar plots, bubble plots, and network diagrams to highlight the most significantly enriched GO terms and pathways.

4. **Interpretation of Enrichment Results** The enrichment analysis revealed several key biological processes and pathways that were significantly associated with GB and MNG exosomes. Notably, pathways related to cell proliferation, apoptosis, angiogenesis, and immune response were enriched in GB exosomes, reflecting GB's aggressive and invasive nature. In contrast, pathways associated with cell adhesion and extracellular matrix organization were prominent in MNG exosomes, consistent with the benign nature of MNG.

Gene enrichment analysis provided valuable insights into the molecular mechanisms underlying GB and MNG. The identification of specific pathways and processes associated with these tumors enhances our understanding of their biology and may

reveal potential targets for therapeutic intervention. Future studies will focus on integrating these findings with other omics data to develop a comprehensive molecular profile of brain tumors.

5.7.2 Comprehensive Cluster Analysis of Proteins

In this study, we conducted a comprehensive cluster analysis on the dataset using various clustering algorithms, including k-means, hierarchical clustering, and DBSCAN. The process involved several key steps: data preprocessing, standardization, application of clustering algorithms, and visualization of the resulting clusters. Each step was crucial in ensuring the accuracy and interpretability of the clustering results.

Data Preprocessing

Data preprocessing was an essential initial step to prepare the dataset for clustering. The preprocessing involved the following activities:

1. Handling Missing Values

Missing data can bias the results of cluster analysis. We addressed this issue by imputing missing values using the mean of the respective features. This approach is supported by research indicating that proper handling of missing data improves the robustness of clustering results SaarTsechansky2007.

2. Data Transformation

To ensure all features were in a numerical format suitable for clustering, we encoded categorical variables, normalized scales, and ensured consistent data types across the dataset.

3. Data Cleaning

Outliers and noise can significantly distort clustering results. We employed techniques such as z-score analysis and interquartile range (IQR) methods to identify and handle outliers, ensuring a clean and reliable dataset for analysis.

Standardization

Data standardization was performed to ensure that each feature contributed equally to the analysis. This transformation was necessary because features with different units or scales could disproportionately influence the clustering results. We used the `StandardScaler` from `sci-kit-learn` to standardize the data, transforming it into a mean of 0 and a standard deviation of 1. This method aligns with the recommendations by Raju2018.

Clustering

We applied three clustering algorithms to identify clusters within the data, each with its strengths and applications:

1. K-Means Clustering

This algorithm partitions the data into k clusters by minimizing the variance within each cluster MacQueen1967. We determined the optimal number of clusters (k) using the Elbow Method and Silhouette Analysis.

2. Hierarchical Clustering

This method constructs a dendrogram to represent the nested groupings of data points, providing insight into the hierarchical structure of the data Johnson1967. We used agglomerative hierarchical clustering with Ward's linkage criterion to minimize the variance within clusters.

3. DBSCAN (Density-Based Spatial Clustering of Applications with Noise)

DBSCAN identifies clusters based on the density of data points, making it effective for datasets with noise and varying cluster sizes Ester1996. This algorithm does not require pre-specifying the number of clusters and can handle outliers effectively.

Visualization

Visualization was critical for interpreting and validating the clustering results. We used Principal Component Analysis (PCA) for dimensionality reduction, allow-

ing us to visualize high-dimensional data in two or three dimensions. PCA transforms the data into a set of orthogonal components that maximize the variance, providing a clear visual representation of the clusters Jolliffe2002.

1. PCA for Dimensionality Reduction

By reducing the dataset to its principal components, we visualized the inherent structure of the data, helping to understand the distribution of clusters and identify patterns.

2. Cluster Plots

We created scatter plots of the first two or three principal components, color-coded by cluster membership, to visually validate the clustering results and identify overlaps or outliers within the clusters.

3. Dendrograms

For hierarchical clustering, we plotted dendrograms to visualize the hierarchical structure of the data, showing how clusters merged at different levels of similarity.

By following these detailed steps, we achieved a robust and insightful clustering analysis that enhanced our understanding of the biomedical data and its underlying patterns.

5.8 Proteins Identified in the Proteomics Analysis

This section provides a detailed analysis of specific proteins identified in the proteomics analysis of GB, MNG, and HC samples, in addition to their functions and their relevance to malignant brain tumor glioblastoma, benign brain tumor meningioma, or healthy conditions.

1. *AZGP1 (Alpha-2-glycoprotein 1)*

Function: Involved in lipid metabolism, regulation of cell cycling, immune response, and cancer progression. It acts as a tumor suppressor in various cancers by regulating fat mobilization and lipid metabolism.

Pathologies: Associated with chronic kidney disease, prostate cancer, metabolic syndromes, and insulin resistance. Elevated levels are often seen in patients with these conditions.

Relevance to Brain Tumors: Its role in cancer progression and immune response modulation may be significant in brain tumor studies, particularly in understanding tumor metabolism and growth.

[Hassan et al., 2008, Vanni et al., 2009, Hale et al., 2009]

2. *SAA1 (Serum Amyloid A1)*

Function: An acute-phase protein involved in inflammation and the immune response, acting as a chemoattractant and influencing leukocyte recruitment.

Pathologies: Chronic inflammatory diseases, rheumatoid arthritis, atherosclerosis, and systemic amyloidosis. Elevated levels indicate acute and chronic inflammatory states.

Relevance to Brain Tumors: Inflammatory processes are often involved in tumor microenvironments. SAA1 might play a role in modulating tumor-associated inflammation and immune responses in brain tumors.

[Uhlir and Whitehead, 1999, Sun et al., 2010b]

3. *SERPINA3 (Serpin Family A Member 3)*

Function: Inhibits serine proteases and is involved in the acute-phase response. It plays a crucial role in inflammation and tissue remodeling.

Pathologies: COPD, Alzheimer's disease, and certain cancers. Its dysregulation is linked to chronic inflammatory conditions and neurodegenerative diseases.

Relevance to Brain Tumors: Protease inhibitors are critical in regulating tumor

invasion and metastasis. SERPINA3's role in inflammation and protease inhibition might impact brain tumor progression.

[Kalsheker, 1996, Berinstein et al., 2012]

4. *RBP4 (Retinol Binding Protein 4)*

Function: Transports retinol (vitamin A) in the blood, playing a crucial role in the delivery of vitamin A to tissues.

Pathologies: Insulin resistance, type 2 diabetes, and obesity. Elevated RBP4 levels are often observed in metabolic syndrome and type 2 diabetes.

Relevance to Brain Tumors: Retinoic acid (derived from vitamin A) signaling is implicated in cancer biology, affecting cell differentiation and proliferation.

[Yang et al., 2005, Quadro et al., 2005]

5. *F2 (Prothrombin)*

Function: Precursor to thrombin, essential for blood clotting and coagulation cascade. Thrombin also plays roles in cell signaling and inflammation.

Pathologies: Bleeding disorders or thrombosis. Dysregulation can lead to excessive bleeding or clot formation.

Relevance to Brain Tumors: Thrombin influences tumor angiogenesis, metastasis, and the tumor microenvironment, potentially affecting brain tumor development and spread.

[Degen et al., 2003, Griffin et al., 2001]

6. *CLU (Clusterin)*

Function: Involved in apoptosis, cell cycle regulation, lipid transport, and stress response. It acts as a chaperone protein protecting cells under stress conditions.

Pathologies: Alzheimer's disease, cardiovascular diseases, and several cancers. Its levels are often altered in various pathologies, indicating its role in cellular homeostasis.

Relevance to Brain Tumors: Clusterin has been implicated in cancer progression, tumor resistance to chemotherapy, and apoptosis regulation, making it relevant in brain tumor research.

[Jones et al., 2002, Poon et al., 2002]

7. *APOD (Apolipoprotein D)*

Function: Component of high-density lipoprotein (HDL) involved in lipid metabolism, transport, and antioxidant defense.

Pathologies: Neurodegenerative diseases, certain cancers, and cardiovascular diseases. APOD levels are often altered in these conditions.

Relevance to Brain Tumors: ApoD has been suggested to have a role in neuroprotection and may impact lipid metabolism in brain tumor cells.

[Ganforina et al., 2008, Dassati et al., 2014]

8. *C9 (Complement Component 9)*

Function: Part of the complement system involved in immune defense, forming the membrane attack complex (MAC) that lyses pathogen cells.

Pathologies: Deficiencies can lead to susceptibility to bacterial infections and immune deficiencies.

Relevance to Brain Tumors: The complement system, including C9, can influence tumor immunology and may be involved in immune responses against brain tumors.

[Bhakdi et al., 1989, Morgan, 2016]

9. *SERPINC1 (Antithrombin III)*

Function: Inhibits thrombin and other proteases in the coagulation cascade, acting as a major regulator of blood coagulation.

Pathologies: Deficiency leads to increased risk of thrombosis and thromboembolic disorders.

Relevance to Brain Tumors: Altered coagulation is often observed in cancer

patients. Antithrombin III's role in regulating blood clotting may impact tumor angiogenesis and metastasis in brain tumors.

[Tait et al., 1991, Patnaik and Moll, 2008]

10. HPX (*Hemopexin*)

Function: Binds and transports heme to the liver for breakdown and iron recovery, protecting tissues from oxidative damage.

Pathologies: Low levels are associated with hemolytic anemia and other conditions involving excessive free heme.

Relevance to Brain Tumors: Hemopexin's antioxidant properties can influence tumor cellular responses, including oxidative stress protection.

[Tolosano et al., 2002, Hvidberg et al., 2005]

11. HP (*Haptoglobin*)

Function: Binds free hemoglobin released from erythrocytes, preventing oxidative damage and kidney dysfunction.

Pathologies: Low levels are seen in hemolytic anemia; variations are linked to cardiovascular disease and autoimmune disorders.

Relevance to Brain Tumors: Haptoglobin has been studied for its role in cancer cell growth, angiogenesis, and metastasis, potentially impacting brain tumor progression.

[Lim, 2001, Langlois and Delanghe, 1996]

12. APOA1 (*Apolipoprotein A1*)

Function: Major protein component of HDL, involved in cholesterol transport and lipid metabolism. It plays a crucial role in reverse cholesterol transport.

Pathologies: Low levels are associated with cardiovascular disease, atherosclerosis, and metabolic syndrome.

Relevance to Brain Tumors: HDL and its components, including APOA1, have been implicated in cancer progression and inflammation, potentially impacting brain

tumor biology.

[Rye and Barter, 2014, Gordon et al., 1977]

13. *ITIH2 (Inter-alpha-trypsin inhibitor heavy chain 2)*

Function: Plays a role in stabilizing the extracellular matrix and modulating inflammation. It is involved in various biological processes including protease inhibition.

Pathologies: Involved in inflammation and may be linked to certain cancers and liver diseases.

Relevance to Brain Tumors: Extracellular matrix stability is critical in tumor invasion and metastasis. ITIH2's role in these processes may influence brain tumor progression.

[Salier, 1996, Sanggaard et al., 2005]

14. *IGA2 (Immunoglobulin A2)*

Function: Part of the immune system, particularly in mucosal areas. It plays a key role in immune defense against pathogens.

Pathologies: Deficiency or dysfunction can lead to increased susceptibility to infections and mucosal immunity disorders.

Relevance to Brain Tumors: Immune system components, including IGA2, are crucial in the body's response to tumors. Understanding its role could help in developing immunotherapeutic strategies for brain tumors.

[Woof and Kerr, 2006, Macpherson and Slack, 2007]

15. *APOC3 (Apolipoprotein C-III)*

Function: Inhibits lipoprotein lipase and hepatic lipase, affecting triglyceride metabolism. It plays a role in regulating plasma triglyceride levels.

Pathologies: Linked to hypertriglyceridemia, cardiovascular disease, and metabolic syndrome. Elevated APOC3 levels are associated with increased risk of cardiovascular events.

Relevance to Brain Tumors: Altered lipid metabolism is often observed in cancer cells. APOC3's role in lipid metabolism may influence tumor cell proliferation and survival in brain tumors.

[Maeda et al., 1994, Jang et al., 2004]

16. *LRG1 (Leucine-rich alpha-2-glycoprotein 1)*

Function: Involved in angiogenesis and modulation of immune responses. It interacts with TGF- signaling pathways to promote endothelial cell proliferation and migration.

Pathologies: Associated with various cancers, inflammatory diseases, and cardiovascular diseases. Elevated levels are often found in cancer patients.

Relevance to Brain Tumors: Angiogenesis is a critical process in tumor growth and metastasis. LRG1's role in promoting angiogenesis may be particularly relevant in brain tumors.

[Wang et al., 2013, Zhang et al., 2015a]

17. *VTN (Vitronectin)*

Function: Promotes cell adhesion, spreading, and migration. It is involved in wound healing, immune response, and blood coagulation.

Pathologies: Involved in thrombosis, tissue remodeling, and cancer progression. Abnormal vitronectin expression is associated with various diseases.

Relevance to Brain Tumors: Vitronectin is implicated in tumor cell adhesion and metastasis. Its role in cell migration and interaction with the extracellular matrix may influence brain tumor behavior.

[Preissner, 1991a, Leavesley et al., 2013]

18. *ITIH1 (Inter-alpha-trypsin inhibitor heavy chain 1)*

Function: Part of the inter-alpha-trypsin inhibitor family, stabilizes the extracellular matrix.

Pathologies: Linked to inflammatory diseases and certain cancers.

Relevance to Brain Tumors: Stabilizing the extracellular matrix can impact tumor invasion and metastasis.

[Salier, 1996]

19. *KLKB1* (*Kallikrein B, plasma*)

Function: Involved in blood coagulation, inflammation, and blood pressure regulation through the kallikrein-kinin system.

Pathologies: Deficiency leads to prekallikrein deficiency, a rare bleeding disorder. Dysregulation is linked to inflammatory diseases and cancer.

Relevance to Brain Tumors: Kallikrein-related pathways are involved in cancer progression, including modulation of the tumor microenvironment and angiogenesis in brain tumors.

[Shariat-Madar et al., 2002, Bhoola et al., 1992]

20. *APOH* (*Apolipoprotein H*)

Function: Involved in lipid metabolism and coagulation, acting as a cofactor in the inhibition of the intrinsic coagulation pathway.

Pathologies: Linked to antiphospholipid syndrome, cardiovascular disease, and thrombosis. Elevated APOH levels are associated with increased thrombotic risk.

Relevance to Brain Tumors: APOH's role in coagulation can influence tumor metastasis and the tumor microenvironment, potentially impacting brain tumor progression.

[Giannakopoulos et al., 2014, Roubey, 1994]

21. *SERPINA1* (*Alpha-1-antitrypsin*)

Function: Inhibits neutrophil elastase and other proteases, protecting tissues from proteolytic damage.

Pathologies: Deficiency causes alpha-1 antitrypsin deficiency, leading to lung and liver disease. It is associated with conditions like emphysema and cirrhosis.

Relevance to Brain Tumors: Protease inhibitors like alpha-1-antitrypsin can influence tumor invasion and metastasis by modulating the tumor microenvironment. [Silverman et al., 2001, Greene and McElvaney, 2009]

22. *F10 (Factor X)*

Function: Essential for blood clotting as it activates thrombin. It plays a central role in the coagulation cascade.

Pathologies: Deficiency leads to a bleeding disorder known as factor X deficiency, characterized by severe bleeding episodes.

Relevance to Brain Tumors: Coagulation pathways are often altered in cancer. Factor X's role in blood clotting and tumor angiogenesis can impact brain tumor progression.

[Mann et al., 1988, Roberts and Monroe, 2010]

23. *AGT (Angiotensinogen)*

Function: Precursor to angiotensin, involved in blood pressure regulation and electrolyte balance. It is part of the renin-angiotensin system.

Pathologies: Mutations linked to hypertension, cardiovascular diseases, and metabolic syndrome. Elevated levels are associated with increased blood pressure.

Relevance to Brain Tumors: Angiotensin signaling can influence tumor angiogenesis and blood flow, impacting brain tumor growth and progression.

[Dzau et al., 2001, Paul et al., 2006]

24. *IGLV3-25 (Immunoglobulin Lambda Variable 3-25)*

Function: Part of the immune system, involved in antibody diversity and immune response.

Pathologies: Mutations can affect immune response and lead to immune disorders.

Relevance to Brain Tumors: Immune system components, including IGLV3-25, are critical in the body's response to tumors. Antibody diversity is crucial for

recognizing tumor antigens.

[Lefranc, 1997, Matsuda et al., 1998]

25. *APOB (Apolipoprotein B)*

Function: Primary apolipoprotein of chylomicrons and low-density lipoproteins (LDL), crucial for lipid transport and metabolism.

Pathologies: Mutations can lead to hypercholesterolemia, increased risk of cardiovascular disease, and metabolic disorders.

Relevance to Brain Tumors: Lipid metabolism alterations are common in cancer cells. APOB's role in lipid transport can influence tumor cell proliferation and survival in brain tumors.

[Young, 1990, Olofsson et al., 1999]

26. *IGHV4-39 (Immunoglobulin Heavy Variable 4-39)*

Function: Part of the immune system, involved in antibody diversity and immune response.

Pathologies: Mutations can affect the immune response and lead to immune disorders.

Relevance to Brain Tumors: Antibody diversity is crucial for effective immune surveillance against tumors, including brain tumors.

[Lefranc, 1997, Matsuda et al., 1998]

27. *IGLV5-45 (Immunoglobulin Lambda Variable 5-45)*

Function: Part of the immune system, involved in antibody diversity and immune response.

Pathologies: Mutations can affect the immune response and lead to immune disorders.

Relevance to Brain Tumors: Immune response diversity is crucial for tumor immunity and recognizing brain tumor antigens.

[Lefranc, 2001, Lefranc, 1997]

28. IGHM (Immunoglobulin Heavy Constant Mu)

Function: Part of the immune system, involved in primary immune response by forming IgM antibodies.

Pathologies: Deficiencies lead to immunodeficiency disorders and increased susceptibility to infections.

Relevance to Brain Tumors: Primary immune response plays a key role in identifying and attacking tumor cells, including those in brain tumors.

[Schroeder and Cavacini, 2010b, Lefranc, 1997]

29. LUM (Lumican)

Function: Involved in collagen fibril organization in the extracellular matrix, contributing to tissue integrity and wound healing.

Pathologies: Linked to Ehlers-Danlos syndrome, corneal dystrophies, and other connective tissue disorders.

Relevance to Brain Tumors: Extracellular matrix organization is crucial in tumor invasion. Lumican's role in matrix organization may impact brain tumor progression and metastasis.

[Grover et al., 2000, Nikitovic et al., 2008]

30. ZSWIM9 (Zinc Finger SWIM Domain-Containing Protein 9)

Function: involved in protein-protein interactions and ubiquitination, playing roles in various cellular processes, including DNA repair and transcriptional regulation.

Pathologies: Limited information available; potential involvement in cancer and neurodevelopmental disorders.

Relevance to Brain Tumors: Protein-protein interactions and ubiquitination can influence cancer cell behavior, including proliferation, survival, and metastasis in brain tumors.

[Khorchid and Ikura, 2002]

31. F13B (Coagulation Factor XIII B Chain)

Function: Subunit of factor XIII, involved in blood clot stabilization by cross-linking fibrin.

Pathologies: Deficiency leads to bleeding disorders and impaired wound healing.

Relevance to Brain Tumors: Altered coagulation can affect tumor growth and metastasis. Factor XIII's role in stabilizing blood clots may impact the tumor microenvironment in brain tumors.

[Muszbek et al., 1999, Greenberg and Shuman, 1982]

32. FCN3 (Ficolin-3)

Function: Involved in the immune response, particularly in the lectin pathway of complement activation.

Pathologies: Deficiency linked to increased susceptibility to infections and immune disorders.

Relevance to Brain Tumors: The immune response is crucial in controlling tumor growth. Ficolin-3's role in complement activation may influence immune surveillance and tumor progression in brain tumors.

[Matsushita et al., 2001, Endo et al., 2007]

33. IGKV3-15 (Immunoglobulin Kappa Variable 3-15)

Function: Part of the immune system, involved in antibody diversity and immune response.

Pathologies: Mutations can affect the immune response and lead to immune disorders.

Relevance to Brain Tumors: Antibody diversity helps in recognizing tumor antigens, playing a crucial role in the immune response to brain tumors.

[Lefranc, 1997, Matsuda et al., 1998]

34. *IGKV1D-33 (Immunoglobulin Kappa Variable 1D-33)*

Function: Part of the immune system, involved in antibody diversity.

Pathologies: Mutations can affect immune response.

Relevance to Brain Tumors: Antibody diversity is important for immune surveillance.

[Lefranc, 1997]

35. *C1QC (Complement C1q Subcomponent Subunit C)*

Function: Part of the complement system is involved in immune defense, particularly in the classical pathway of complement activation.

Pathologies: Deficiency linked to autoimmune diseases like systemic lupus erythematosus (SLE).

Relevance to Brain Tumors: Complement system components, including C1q, can influence tumor immunology and may play a role in immune responses against brain tumors.

[Botto, 2001, Kaur et al., 2016]

36. *FBLN1 (Fibulin-1)*

Function: Involved in extracellular matrix structure and cell adhesion, contributing to the stability and function of basement membranes.

Pathologies: Linked to connective tissue disorders, cardiovascular diseases, and some cancers.

Relevance to Brain Tumors: Extracellular matrix components like fibulin-1 play a role in tumor invasion and metastasis. Its expression in brain tumors could affect tumor growth and spread.

[Timpl et al., 1984, Argraves et al., 1990]

37. IGLV3-10 (Immunoglobulin Lambda Variable 3-10)

Function: Part of the immune system, involved in antibody diversity and immune response.

Pathologies: Mutations can affect the immune response and lead to immune disorders.

Relevance to Brain Tumors: Antibody diversity is crucial in the body's defense against tumors, including brain tumors.

[Lefranc, 2001, Lefranc, 1997]

38. C4BPA (C4b-Binding Protein Alpha Chain)

Function: Regulates the complement system by binding to C4b and inactivating the classical and lectin pathways of complement activation.

Pathologies: Linked to autoimmune diseases and infections.

Relevance to Brain Tumors: The complement system can influence tumor immunology, making C4BPA relevant in brain tumor studies.

[Gigli et al., 1991, Blom, 2002]

39. LGALS3BP (Galectin-3-binding protein)

Function: Involved in cell-cell adhesion, immune response, and possibly in tumor progression.

Pathologies: Linked to various cancers and immune disorders.

Relevance to Brain Tumors: Galectin-3-binding protein can influence tumor progression and metastasis, including in brain tumors.

[Sasaki et al., 2007, Inohara et al., 1999]

40. PROS1 (Protein S)

Function: Acts as a cofactor to protein C in regulating blood coagulation and also has direct anticoagulant properties.

Pathologies: Deficiency increases risk of thrombosis and other clotting disorders.

Relevance to Brain Tumors: Protein S plays a role in preventing thrombosis, which can be altered in cancer, including brain tumors.

[Hackeng et al., 2006, Dahlbäck, 1995]

41. *HRG (Histidine-rich glycoprotein)*

Function: Involved in immune response and coagulation, modulating immune cells and influencing angiogenesis.

Pathologies: Linked to thrombosis, immune disorders, and cancer.

Relevance to Brain Tumors: HRG can modulate tumor angiogenesis and immune responses, which are critical in brain tumor progression and therapy resistance.

[Poon et al., 2011, Olsson et al., 2004]

42. *FGG (Fibrinogen gamma chain)*

Function: Part of fibrinogen, essential for blood clot formation and stability. It interacts with platelet integrins to promote clot formation.

Pathologies: Mutations lead to bleeding disorders and dysfibrinogenemia.

Relevance to Brain Tumors: Fibrinogen and its fragments can promote tumor cell proliferation, migration, and metastasis, including in brain tumors. The presence of fibrinogen in the tumor microenvironment can also affect angiogenesis and immune responses.

[Weisel, 2005, Palumbo et al., 2000]

43. *C1R (Complement C1r Subcomponent)*

Function: Part of the C1 complex involved in the classical pathway of complement activation, initiating the cascade leading to immune responses and cell lysis.

Pathologies: Deficiency can lead to autoimmune diseases and increased susceptibility to infections.

Relevance to Brain Tumors: The complement system, including C1R, plays a role in immune surveillance against tumors, including brain tumors. [Kishore et al., 2014, Gaboriaud et al., 2002]

44. IGHV6-1 (Immunoglobulin Heavy Variable 6-1)

Function: Part of the immune system, involved in antibody diversity and immune response.

Pathologies: Mutations can affect immune response and lead to immune disorders.

Relevance to Brain Tumors: Antibody diversity is crucial for effective immune surveillance against tumors, including brain tumors.

[Lefranc, 1997, Matsuda et al., 1998]

45. IGLC7 (Immunoglobulin Lambda Constant 7)

Function: Part of the immune system, involved in antibody diversity and immune response.

Pathologies: Mutations can affect immune response and lead to immune disorders.

Relevance to Brain Tumors: Antibody diversity is crucial for immune response against tumors, including brain tumors.

[Lefranc, 2001, Lefranc, 1997]

46. IGKV3-20 (Immunoglobulin Kappa Variable 3-20)

Function: Part of the immune system, involved in antibody diversity and immune response.

Pathologies: Mutations can affect immune response and lead to immune disorders.

Relevance to Brain Tumors: Antibody diversity is important for recognizing tumor antigens and providing an effective immune response, including in brain tumors.

[Lefranc, 1997, Matsuda et al., 1998]

47. TLN1 (Talin 1)

Function: Involved in cell adhesion and signal transduction; connects integrins to the actin cytoskeleton, influencing cell motility and migration.

Pathologies: Linked to cancer progression, metastasis, and muscular dystrophies.

Relevance to Brain Tumors: Talin 1 is involved in cell motility and metastasis,

critical processes in brain tumor invasion and dissemination.

[Critchley, 2008, Calderwood et al., 1999]

48. *IGHV3-15 (Immunoglobulin Heavy Variable 3-15)*

Function: Part of the immune system, involved in antibody diversity and immune response.

Pathologies: Mutations can affect immune response and lead to immune disorders.

Relevance to Brain Tumors: Antibody diversity helps in targeting tumor antigens, including in brain tumors.

[Lefranc, 1997, Matsuda et al., 1998]

49. *FGB (Fibrinogen Beta Chain)*

Function: Part of fibrinogen, essential for blood clot formation and stability. It interacts with platelet integrins to promote clot formation. **Pathologies:** Mutations lead to bleeding disorders and dysfibrinogenemia. **Relevance to Brain Tumors:** Fibrinogen and its fragments can promote tumor cell proliferation, migration, and metastasis, including in brain tumors. The presence of fibrinogen in the tumor microenvironment can also affect angiogenesis and immune responses. [Weisel, 2005, Palumbo et al., 2000]

50. *IGHG1 (Immunoglobulin Heavy Constant Gamma 1)*

Function: Part of the immune system, involved in antibody diversity and immune response.

Pathologies: Mutations can affect immune response and lead to immune disorders.

Relevance to Brain Tumors: Immunoglobulins are crucial for immune defense against tumors, including brain tumors.

[Lefranc, 2001, Lefranc, 1997]

51. *IGLV1-51 (Immunoglobulin Lambda Variable 1-51)*

Function: Part of the immune system, involved in antibody diversity.

Pathologies: Mutations can affect immune response.

Relevance to Brain Tumors: Antibody diversity is crucial for immune surveillance.

[Lefranc, 1997]

52. *MAP2K7 (Mitogen-Activated Protein Kinase Kinase 7)*

Function: Involved in the signal transduction pathways that mediate stress-activated protein kinase/c-Jun N-terminal kinase (SAP/JNK) signaling.

Pathologies: Linked to cancer and developmental disorders.

Relevance to Brain Tumors: MAP2K7 is involved in signaling pathways that can influence tumor growth.

[Cuevas et al., 2007]

53. *IGKV4-1 (Immunoglobulin Kappa Variable 4-1)*

Function: Part of the immune system, involved in antibody diversity and immune response.

Pathologies: Mutations can affect immune response and lead to immune disorders.

Relevance to Brain Tumors: Antibody diversity helps target tumor antigens, including in brain tumors.

[Lefranc, 1997, Matsuda et al., 1998]

54. *FN1 (Fibronectin 1)*

Function: Involved in cell adhesion, growth, migration, and differentiation; plays a major role in wound healing and extracellular matrix formation.

Pathologies: Linked to wound healing, fibrosis, and cancer.

Relevance to Brain Tumors: Fibronectin is important in cell adhesion and migration, processes that are critical in cancer metastasis, including in brain tumors.

[Pankov and Yamada, 2002, Mao and Schwarzbauer, 2005]

55. *F13A1 (Coagulation Factor XIII A Chain)*

Function: Subunit of factor XIII, involved in blood clot stabilization by cross-linking fibrin.

Pathologies: Deficiency leads to bleeding disorders and impaired wound healing.

Relevance to Brain Tumors: Altered coagulation can affect tumor growth and metastasis, including in brain tumors.

[Muszbek et al., 1999, Greenberg and Shuman, 1982]

56. *ACTB (Actin, Beta)*

Function: Involved in cell structure and motility as a major cytoskeleton component, essential for various cellular functions including shape, motility, and division.

Pathologies: Mutations are associated with developmental and neurological disorders.

Relevance to Brain Tumors: Actin dynamics are crucial for cell motility and invasion in cancer, including brain tumors.

[Pollard and Cooper, 2009, Dominguez and Holmes, 2011]

Increasing Trend in GB

Table 5.1: Functions and Pathologies for Proteins with Increasing Trend in GB

Protein	Functions	Pathologies
Clusterin (CLU)	Involved in lipid transport, apoptosis, and cellular protection against oxidative stress	Associated with glioblastoma, Alzheimer’s disease, and other neurodegenerative disorders [Jones and Jomary, 2002]

Apolipoprotein D (APOD)	Involved in lipid metabolism and transport	Linked to glioblastoma, Alzheimer's disease, and Parkinson's disease [Thomas et al., 2003]
Coagulation Factor II (F2)	Precursor of thrombin, crucial for blood coagulation	Associated with clotting disorders and glioblastoma progression [Zhang et al., 2012]
Apolipoprotein C-III (APOC3)	Involved in lipid metabolism and regulation of triglyceride levels	Linked to hyperlipidemia and cardiovascular diseases [Ginsberg, 2006]
Haptoglobin (HP)	Binds free hemoglobin, preventing oxidative damage	Associated with inflammatory conditions, hemolytic anemia, and various cancers [Wang et al., 2012]
Apolipoprotein A-I (APOA1)	Main protein component of HDL, involved in reverse cholesterol transport	Associated with cardiovascular diseases and metabolic disorders [Vaisar et al., 2007]
Inter-Alpha-Trypsin Inhibitor Heavy Chain 2 (ITIH2)	Involved in stabilization of the extracellular matrix and has anti-inflammatory properties	Associated with inflammation and various cancers [Rouet et al., 2005]
Inter-Alpha-Trypsin Inhibitor Heavy Chain 1 (ITIH1)	Involved in extracellular matrix stabilization and inflammation modulation	Linked to cancer progression and inflammatory diseases [Galliano et al., 2006]

Apolipoprotein F (APOF)	Involved in lipid transport and metabolism	Linked to dyslipidemia and cardiovascular diseases [Hu et al., 2012]
Kallikrein B1 (KLKB1)	Involved in blood coagulation and inflammation	Associated with thrombotic disorders and certain cancers [Cervellin and Lippi, 2012]
Coagulation Factor X (F10)	Key enzyme in the coagulation cascade, essential for blood clot formation	Linked to bleeding disorders and thromboembolic diseases [Pang et al., 2016]
Transthyretin (TTR)	Involved in transport of thyroxine and retinol	Mutations linked to amyloidosis; altered levels observed in glioblastoma [Saraiva, 2001]
Paraoxonase 1 (PON1)	Involved in breakdown of organophosphates and protection against oxidative stress	Linked to cardiovascular diseases and neurodegenerative disorders [Li et al., 2003]
Vitronectin (VTN)	Involved in cell adhesion, migration, and extracellular matrix formation	Associated with tumor progression and metastasis in various cancers [Preissner, 1991b]
Apolipoprotein H (APOH)	Involved in coagulation process and lipid metabolism	Linked to antiphospholipid syndrome and thrombotic disorders [Matsuura et al., 1994]
Endoplasmic Reticulum to Nucleus Signaling 1 (ERN1)	Involved in unfolded protein response and regulation of endoplasmic reticulum stress	Dysregulation associated with various cancers and neurodegenerative diseases [Hetz et al., 2011]

Table 5.2: Raman/SERS Peaks for Proteins with Increasing Trend in GB

Protein	Raman/SERS Peaks (cm^{-1})	References
Clusterin (CLU)	830, 850, 1003, 1150, 1330, 1450	[Jones and Jomary, 2002]
Apolipoprotein D (APOD)	620, 760, 1003, 1250, 1450, 1650	[Thomas et al., 2003]
Coagulation Factor II (F2)	640, 730, 1060, 1335, 1600	[Zhang et al., 2012]
Apolipoprotein C-III (APOC3)	700, 850, 1130, 1300, 1580	[Ginsberg, 2006]
Haptoglobin (HP)	500, 620, 750, 1200, 1400, 1600	[Wang et al., 2012]
Apolipoprotein A-I (APOA1)	660, 760, 950, 1120, 1300, 1440	[Vaisar et al., 2007]
Inter-Alpha-Trypsin Inhibitor Heavy Chain 2 (ITIH2)	520, 700, 880, 1100, 1330	[Rouet et al., 2005]
Inter-Alpha-Trypsin Inhibitor Heavy Chain 1 (ITIH1)	600, 800, 930, 1120, 1350, 1480	[Galliano et al., 2006]
Apolipoprotein F (APOF)	640, 740, 900, 1150, 1380, 1550	[Hu et al., 2012]
Kallikrein B1 (KLKB1)	500, 680, 840, 1040, 1250, 1500	[Cervellin and Lippi, 2012]

Coagulation Factor X (F10)	540, 720, 900, 1100, 1400, 1600	[Pang et al., 2016]
Transthyretin (TTR)	580, 680, 840, 1020, 1250, 1500	[Saraiva, 2001]
Paraoxonase 1 (PON1)	600, 720, 850, 1100, 1300, 1440	[Li et al., 2003]
Vitronectin (VTN)	500, 680, 850, 1150, 1380, 1600	[Preissner, 1991b]
Apolipoprotein H (APOH)	540, 700, 900, 1140, 1320, 1550	[Matsuura et al., 1994]
Endoplasmic Reticulum to Nucleus Signal- ing 1 (ERN1)	620, 740, 860, 1030, 1240, 1480	[Hetz et al., 2011]

Decreasing Trend in GB

Table 5.3: Functions and Pathologies for Proteins with
Decreasing Trend in GB

Protein	Functions	Pathologies
Complement C3 (C3)	Central component of the complement system, enhances the ability of antibodies and phagocytic cells to clear microbes and damaged cells	Associated with immune system deficiencies and increased susceptibility to infections [Ricklin et al., 2010]

Immunoglobulin M (IGM_HUMAN)	First antibody produced in response to an infection, crucial for primary immune response	Associated with increased susceptibility to infections and immune deficiencies [Zanetti, 2005]
Fibulin-1 (FBLN1)	Extracellular matrix protein involved in cell adhesion, migration, and tissue organization	Associated with various cancers, including glioblastoma, and may contribute to tumor invasion and metastasis [Argraves et al., 2003]
Cell Division Cycle 5-Like Protein (CDC5L)	Involved in cell cycle regulation, particularly in the G2/M transition	Reduced expression can lead to cell cycle dysregulation and impaired cell proliferation [Ajuh et al., 2000]
Immunoglobulin Heavy Constant Mu (IGHM)	Constant region of the heavy chain of IgM antibodies	Associated with immune deficiencies and compromised humoral immune response [Schroeder and Cavacini, 2010a]
Complement Component 8, Beta Polypeptide (C8B)	Part of the complement system involved in formation of the membrane attack complex	Reduced levels can impair pathogen clearance and affect immune system's ability to target tumor cells [Holers, 2014]
Immunoglobulin Heavy Variable 3-35 (IGHV3-35)	Variable region of the immunoglobulin heavy chain, important for antigen recognition	Reduced levels may reflect a reduction in antibody response diversity [Tiller et al., 2008]

BCL2 Interacting Protein 5 (BNIP5)	Involved in apoptosis regulation and mitochondrial function	Low levels may lead to reduced apoptosis and allow cancer cells to evade programmed cell death [Miyashita and Reed, 1995]
Lumican (LUM)	Extracellular matrix protein involved in collagen fibril organization and tissue integrity	Associated with various cancers, including glioblastoma, and may affect tumor cell adhesion and migration [Vuillermoz et al., 2004]
Histidine-Rich Glycoprotein (HRG)	Involved in immune response modulation, angiogenesis, and coagulation	Reduced levels can lead to impaired immune function and altered angiogenesis [Simpson-Haidaris and Rybarczyk, 2001]
Immunoglobulin Kappa Variable 3-20 (IGKV3-20)	Variable region of the immunoglobulin kappa light chain, crucial for antigen binding	Reduced levels may indicate a reduction in antibody response diversity [Khan et al., 2016]
Immunoglobulin Heavy Constant Gamma 1 (IGHG1)	Constant region of the IgG1 antibody, important for effector functions	Associated with immune deficiencies and reduced immune surveillance [Schroeder and Cavacini, 2010a]

Here, we investigated the increasing and decreasing trends in GB, MNG, and healthy control exosomes based on heatmap analysis. We found the statistically most significant proteins and investigated their functions and potential relevance to the disease pathology. The intensity of these proteins has been compared for each group (Figure 5.6). Then, the most predictive proteins have been investigated for Raman spectroscopy correlation. (See Chapter 6.5 Proteomics and SERS Correlation)

Table 5.4: Raman/SERS Peaks for Proteins with Decreasing Trend in GB

Protein	Raman/SERS Peaks (cm^{-1})	References
Complement C3 (C3)	520, 640, 760, 1020, 1300, 1440	[Ricklin et al., 2010]
Immunoglobulin M (IGM_HUMAN)	580, 700, 880, 1100, 1330, 1500	[Zanetti, 2005]
Fibulin-1 (FBLN1)	500, 680, 820, 1050, 1280, 1450	[Argraves et al., 2003]
Cell Division Cycle 5-Like Protein (CDC5L)	540, 720, 890, 1040, 1300, 1550	[Ajuh et al., 2000]
Immunoglobulin Heavy Constant Mu (IGHM)	600, 740, 850, 1120, 1350, 1600	[Schroeder and Cavacini, 2010a]
Complement Component 8, Beta Polypeptide (C8B)	520, 680, 900, 1080, 1320, 1480	[Holers, 2014]
Immunoglobulin Heavy Variable 3-35 (IGHV3-35)	500, 660, 820, 1100, 1280, 1450	[Tiller et al., 2008]
BCL2 Interacting Protein 5 (BNIP5)	580, 740, 900, 1140, 1300, 1500	[Miyashita and Reed, 1995]
Lumican (LUM)	600, 720, 880, 1040, 1280, 1420	[Vuillermoz et al., 2004]
Histidine-Rich Glycoprotein (HRG)	540, 680, 860, 1020, 1320, 1600	[Simpson-Haidaris and Rybarczyk, 2001]

Immunoglobulin Kappa Variable 3-20 (IGKV3-20)	520, 700, 850, 1080, 1240, 1400	[Khan et al., 2016]
Immunoglobulin Heavy Constant Gamma 1 (IGHG1)	540, 720, 880, 1040, 1300, 1500	[Schroeder and Cavacini, 2010a]

Increasing Trend in MNG

Table 5.5: Functions and Pathologies for Proteins with Increasing Trend in Meningioma

Protein	Functions	Pathologies
IGLV2-14	Part of the immunoglobulin lambda light chain involved in the adaptive immune response	Associated with various immune disorders and cancers (B cell malignancies) [Collins and Jackson, 2021]
Biotinidase (BTD)	Involved in the recycling of biotin, an essential vitamin for carboxylase function	Deficiency leads to metabolic disorders; elevated levels associated with certain cancers and metabolic dysregulation [Wolf, 2011]
Lipopolysaccharide Binding Protein (LBP)	Binds bacterial lipopolysaccharides and presents them to immune cells	Associated with inflammatory conditions, sepsis, and cancer-related inflammation [Schumann et al., 1990]

Table 5.6: Raman/SERS Peaks for Proteins with Increasing Trend in Meningioma

Protein	Raman/SERS Peaks (cm^{-1})	References
IGLV2-14	500, 640, 800, 1040, 1280, 1440	[Collins and Jackson, 2021]
Biotinidase (BTD)	600, 720, 850, 1100, 1330, 1550	[Wolf, 2011]
Lipopolysaccharide-Binding Protein (LBP)	520, 680, 820, 1050, 1300, 1450	[Schumann et al., 1990]

Decreasing Trend in MNG

Table 5.7: Functions and Pathologies for Proteins with Decreasing Trend in Meningioma

Protein	Functions	Pathologies
Kininogen-1 (KNG1)	Involved in regulation of blood coagulation and inflammation	Associated with impaired coagulation and inflammatory responses [Colman, 2006]
Apolipoprotein L1 (APOL1)	Involved in lipid transport and innate immunity	Mutations linked to kidney disease and cardiovascular diseases; decreased levels may affect tumor cell metabolism and immune response [Genovese et al., 2010]

IGLV1-47	Part of the immunoglobulin lambda light chain involved in the adaptive immune response	Associated with various immune disorders and cancers (B cell malignancies) [Collins and Jackson, 2021]
Plasminogen (PLG)	Involved in the breakdown of fibrin in blood clots	Associated with impaired fibrinolysis and cancer progression [Cesarman et al., 1994]
Complement C1s (C1S)	Serine protease part of the complement system	Impaired immune response allowing tumor cells to evade immune surveillance [Noris et al., 2013]
Ficolin-3 (FCN3)	Involved in the lectin pathway of the complement system	Impaired immune function affecting cancer progression [Matsushita, 2012]

Table 5.8: Raman/SERS Peaks for Proteins with Decreasing Trend in Meningioma

Protein	Raman/SERS Peaks (cm^{-1})	References
Kininogen-1 (KNG1)	540, 700, 860, 1080, 1350, 1580	[Colman, 2006]
Apolipoprotein L1 (APOL1)	500, 660, 820, 1020, 1280, 1450	[Genovese et al., 2010]
IGLV1-47	620, 760, 900, 1150, 1330, 1500	[Collins and Jackson, 2021]
Plasminogen (PLG)	540, 720, 880, 1100, 1320, 1480	[Cesarman et al., 1994]

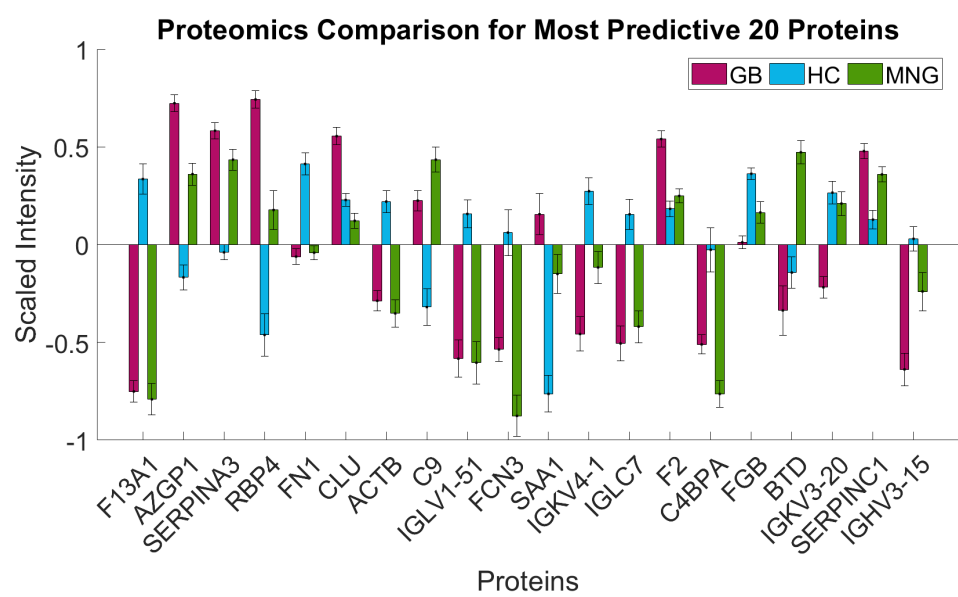


Figure 5.6: Proteomics Comparison for the Most Predictive 20 Proteins

Complement (C1S)	C1s	600, 740, 880, 1140, 1360, 1520	[Noris et al., 2013]
Ficolin-3 (FCN3)		520, 700, 850, 1080, 1260, 1400	[Matsushita, 2012]

Chapter 6

SURFACE ENHANCED RAMAN SPECTROSCOPY ANALYSIS OF PLASMA EXOSOMES

This chapter details the methodology of Surface-Enhanced Raman Spectra measurements and computational analysis used to predict patient conditions from Surface-Enhanced Raman Spectroscopy (SERS) profiles of brain tumor plasma exosome samples. Our study focused on identifying three states: healthy control (HC), glioblastoma (GB), and meningioma (MNG). The computational analysis involved extensive pre-processing of the SERS profiles, distribution analysis of the processed spectra, and prediction of patient states using various AI models. We conducted this analysis on a comprehensive dataset comprising 32,193 spectra, ensuring a balanced representation across the three classes.

Dr. Ugur Parlattan and Chris Nguyen are acknowledged for their support in Raman spectra acquisition. Tim Valencony is greatly acknowledged for his efforts in spectra pre-processing and AI-classification of the SERS spectra. The spectra pre-processing and AI-classification were conducted by Tim Valencony, following continuous discussions and under the guidance of Dr. Ugur Parlattan.

6.1 Motivation for Raman Spectroscopy in Gliomas

Raman spectroscopy is a powerful analytical tool widely used to characterize biological tissues and materials due to its non-invasive nature and high sensitivity to molecular composition [Kneipp et al., 1997]. In the context of gliomas, Raman spectroscopy offers several advantages. It provides detailed molecular fingerprints of tissues, enabling the identification of biochemical changes associated with tumorigenesis [Sahu et al., 2013]. This technique is particularly useful for distinguishing

between healthy and cancerous tissues based on their unique spectral signatures [Zhang et al., 2013]. Additionally, Raman spectroscopy can be combined with advanced computational methods, such as machine learning algorithms, to enhance the accuracy and reliability of tumor classification [Shipp et al., 2019]. By analyzing the subtle variations in Raman spectra, researchers can detect early-stage tumors, monitor disease progression, and evaluate treatment responses [Liu et al., 2019]. Overall, using Raman spectroscopy in glioma analysis facilitates precise and real-time diagnosis, making it a valuable tool in the clinical management of brain tumors.

6.2 Spectra Pre-Processing

Effective pre-processing of SERS profiles is crucial to optimize the performance of AI models. We implemented several pre-processing steps using the RamanSPy library in Python to enhance the spectral data quality. These steps are designed to remove noise and irrelevant information, thereby improving the accuracy of the subsequent analysis.

6.2.1 Raw Spectra Spectra Cropping

The first step in pre-processing involved cropping the similarly repeating parts off every raw spectrum. These redundant sections, which are not unique and are found in every spectrum regardless of the state class, were removed. By focusing on the distinctive parts of the spectra, we aimed to highlight the unique molecular signatures associated with each state, facilitating more accurate classification.

6.2.2 Cosmic Ray Removal

Cosmic rays can introduce extraneous spikes in the spectra, distorting the data and potentially leading to incorrect interpretations. We employed the Whitaker-Hayes algorithm to address this issue, effectively identifying and eliminating these unwanted spikes. This algorithm differentiates between the sharp, short-lived spikes

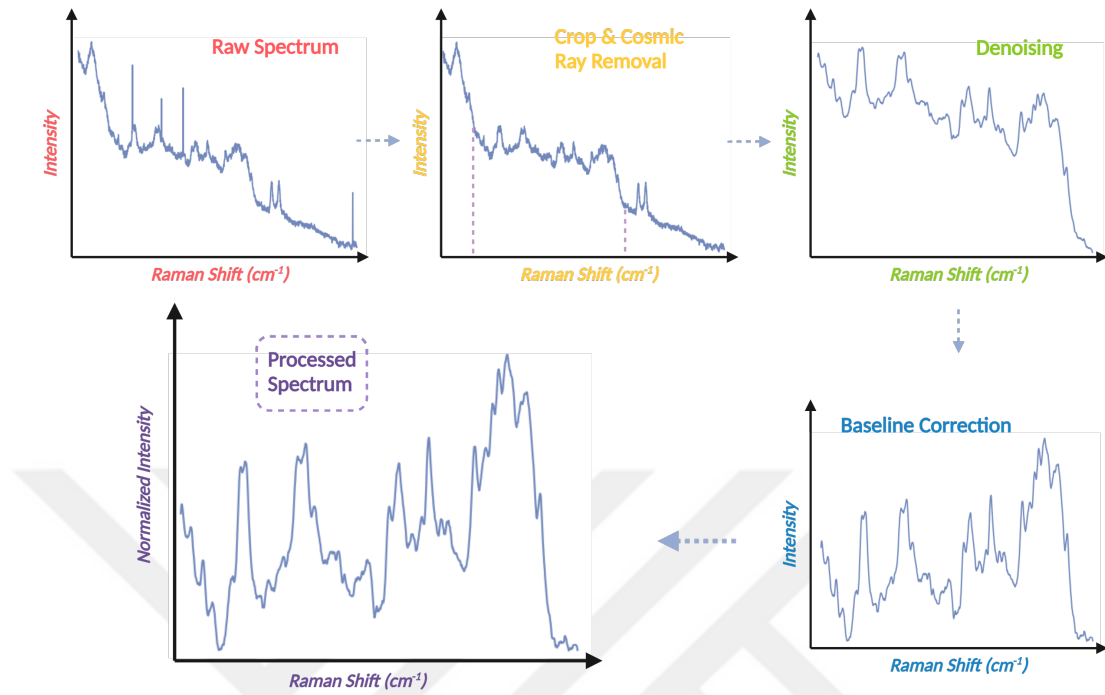


Figure 6.1: Schematic illustration of pre-processing steps for Surface-Enhanced Raman Spectra analysis: Acquiring raw spectra from the plasma exosomes of all study participants, cropping the spectra to narrow it down to the fingerprint region, denoising the spectra using Savitzky-Golay filtering, baseline correction and vector normalization. Created with BioRender.com

caused by cosmic rays and the smoother, longer-lasting spectral features.

6.2.3 Spectra Smoothing-Denoising

Noise reduction is a critical aspect of pre-processing, as it enhances the signal-to-noise ratio (SNR) and preserves the integrity of the spectral data. We applied the Savitzky-Golay algorithm, which is a smoothing filter widely used in spectroscopy. This algorithm fits successive subsets of adjacent data points with a low-degree polynomial by the method of linear least squares, effectively reducing noise while preserving important spectral features.

6.2.4 Baseline Correction

Each spectrum contains a natural fluorescence that can obscure the true Raman signal, affecting the accuracy of the analysis. We used an improved asymmetric least squares baseline correction algorithm to correct this. This algorithm adjusts the baseline of each spectrum, removing the extra fluorescence and enabling a clearer comparison of the spectra by focusing on the actual molecular vibrations rather than background noise.

6.2.5 Spectra Normalization

Normalization is essential for ensuring that the intensity of each spectrum is adjusted to a common scale. This step facilitates easier comparison and analysis of the Raman profiles by eliminating intensity variations unrelated to the actual differences between the classes. By normalizing the spectra, we ensured that the differences observed were due to the intrinsic properties of the samples rather than extraneous factors.

6.3 Spectral Analysis

6.3.1 Distribution Analysis using t-SNE

Following pre-processing, we analyzed the data distribution using a dimensionality reduction technique called t-Distributed Stochastic Neighbor Embedding (t-SNE). This technique is particularly useful for visualizing high-dimensional data in a lower-dimensional space. We performed t-SNE analysis to each group separately to see the spectra distribution for each study participant, then we performed another t-SNE analysis for all groups together to see the distribution of all data among different pathological conditions.

6.3.2 Visualization of Spectral Data

t-SNE is a non-linear dimensionality reduction technique that embeds high-dimensional data in a two or three-dimensional space. The algorithm minimizes the Kullback-

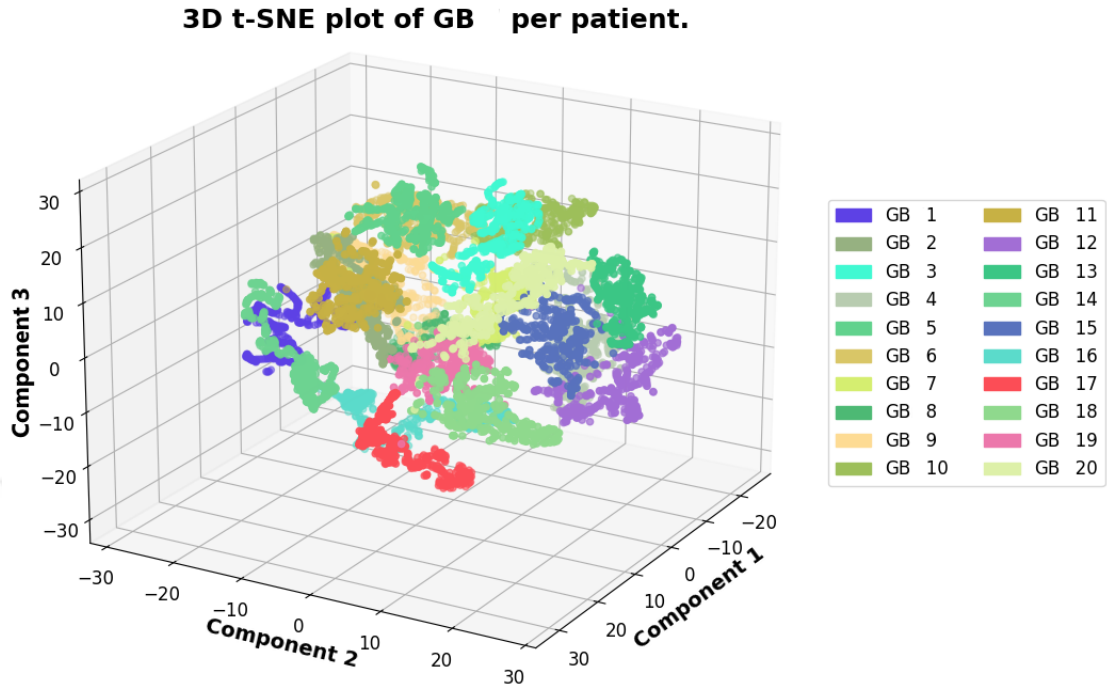


Figure 6.2: t-SNE analysis of glioblastoma (GB) exosome SERS spectra (n=20). Each number refers to a different patient

Leibler divergence between the probability distributions of points in the high-dimensional space and their corresponding low-dimensional representations. This visualization helps identify clusters of similar spectra and understand the separations between different classes. In our study, t-SNE clearly represented the class distributions, highlighting clusters corresponding to HC, GB, and MNG states.

6.3.3 Average Spectra of Plasma Exosomes

The average spectra of plasma exosomes provide a comprehensive overview of the typical Raman shifts and intensities observed in the samples from each class: healthy control (HC), glioblastoma (GB), and meningioma (MNG). By averaging the spectra, we can highlight the common spectral features and identify the characteristic peaks associated with each condition. This section will detail the methodology used to compute the average spectra and discuss the key spectral differences observed among the classes.

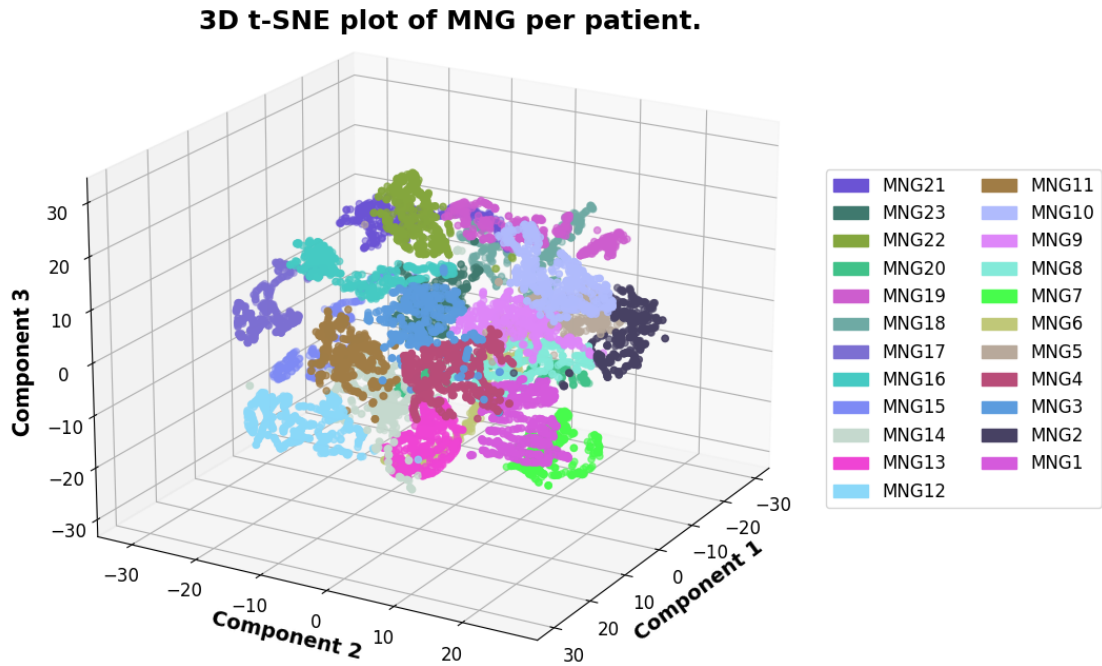


Figure 6.3: t-SNE analysis of meningioma (MNG) exosome SERS spectra (n=24). Each number refers to a different patient

6.3.4 Methodology

To compute the average spectra, we first aligned the individual spectra to ensure consistency in the Raman shift axis. This alignment is crucial to account for slight variations in the measurement conditions and to ensure that the averaging process reflects the true underlying spectral features. After alignment, we averaged the intensity values at each Raman shift across all spectra within each class. This process involved the following steps:

Spectral Alignment: Using a peak alignment algorithm to correct for any shifts in the Raman peaks. **Intensity Averaging:** Calculating the mean intensity at each Raman shift for all spectra within a class.

The resulting average spectra for HC, GB, and MNG samples exhibit distinct patterns that reflect the molecular composition of the exosomes from each condition. Key observations include:

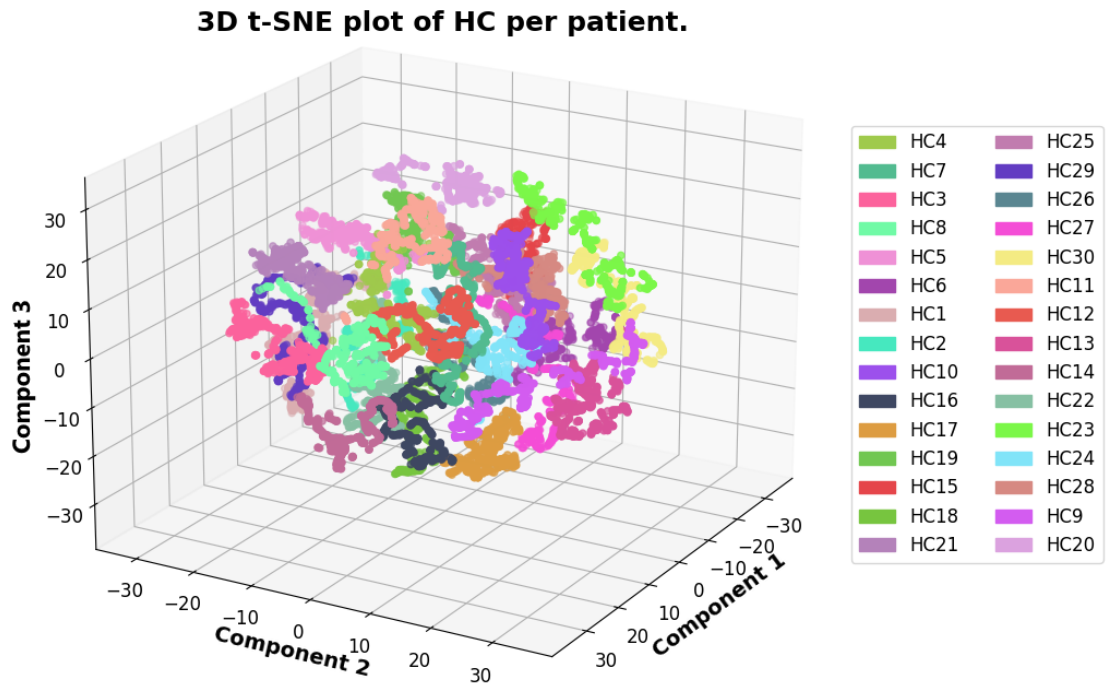


Figure 6.4: t-SNE analysis of healthy control (HC) exosome SERS spectra (n=30). Each number refers to a different study participant

- **HC Samples:** The average spectrum shows prominent peaks corresponding to normal cellular components, such as proteins and lipids.
- **GB Samples:** The spectrum is characterized by additional peaks and increased intensities at specific Raman shifts, indicating the presence of tumor-specific molecular markers.
- **MNG Samples:** The average spectrum displays unique features that differentiate it from both HC and GB, reflecting the benign nature of meningioma tumors.

The comparison of these average spectra provides a clear visual representation of the differences between the classes and serves as a foundation for further analysis and model development.

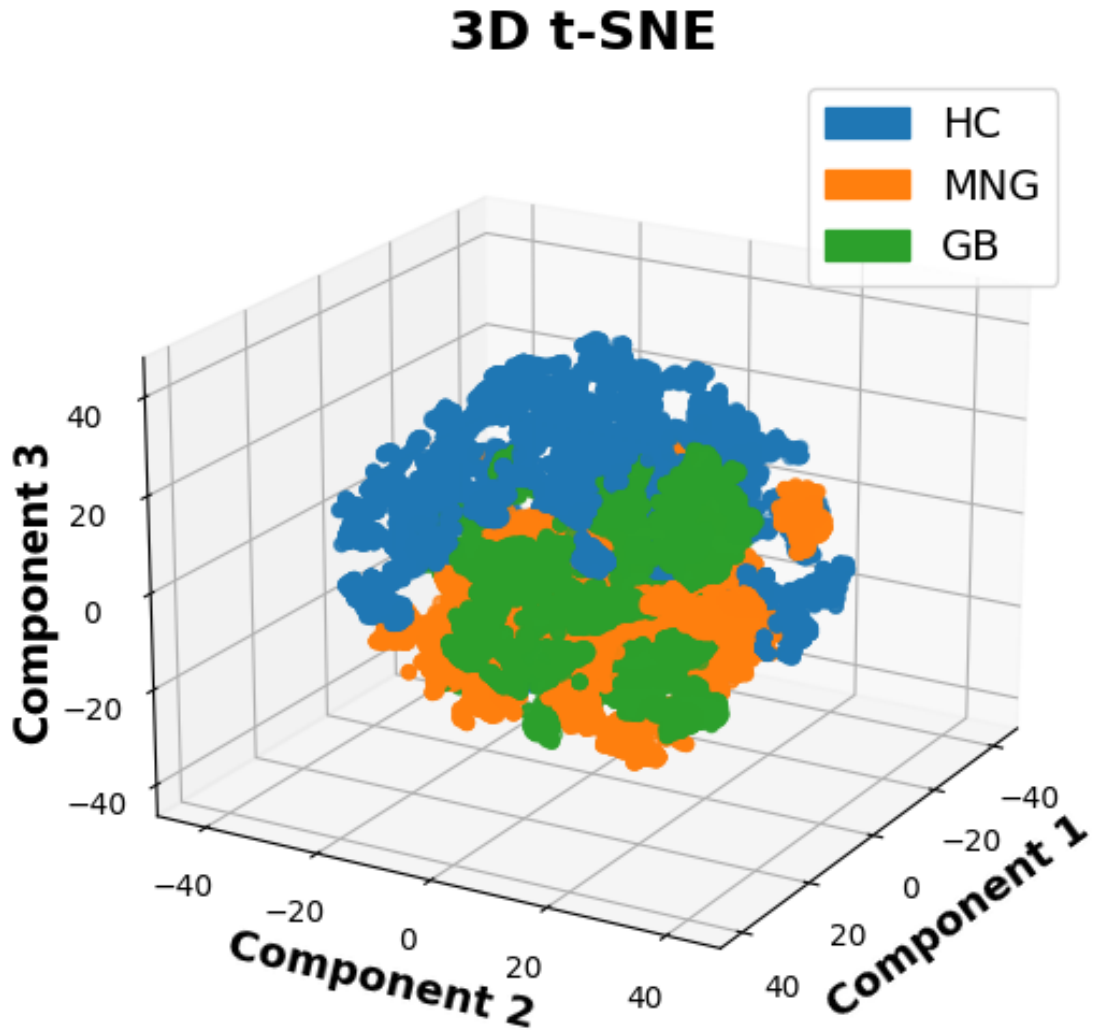


Figure 6.5: t-SNE analysis of all groups including glioblastoma (GB), meningioma (MNG), healthy control (HC) exosome averaged SERS spectra ($n_{GB} = 20$, $n_{MNG} = 24$, $n_{HC} = 30$)

6.4 Regression Analysis

Regression analysis is employed to explore the quantitative relationships between the Raman spectral features and the clinical conditions of the samples. This section will describe the regression models used, the feature selection process, and the outcomes of the analysis.

Regression Models: We utilized several regression models to predict continuous

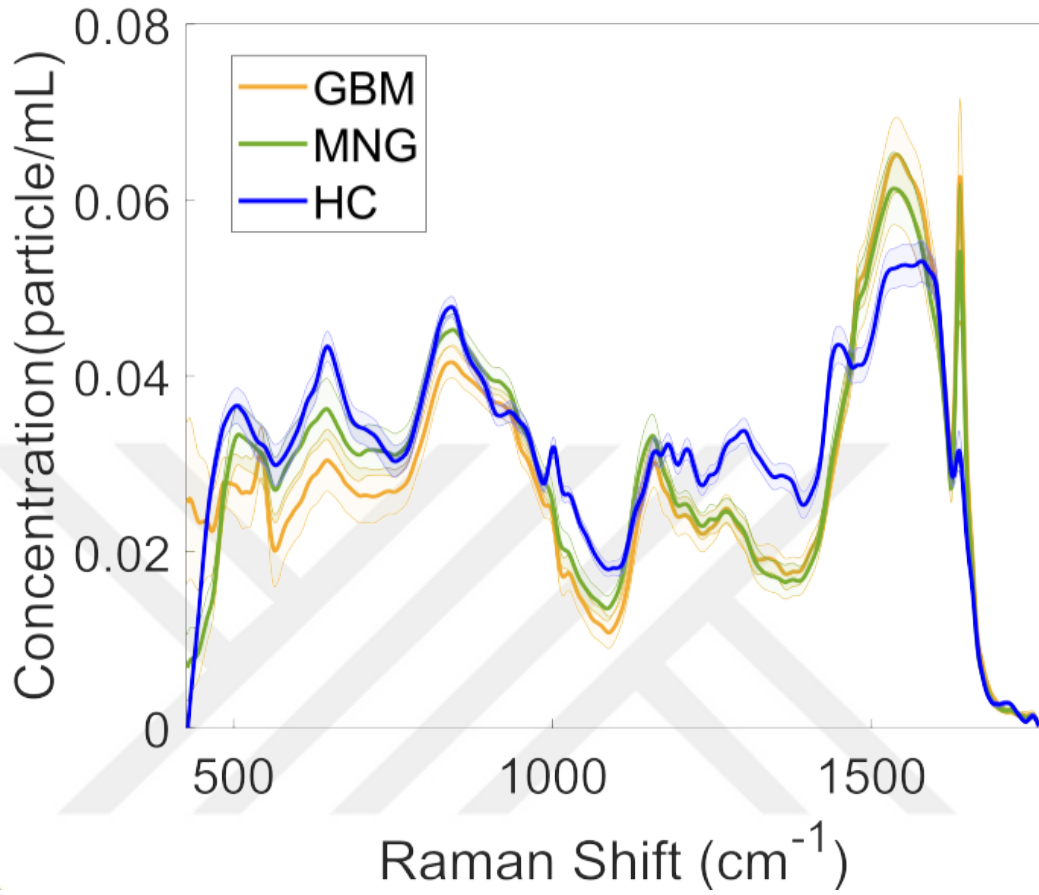


Figure 6.6: Average SERS spectra of glioblastoma (GB), meningioma (MNG), and healthy control (HC) exosomes ($n_{GB} = 20, n_{MNG} = 24, n_{HC} = 30$)

clinical variables from the spectral data. The models included linear regression, Lasso regression, and Ridge regression. Each model was evaluated for its ability to accurately predict variables such as tumor size, stage, and patient survival time based on the SERS spectra.

6.4.1 Feature Selection

Feature selection is a critical step in regression analysis, as it helps identify the most relevant spectral features that contribute to the predictive model. We employed methods such as:

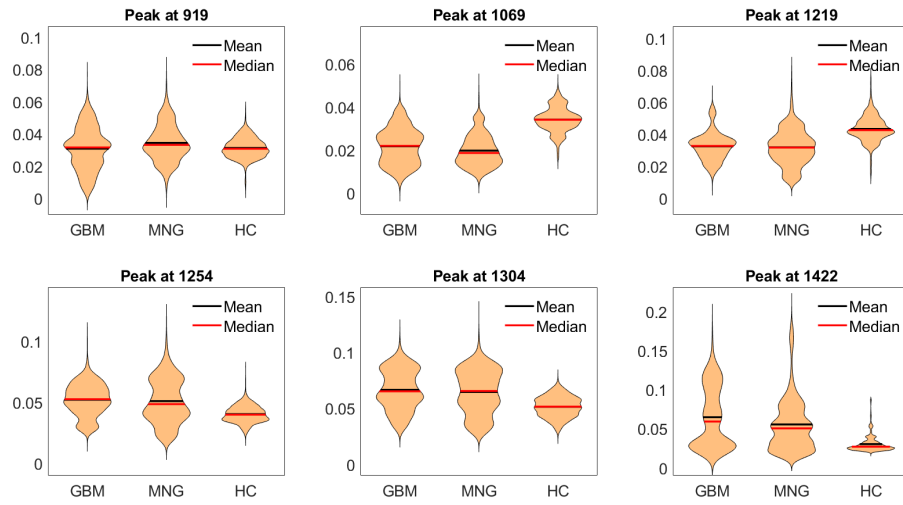


Figure 6.7: Violin plots of significant peaks differing in between three groups: glioblastoma (GB), meningioma (MNG), and healthy control (HC) exosomes ($n_{GB} = 20, n_{MNG} = 24, n_{HC} = 30$)

6.4.2 Principal Component Analysis (PCA)

To reduce dimensionality and identify the principal components that explain the most variance in the data.

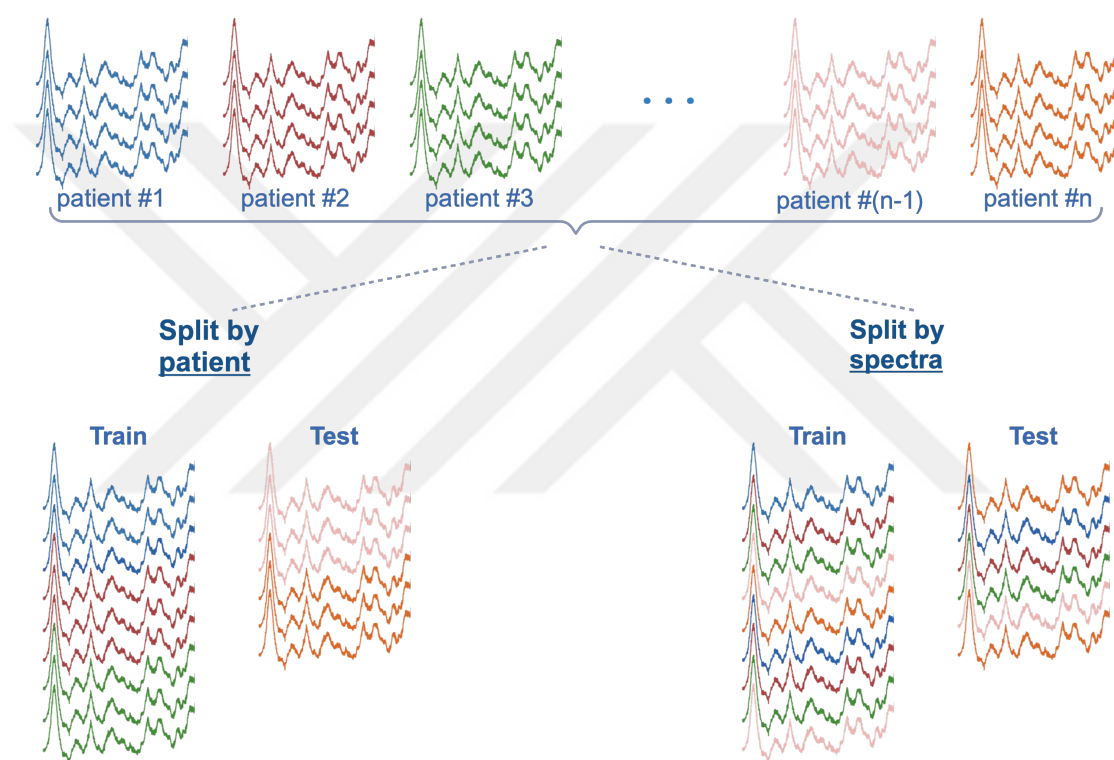
6.4.3 Recursive Feature Elimination (RFE)

To iteratively select the most important features by training the model and ranking the features based on their contribution to the prediction accuracy.

The regression models demonstrated varying degrees of success in predicting the clinical variables. Key findings include:

6.4.4 Linear Regression

Provided a baseline performance, with moderate accuracy in predicting the distribution of the specific Raman peak intensity in the data.



Different ways of splitting the data for the training and testing steps of the model.

Figure 6.8: Different type of train-test splits: (a) Spectra-wise classification and (b) Patient-wise classification, which is more suitable for clinical setting

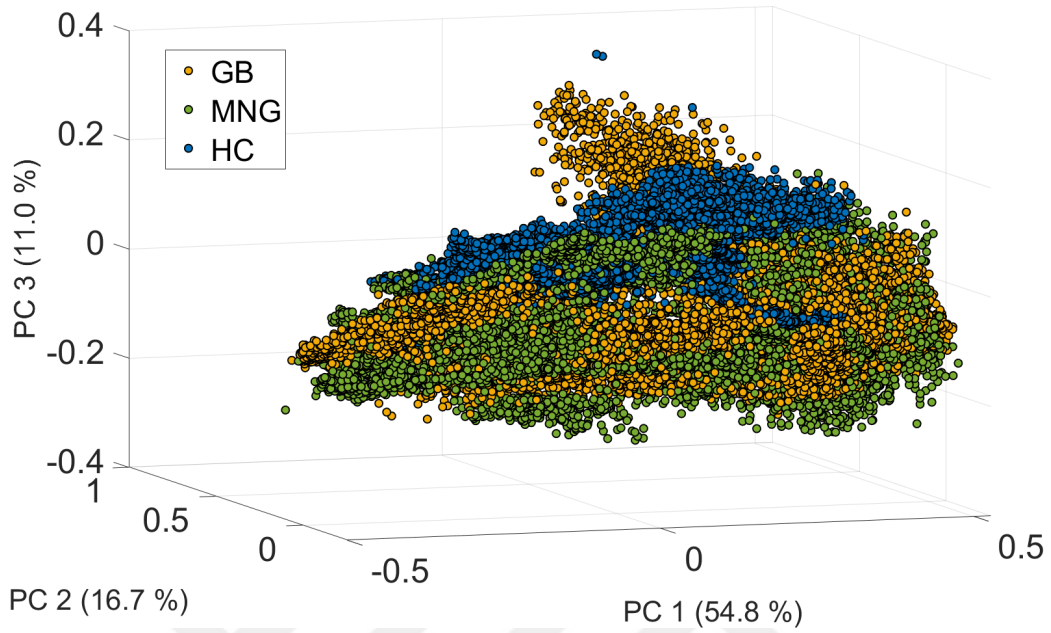


Figure 6.9: Principal Component Analysis (PCA) of glioblastoma (GB), meningioma (MNG), healthy control (HC) exosome averaged SERS spectra ($n_{GB} = 20, n_{MNG} = 24, n_{HC} = 30$)

6.4.5 Lasso Regression

Improved performance by enforcing sparsity in the feature set, highlighting the most significant Raman shifts.

6.4.6 Ridge Regression

Offered the best overall performance by balancing bias and variance, leading to robust predictions across different clinical variables. The regression analysis underscored the potential of SERS spectra as a non-invasive tool for quantitative assessment of tumor characteristics, providing valuable insights for personalized treatment planning.

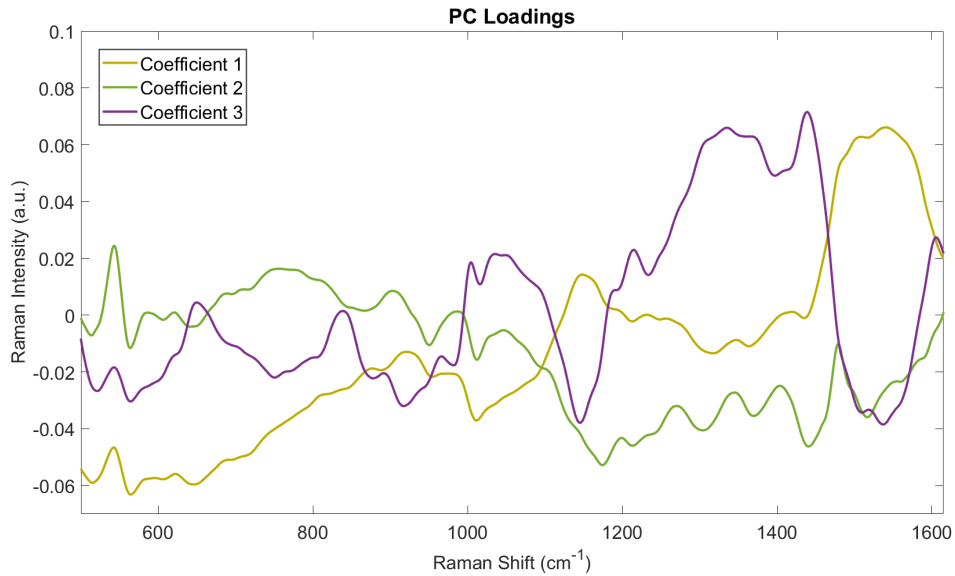


Figure 6.10: Principal Component (PC) coefficients of glioblastoma (GB), meningioma (MNG), healthy control (HC) exosome averaged SERS spectra ($n_{GB} = 20$, $n_{MNG} = 24$, $n_{HC} = 30$)

6.5 Proteomics and SERS Correlation

We correlated our findings in proteomics and sERS analysis in this section. We found Raman or SERS peaks for the distinguishing proteins and searched for these peaks in our SERS findings. Based on our results in proteomics and SERS analyses, we calculated correlation coefficients to understand which Raman/SERS peaks have been attributed to the relevant distinguishing peaks in our findings (See Tables 5.1 - 5.8 in Chapter 5 for SERS peaks attributed to the relevant proteins used in this analysis).

The proteomic analysis of exosomes offered promising insights into the molecular mechanisms underlying neurodegenerative diseases. This study aimed to identify novel biomarkers and therapeutic targets by focusing on malign and benign tumor-related proteins. The correlation of the proteomics results with SERS analyses was examined to better understand the protein signatures identified by both techniques.

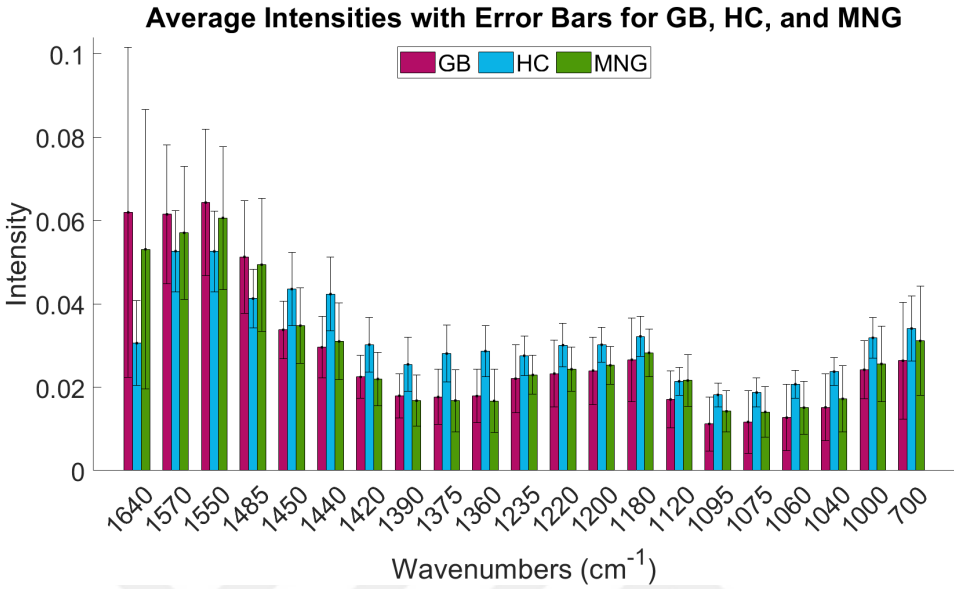


Figure 6.11: Average Intensities with Error Bars for Glioblastoma (GB), Meningioma (MNG), and Healthy Control (HC) Exosome SERS Signatures

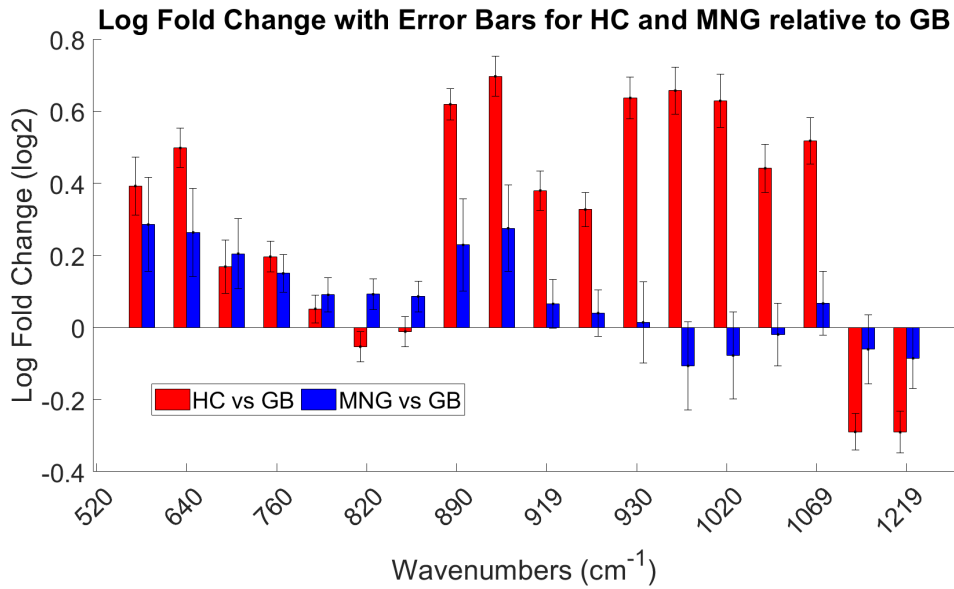


Figure 6.12: Log2Fold Change with Error Bars for Meningioma (MNG) and Healthy Control (HC) Exosome Raman Signatures Compared to Glioblastoma (GB)

Chapter 7

CLASSIFICATION OF BRAIN TUMOR EXOSOME SERS SPECTRA USING ARTIFICIAL INTELLIGENCE MODELS

7.1 *Prediction Models*

Patient states were predicted from the SERS spectra using six pre-identified machine-learning models. Each model was optimized and evaluated based on its performance in predicting the states from the spectral data.

7.1.1 *Quadratic Discriminant Analysis (QDA)*

Quadratic Discriminant Analysis (QDA) is a classification algorithm that models the distribution of each class using a quadratic decision boundary. Unlike linear methods, QDA allows for distinct covariance matrices for each class, providing more flexibility in modeling data. QDA computes discriminant functions to classify new data points based on their likelihood under each class distribution. This approach was used in our study to classify the SERS spectra with high accuracy. The quadratic decision boundaries allow QDA to capture more complex relationships within the data, enhancing classification performance.

7.1.2 *Random Forest (RF) Classifier*

Random Forest (RF) is an ensemble learning method that constructs multiple decision trees during training. Each tree is built using a random subset of the data and features, and the final prediction is made by aggregating the outputs of these individual trees. This approach enhances predictive accuracy and robustness by reducing overfitting and variance. In our study, RF was used to identify relevant

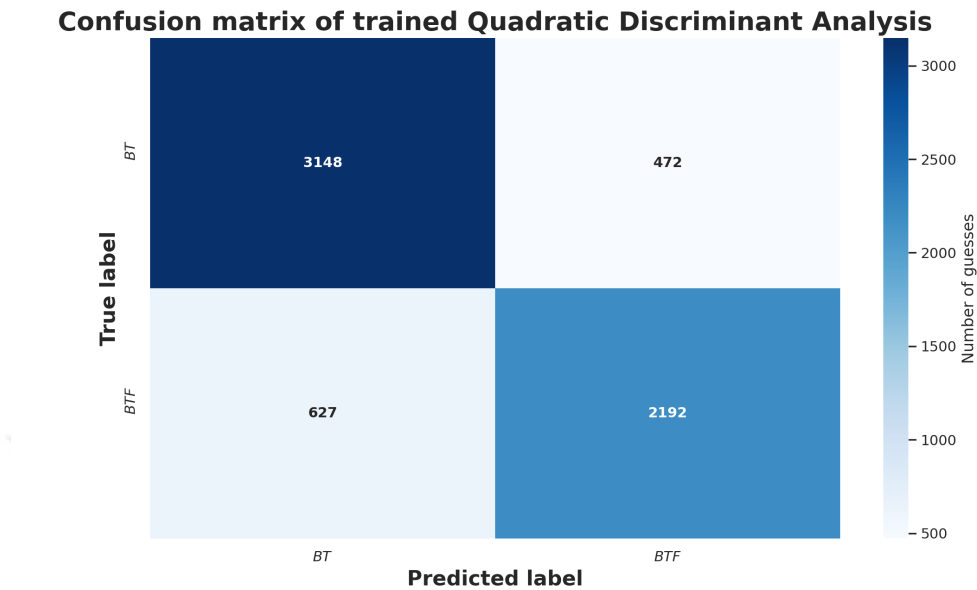


Figure 7.1: Confusion matrix for the first step classification to identify brain tumor free (BTF) individuals and patients with brain tumor (BT) using quadratic discriminant analysis (QDA) for spectra-wise classification

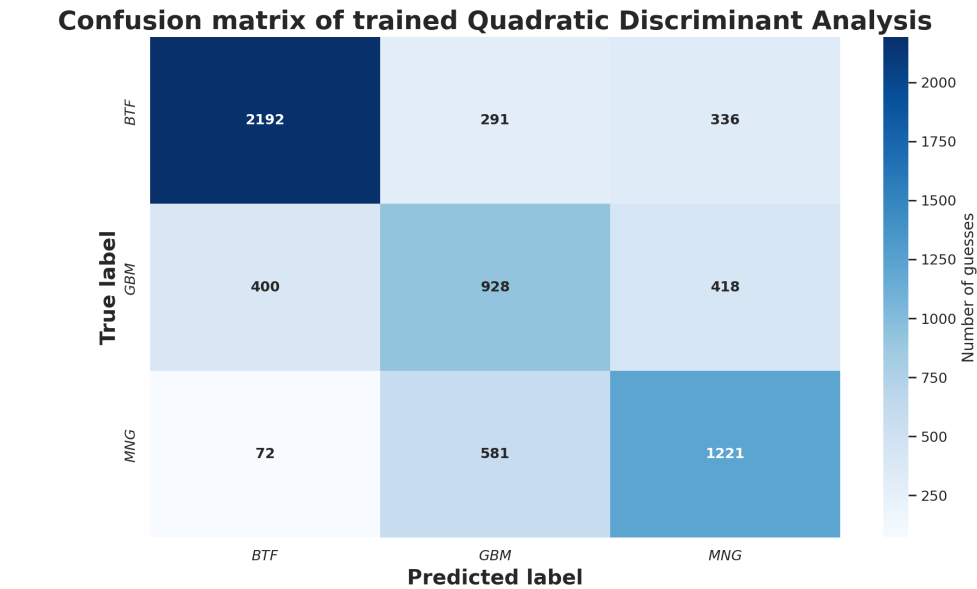


Figure 7.2: Confusion matrix for the second step classification to identify brain tumor free (BTF) individuals and patients with glioblastoma (GB), meningioma (MNG) using quadratic discriminant analysis (QDA) model for spectra-wise classification

Multiclass ROC curve for QDA model with AUC score: 0.7801617812452412

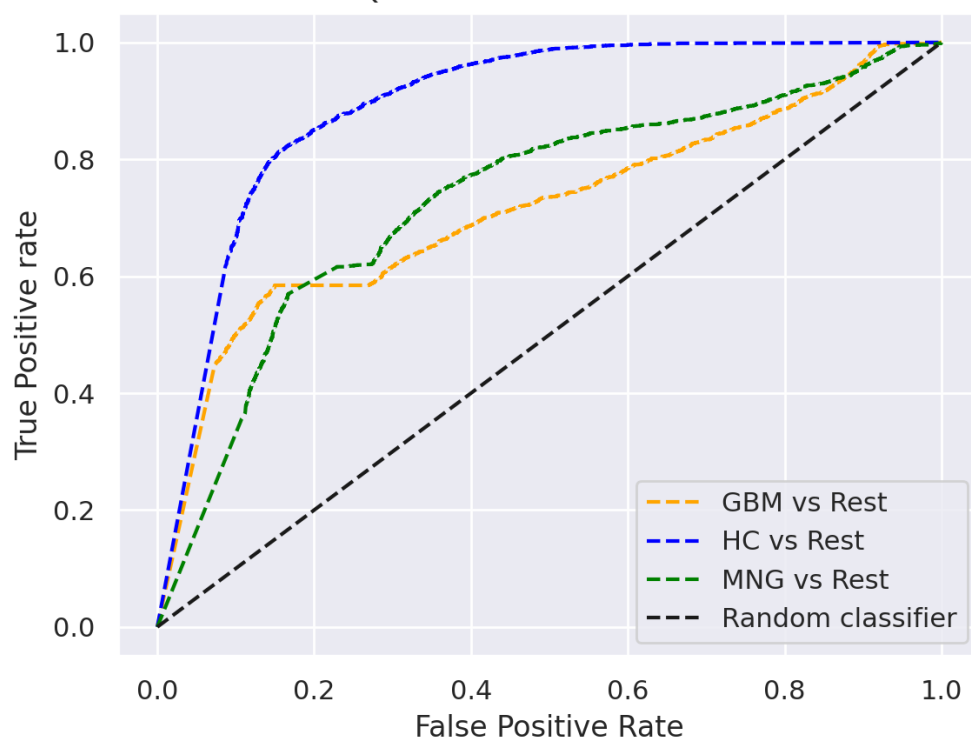


Figure 7.3: ROC Curve for the classification to identify brain tumor-free (BTF) individuals and patients with glioblastoma (GB) or meningioma (MNG) using quadratic discriminant analysis (QDA) model for spectra-wise classification

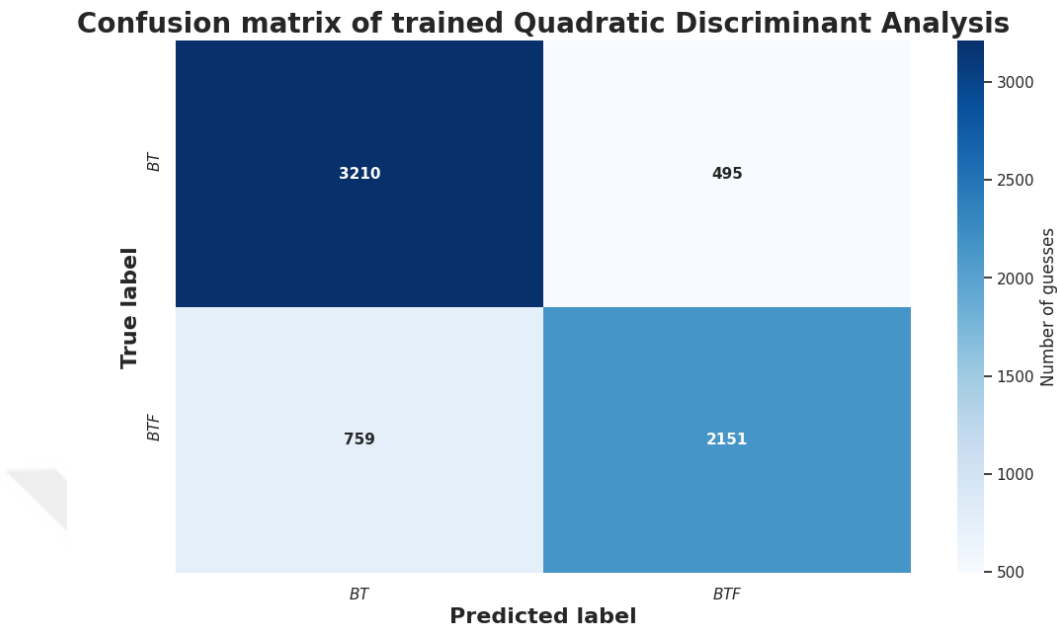


Figure 7.4: Confusion matrix for the first step classification to identify brain tumor free (BTF) individuals and patients with brain tumor (BT) using quadratic discriminant analysis (QDA) for patient-wise classification

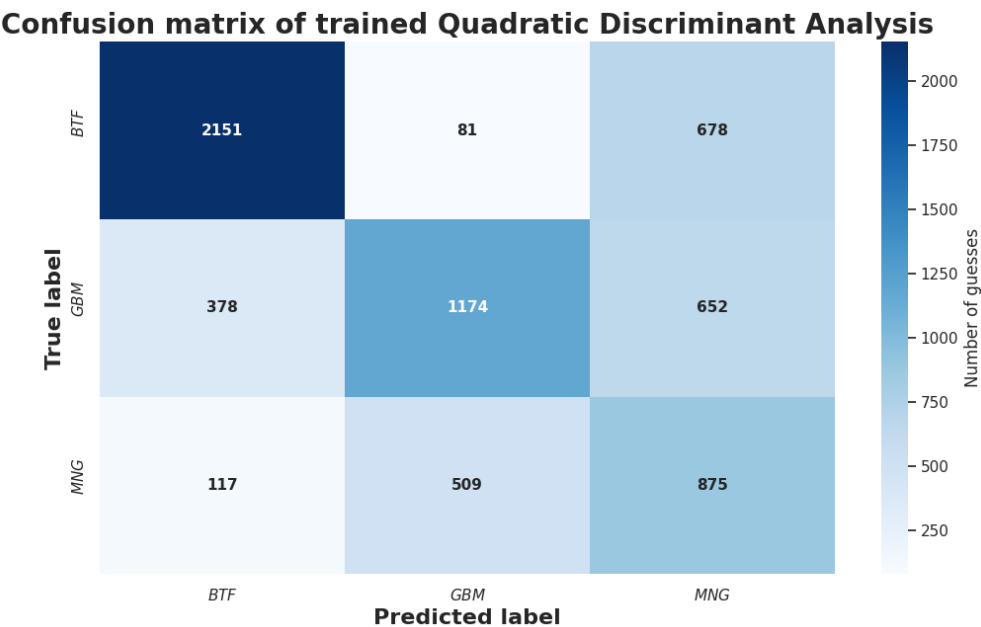


Figure 7.5: Confusion matrix for the second step classification to identify brain tumor free (BTF) individuals and patients with glioblastoma (GB), meningioma (MNG) using quadratic discriminant analysis (QDA) model for patient-wise classification

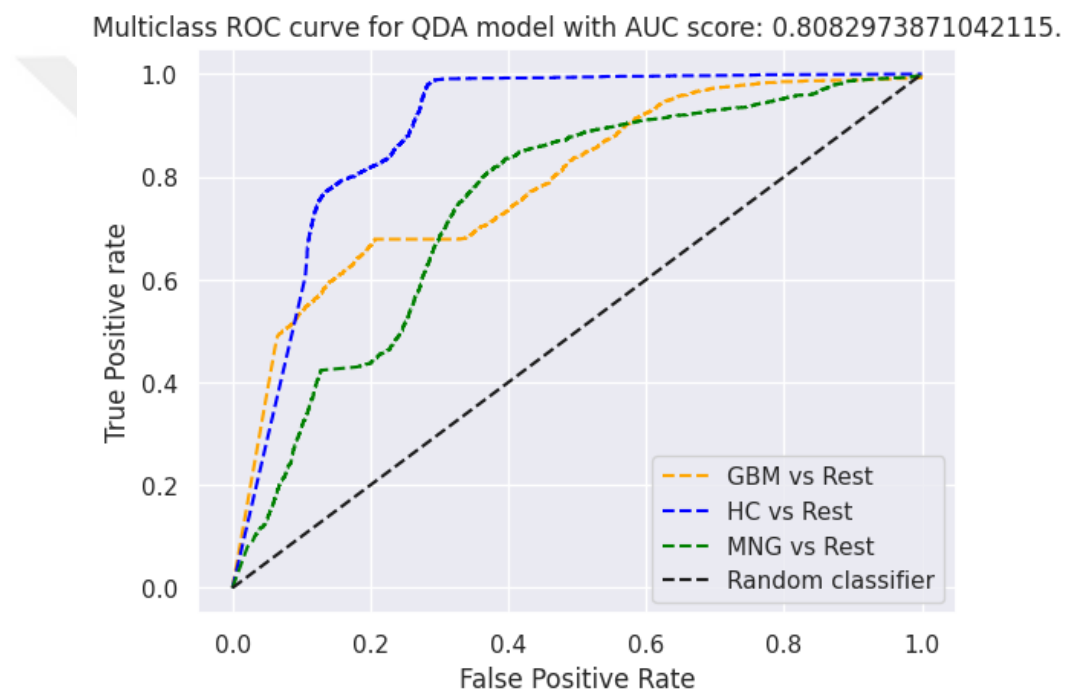


Figure 7.6: ROC Curve for the classification to identify brain tumor-free (BTF) individuals and patients with glioblastoma (GB) or meningioma (MNG) using quadratic discriminant analysis (QDA) model for patient-wise classification

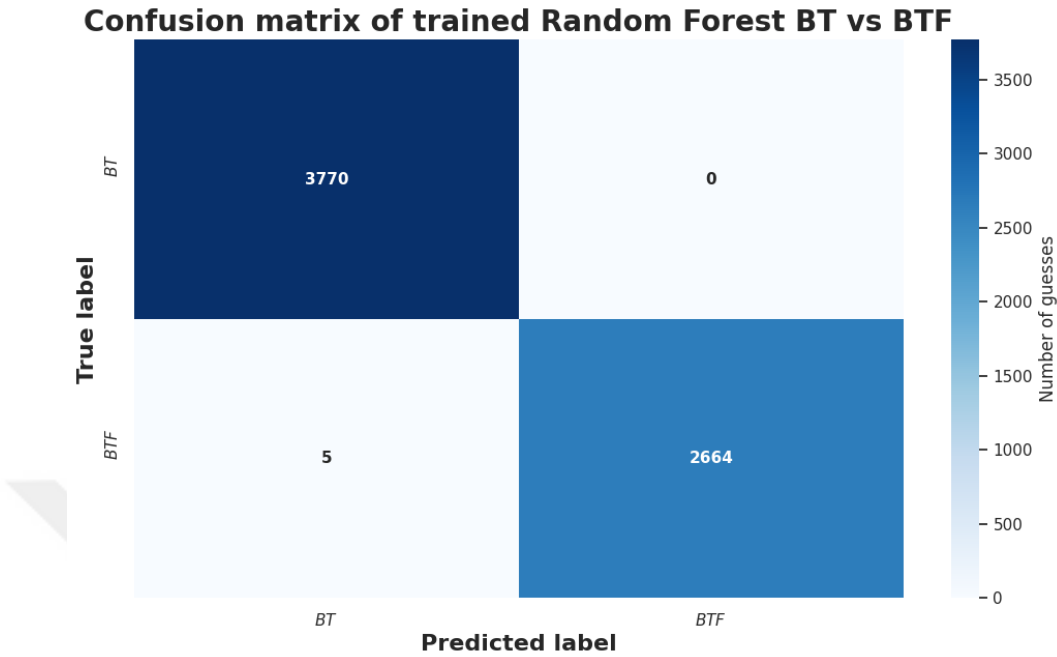


Figure 7.7: Confusion matrix for the first step classification to identify brain tumor free (BTF) individuals and patients with brain tumor (BT) using Random Forest (RF) for spectra-wise classification

differences between the spectra of each class and make accurate predictions based on these differences. The ensemble nature of RF helps capture the complex patterns in the data, providing reliable and interpretable results.

7.1.3 XG Boost (XGB)

XGBoost, short for Extreme Gradient Boosting, is an advanced implementation of gradient boosting algorithms designed for speed and performance. It is particularly effective for classification tasks because it can handle large datasets and complex patterns. This section will explain the principles of XGBoost, the hyperparameter tuning process, and its application in classifying SERS spectra.

Principles of XGBoost: XGBoost operates by constructing an ensemble of decision trees sequentially. Each new tree is built to correct the errors made by the previous trees, gradually improving the model’s accuracy. The key features of XGBoost include:

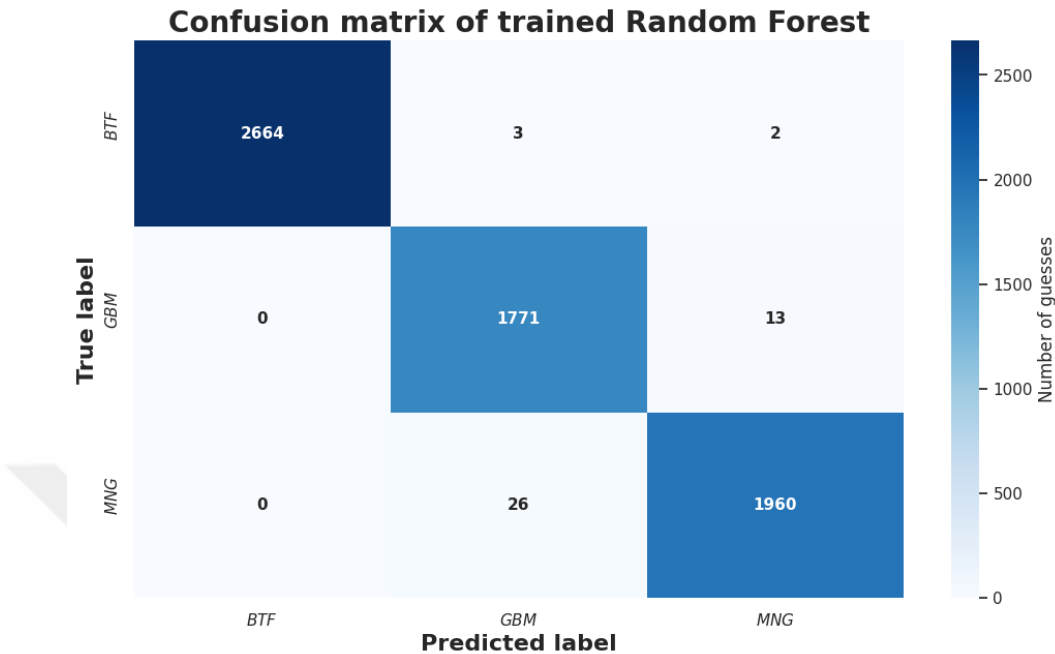


Figure 7.8: Confusion matrix for the second step classification to identify brain tumor free (BTF) individuals and patients with glioblastoma (GB), meningioma (MNG) using Random Forest (RF) model for spectra-wise classification

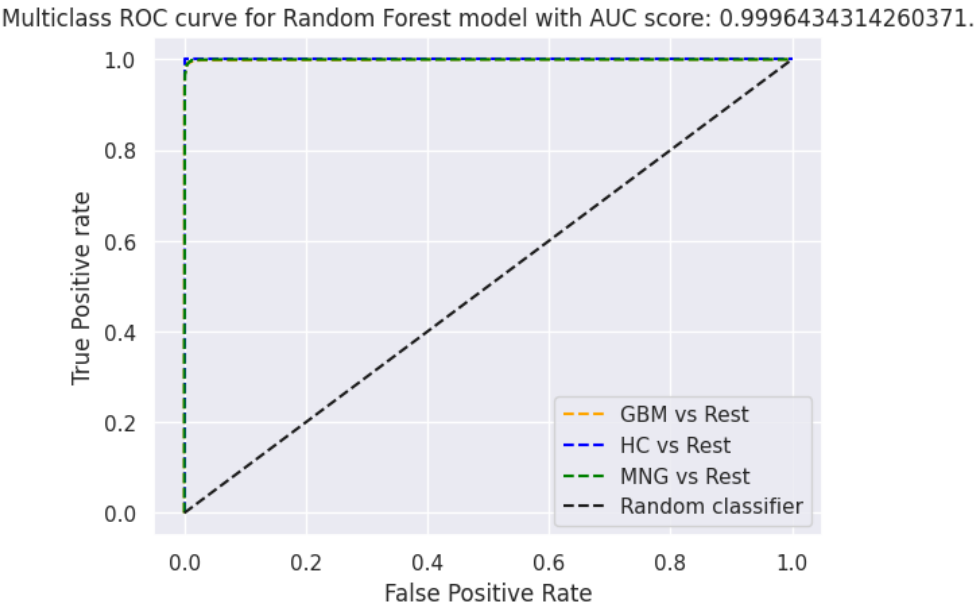


Figure 7.9: ROC Curve for the classification to identify brain tumor-free (BTF) individuals and patients with glioblastoma (GB) or meningioma (MNG) using Random Forest (RF) model for spectra-wise classification

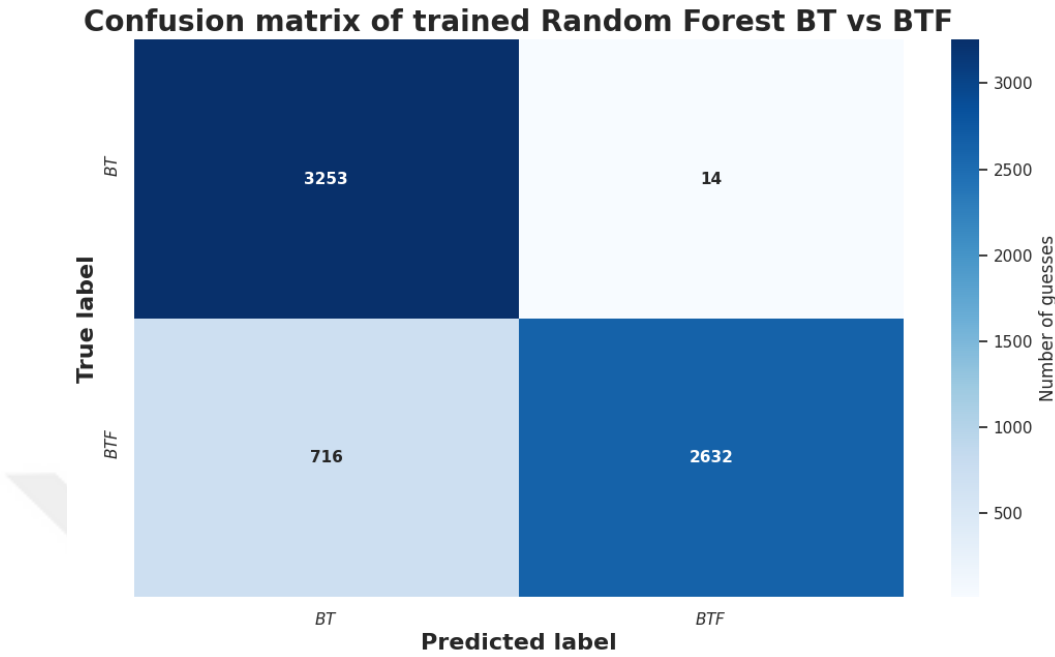


Figure 7.10: Confusion matrix for the first step classification to identify brain tumor free (BTF) individuals and patients with brain tumor (BT) using Random Forest (RF) for patient-wise classification

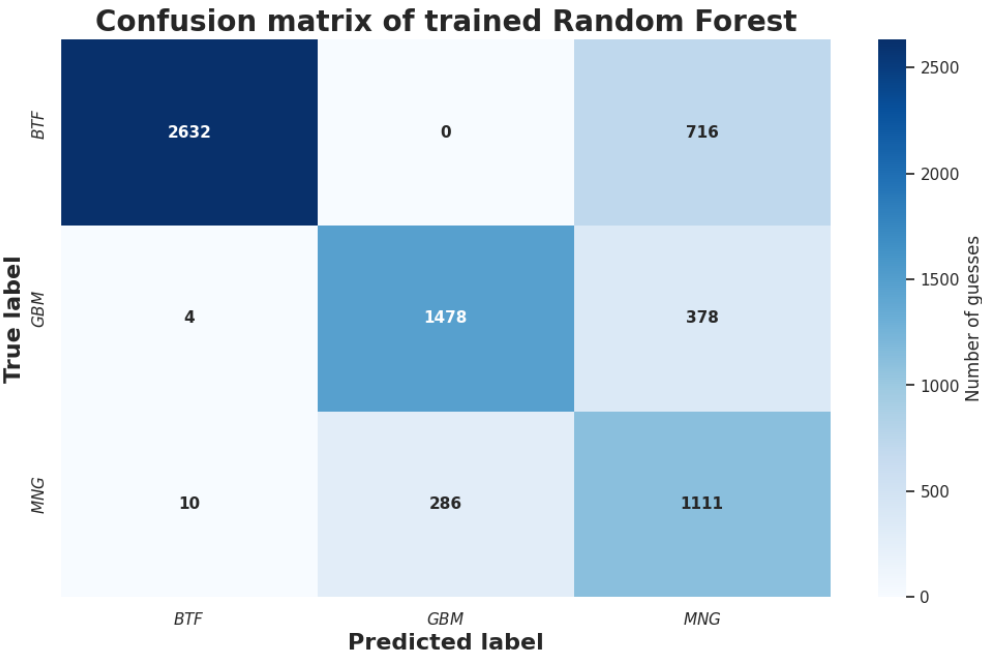


Figure 7.11: Confusion matrix for the second step classification to identify brain tumor free (BTF) individuals and patients with glioblastoma (GB), meningioma (MNG) using Random Forest (RF) model for spectra-wise classification

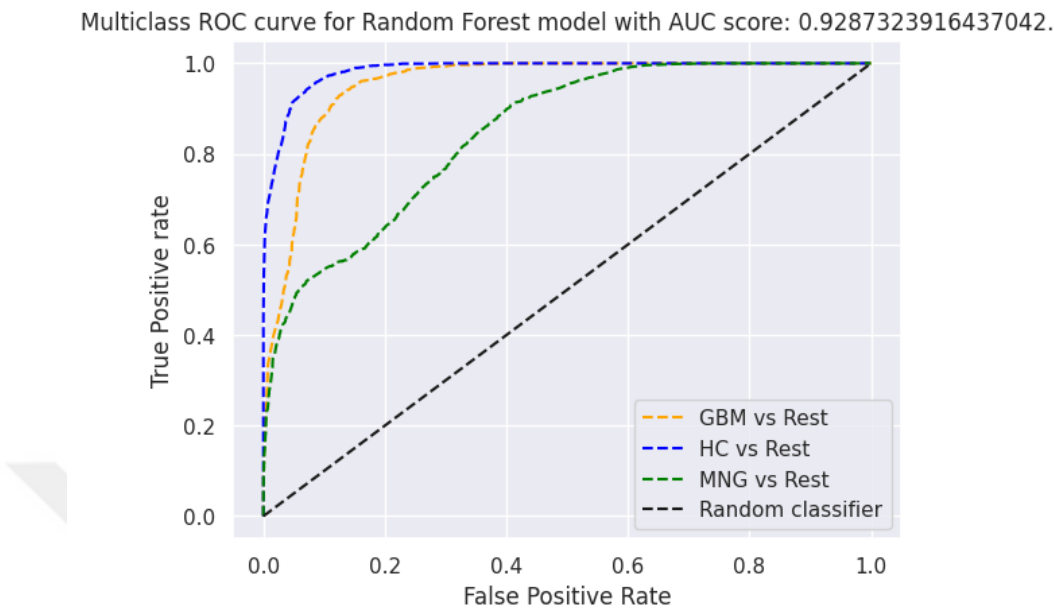


Figure 7.12: ROC Curve for the classification to identify brain tumor-free (BTF) individuals and patients with glioblastoma (GB) or meningioma (MNG) using Random Forest (RF) model for patient-wise classification

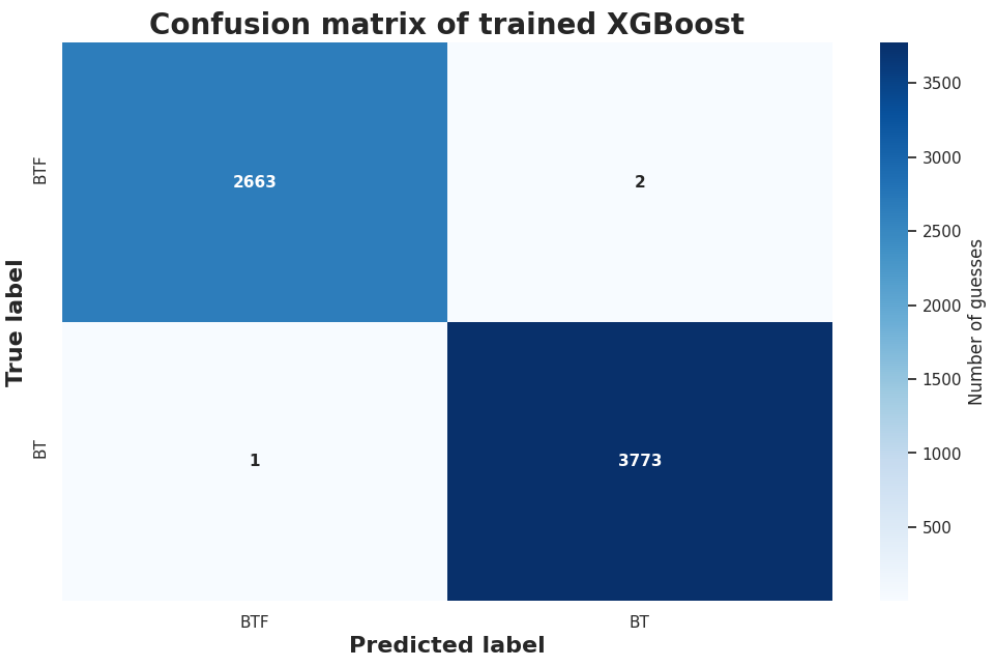


Figure 7.13: Confusion matrix for the first step classification to identify brain tumor free (BTF) individuals and patients with brain tumor (BT) using XG-Boost (XGB) model for spectra-wise classification

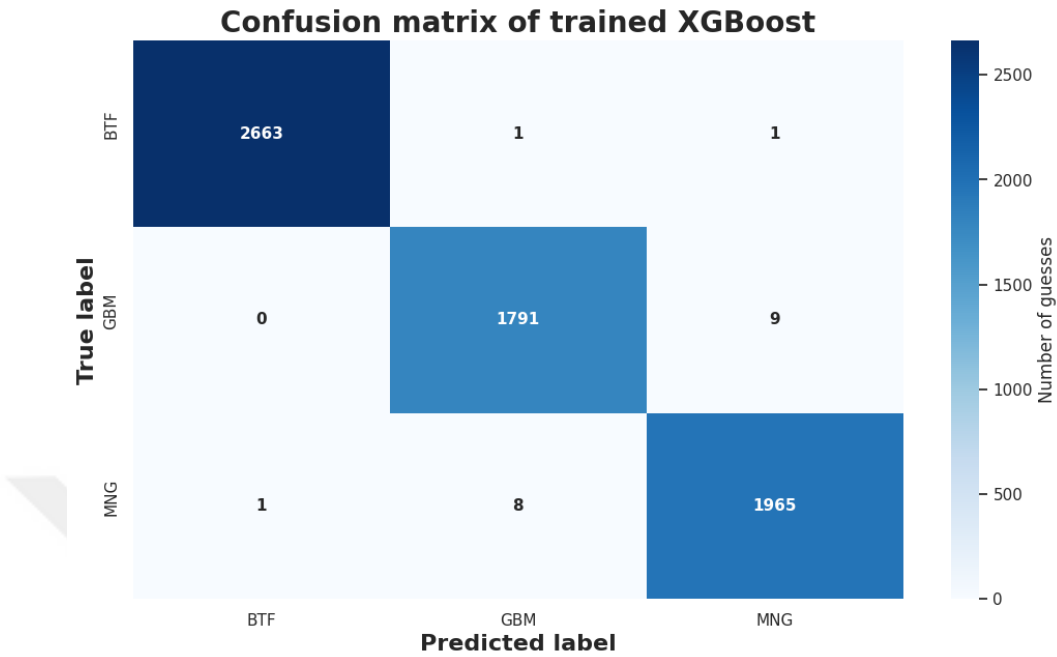


Figure 7.14: Confusion matrix for the second step classification to identify brain tumor free (BTF) individuals and patients with glioblastoma (GB), meningioma (MNG) using XG-Boost (XGB) model for spectra-wise classification

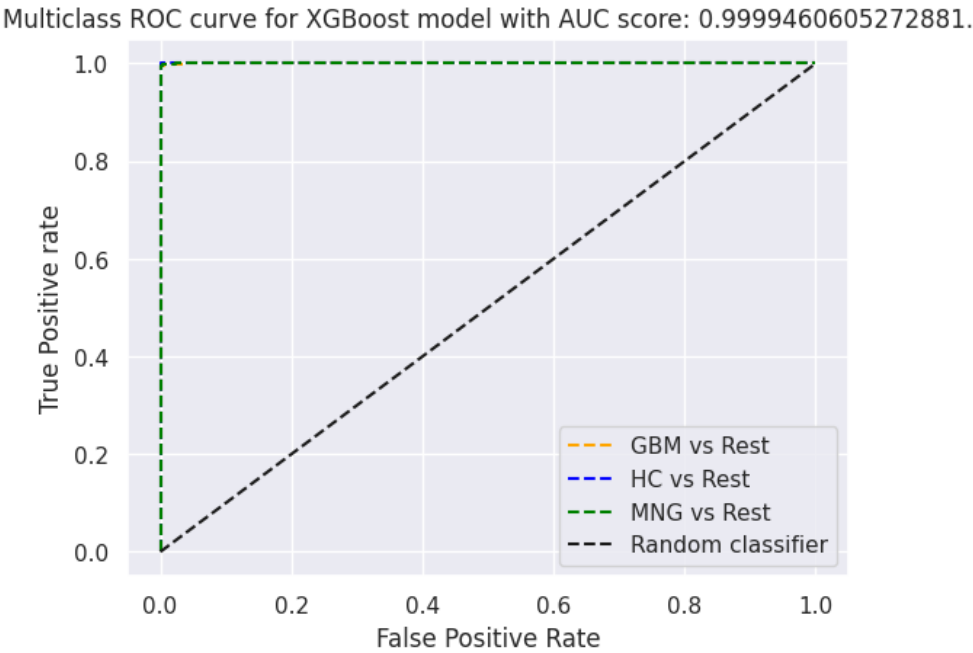


Figure 7.15: ROC Curve for the classification to identify brain tumor-free (BTF) individuals and patients with glioblastoma (GB) or meningioma (MNG) using XG-Boost (XGB) model for spectra-wise classification

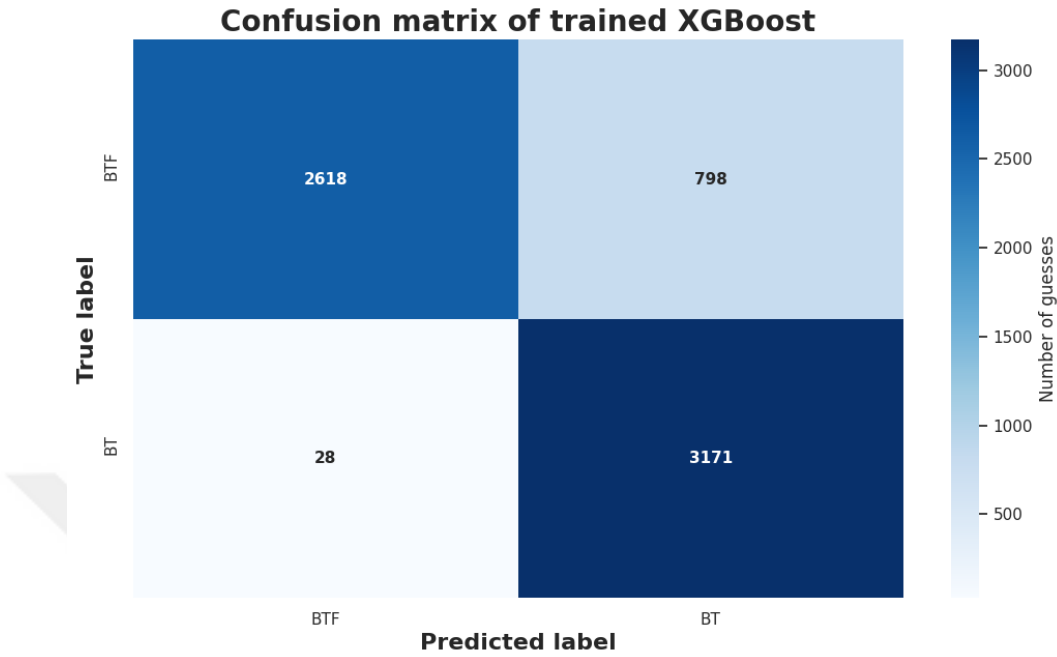


Figure 7.16: Confusion matrix for the first step classification to identify brain tumor free (BTF) individuals and patients with brain tumor (BT) using XG-Boost (XGB) model for patient-wise classification

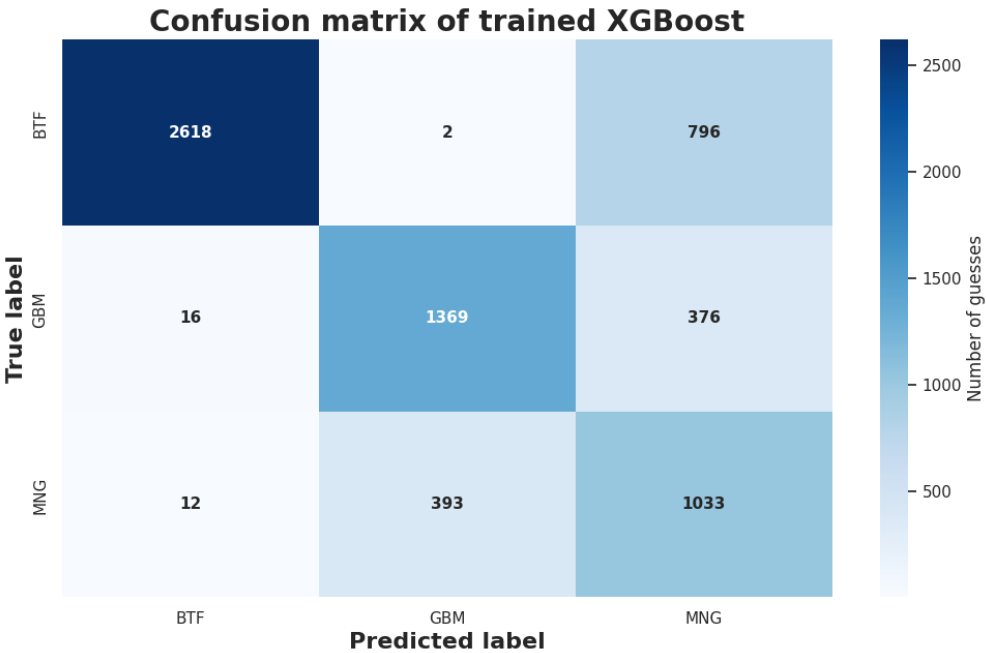


Figure 7.17: Confusion matrix for the second step classification to identify brain tumor free (BTF) individuals and patients with glioblastoma (GB), meningioma (MNG) using XG-Boost (XGB) model for spectra-wise classification

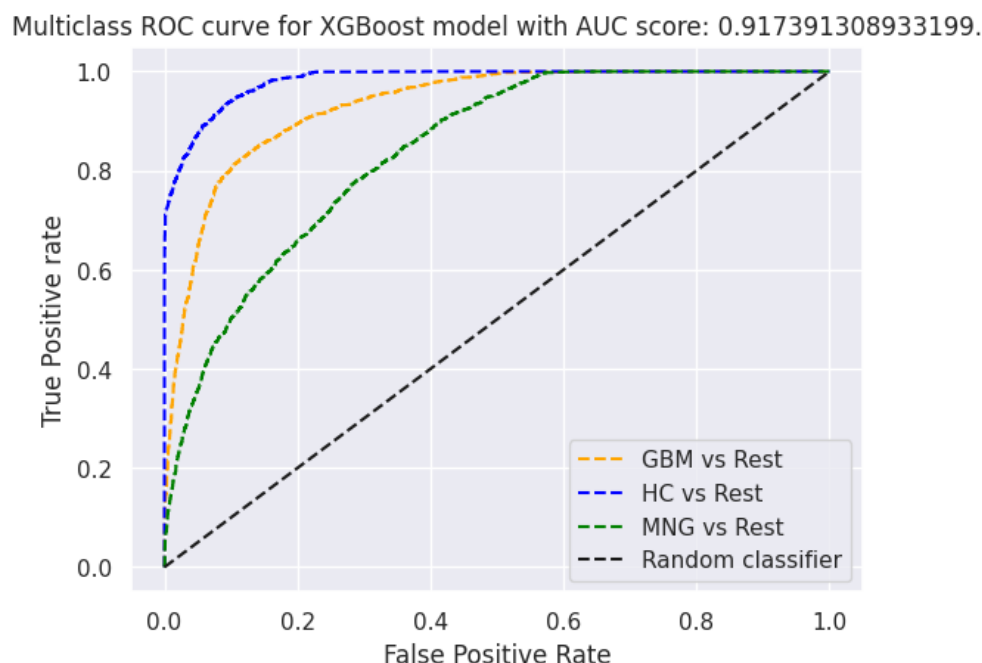


Figure 7.18: ROC Curve for the classification to identify brain tumor-free (BTF) individuals and patients with glioblastoma (GB) or meningioma (MNG) using XGBoost (XGB) model for patient-wise classification

Regularization: To prevent overfitting by penalizing complex models. Tree Pruning: To remove branches that do not contribute significantly to the model's performance. Handling Missing Data: Efficiently manages missing values during training. Hyperparameter Tuning Optimizing the hyperparameters of XGBoost is crucial for achieving high performance. The following parameters were tuned in our study: Learning Rate: Controls the step size at each iteration while moving towards a minimum of the loss function. Max Depth: Determines the maximum depth of the trees, balancing model complexity and overfitting. Subsample: The fraction of samples used for training each tree, which helps in reducing overfitting.

Application and Results XGBoost was applied to classify the SERS spectra into the three identified states: HC, GB, and MNG. The model was trained on 80% of the dataset and tested on the remaining 20%. The performance metrics included accuracy, precision, recall, and F1 score. The results indicated that XGBoost achieved high accuracy and robust performance, making it a suitable choice for this classifi-

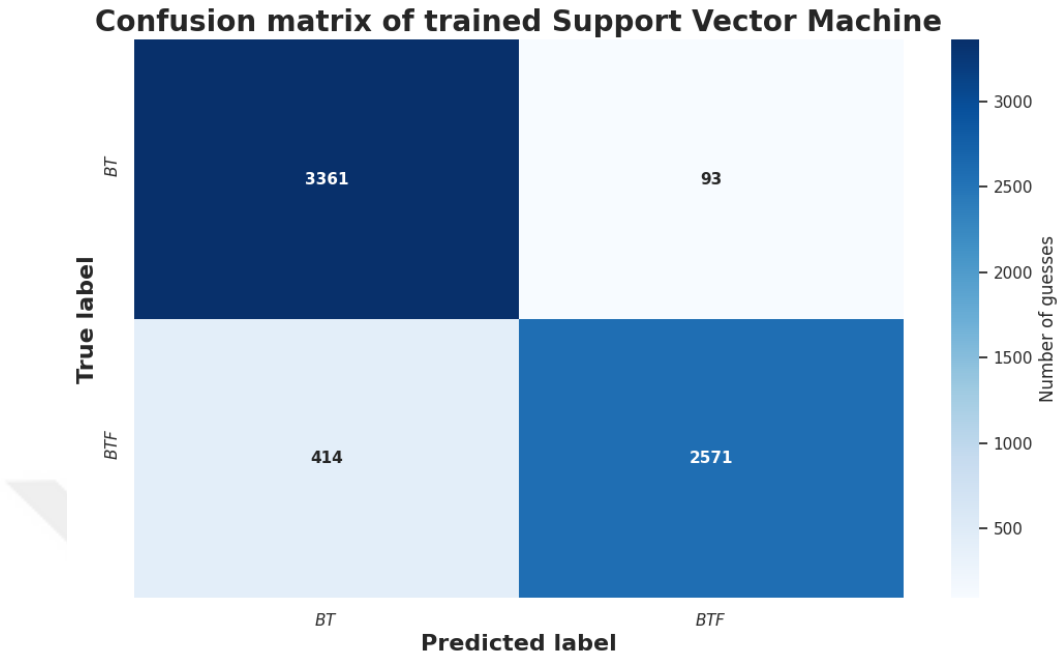


Figure 7.19: Confusion matrix for the first step classification to identify brain tumor free (BTF) individuals and patients with brain tumor (BT) using Support Vector Machine (SVM) model for spectra-wise classification

cation task.

7.1.4 Support Vector Machine (SVM)

Support Vector Machine (SVM) is a powerful classification algorithm that aims to find the optimal hyperplane that best separates different classes in a dataset. It works by identifying support vectors, which are the data points closest to the decision boundary, and maximizing the margin between classes. SVM is particularly effective for high-dimensional data and is known for its robustness and accuracy. Our study used SVM to classify the SERS spectra into the three identified states. The algorithm’s ability to handle high-dimensional feature spaces and its effectiveness in separating non-linearly separable data made it a valuable tool for our analysis.

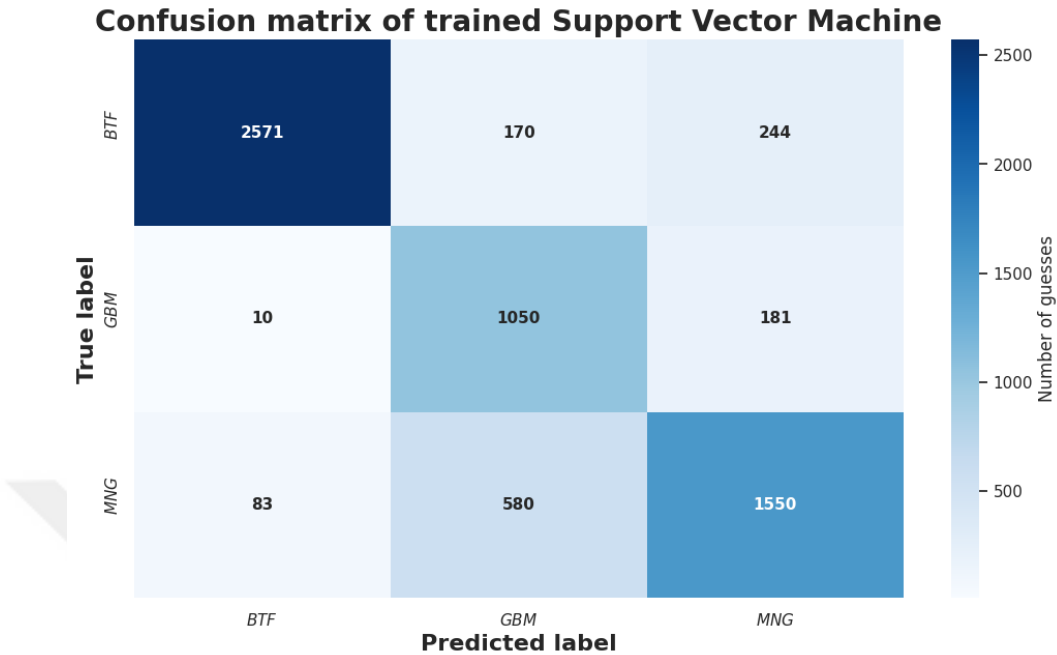


Figure 7.20: Confusion matrix for the second step classification to identify brain tumor free (BTF) individuals and patients with glioblastoma (GB), meningioma (MNG) using Support Vector Machine (SVM) model for spectra-wise classification

7.1.5 Neural Networks (NN)

Neural Networks (NN) are inspired by the structure and function of the human brain. They consist of layers of interconnected nodes (neurons) that process input data through a series of transformations to approximate complex functions. NNs are effective in static data processing tasks such as regression, classification, and pattern recognition. In this study, we used NNs to learn the complex patterns in the SERS spectra and classify them into the three identified states. The network’s multiple-layer architecture allowed it to capture and model the intricate relationships within the spectral data.

7.1.6 Convolutional Neural Networks (CNN)

Convolutional Neural Networks (CNN) are a specialized type of neural network that is particularly effective for image analysis. They use convolutional layers to detect local patterns and pooling layers to reduce the dimensionality of the data. This hier-

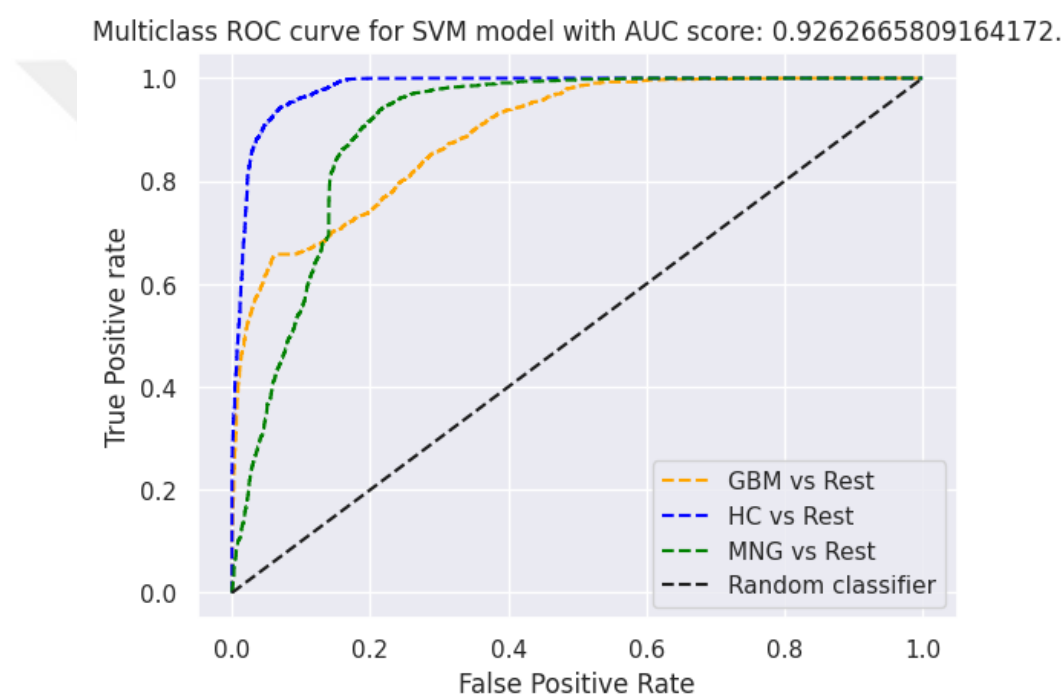


Figure 7.21: ROC Curve for the classification to identify brain tumor-free (BTF) individuals and patients with glioblastoma (GB) or meningioma (MNG) using Support Vector Machine (SVM) model for spectra-wise classification

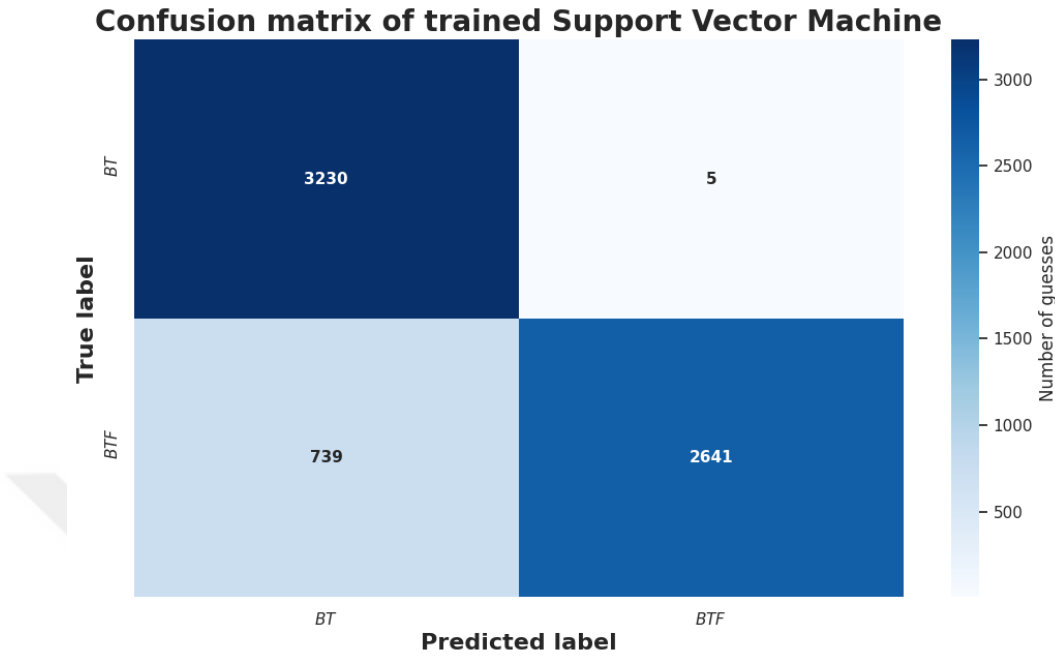


Figure 7.22: Confusion matrix for the first step classification to identify brain tumor free (BTF) individuals and patients with brain tumor (BT) using Support Vector Machine (SVM) model for patient-wise classification

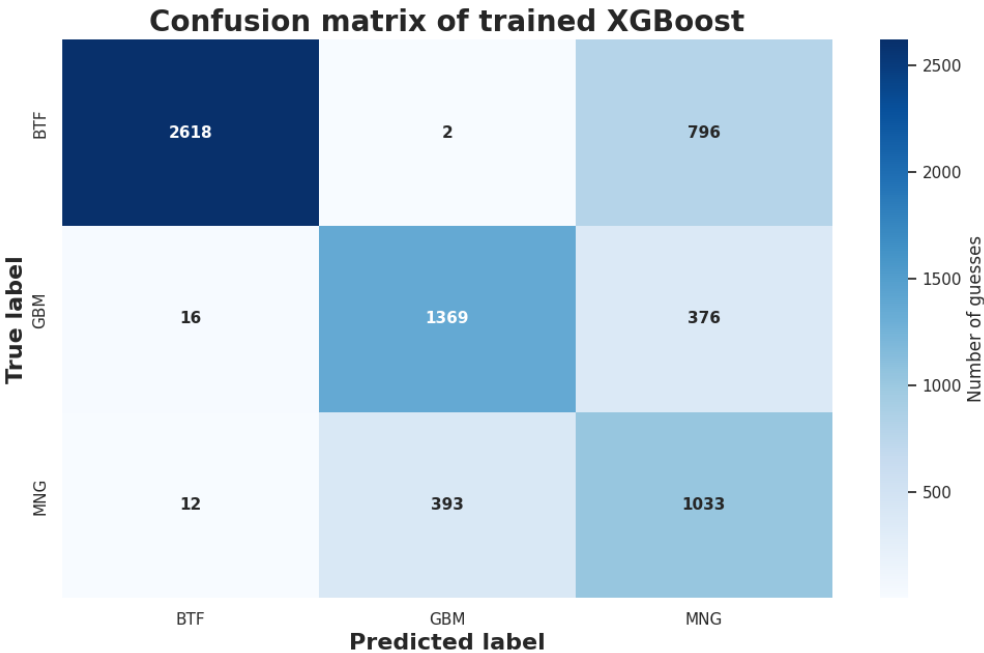


Figure 7.23: Confusion matrix for the second step classification to identify brain tumor free (BTF) individuals and patients with glioblastoma (GB), meningioma (MNG) using Support Vector Machine (SVM) model for spectra-wise classification

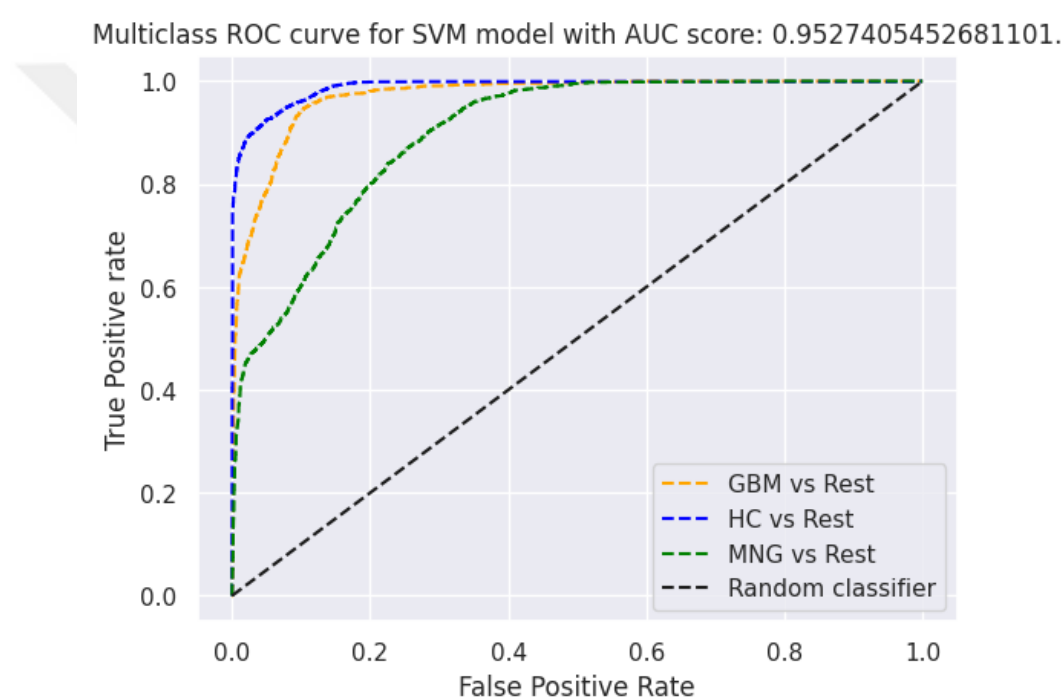


Figure 7.24: ROC Curve for the classification to identify brain tumor-free (BTF) individuals and patients with glioblastoma (GB) or meningioma (MNG) using Support Vector Machine (SVM) model for patient-wise classification

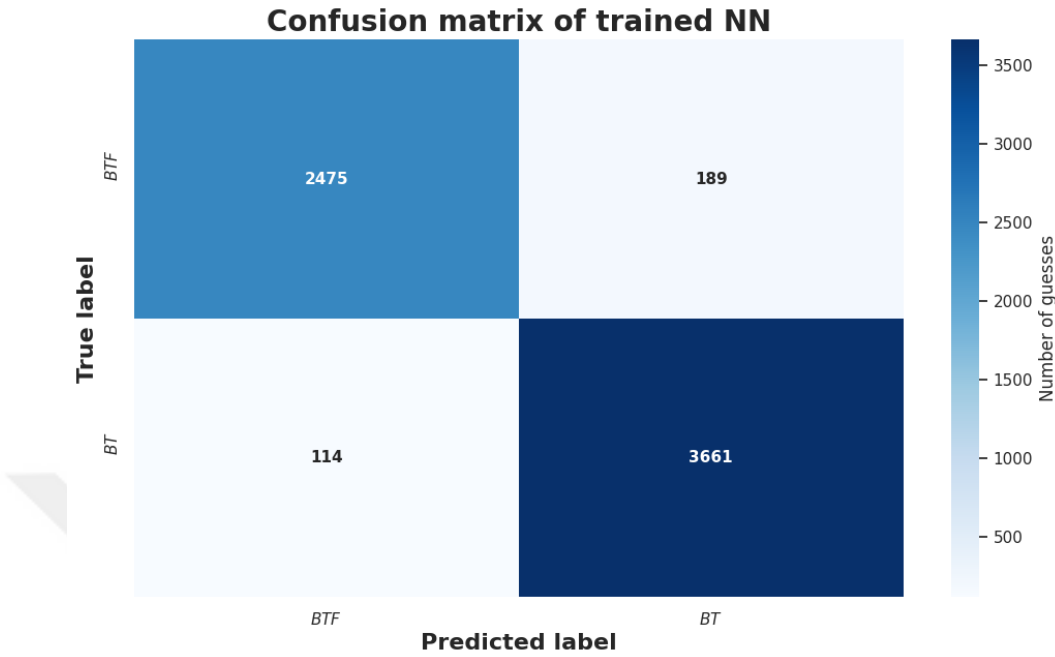


Figure 7.25: Confusion matrix for the first step classification to identify brain tumor free (BTF) individuals and patients with brain tumor (BT) using Neural Networks (NN) for spectra-wise classification

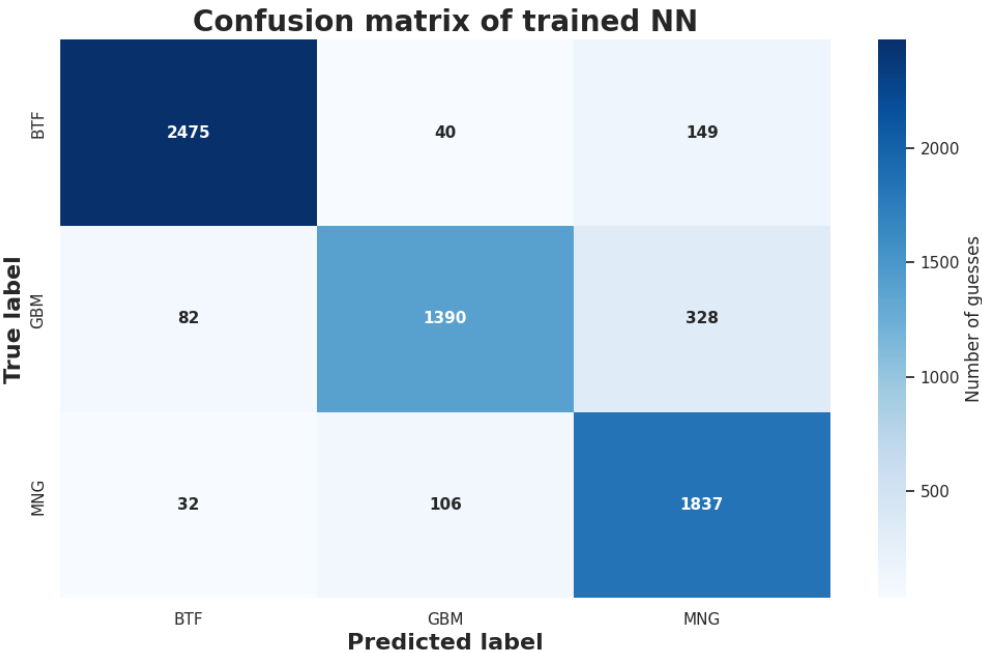


Figure 7.26: Confusion matrix for the second step classification to identify brain tumor free (BTF) individuals and patients with glioblastoma (GB), meningioma (MNG) using Neural Networks (NN) for spectra-wise classification

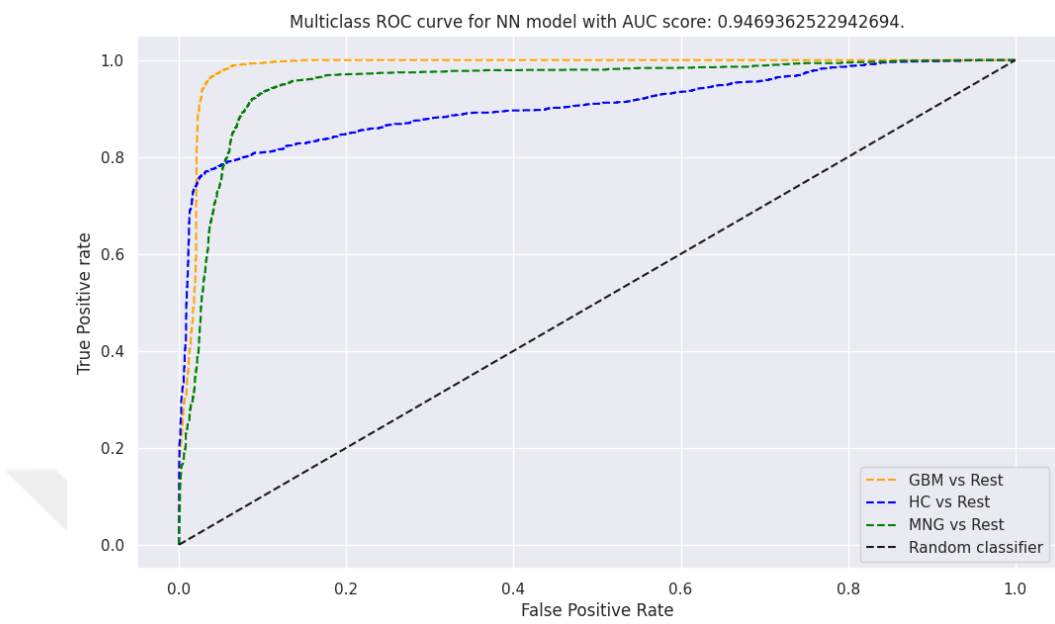


Figure 7.27: ROC Curve for the classification to identify brain tumor-free (BTF) individuals and patients with glioblastoma (GB) or meningioma (MNG) using Neural Networks (NN) for spectra-wise classification

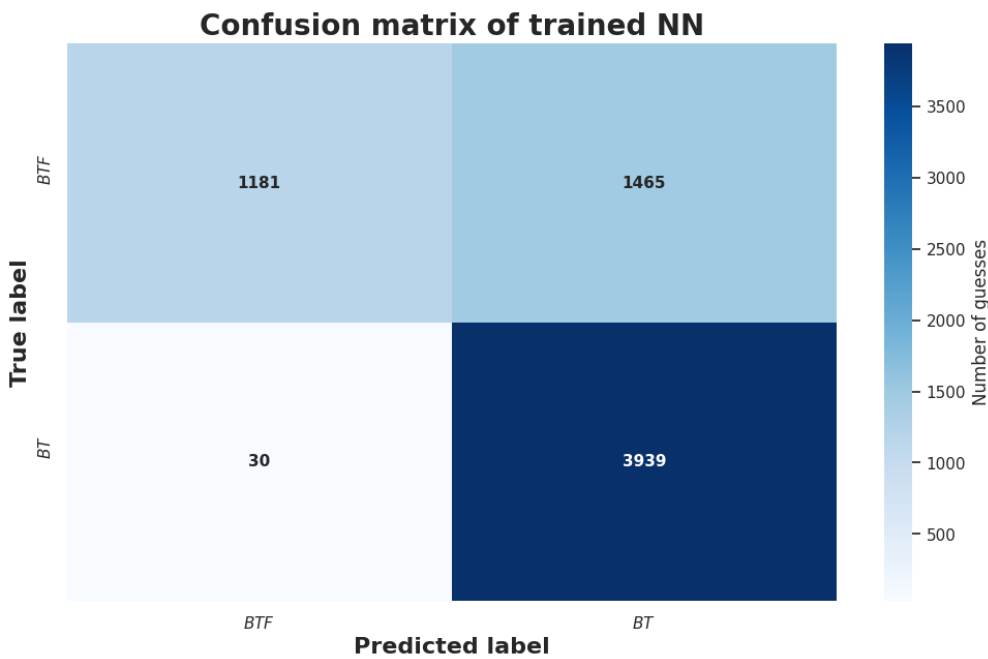


Figure 7.28: Confusion matrix for the first step classification to identify brain tumor free (BTF) individuals and patients with brain tumor (BT) using Neural Networks (NN) for patient-wise classification

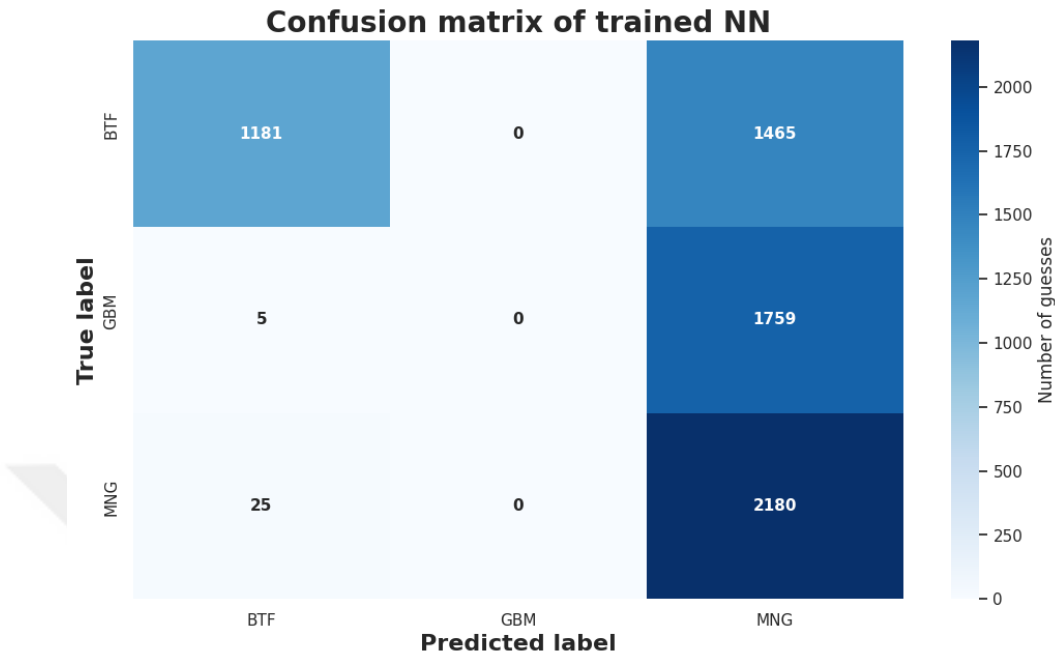


Figure 7.29: Confusion matrix for the second step classification to identify brain tumor free (BTF) individuals and patients with glioblastoma (GB), meningioma (MNG) using Neural Networks (NN) for spectra-wise classification

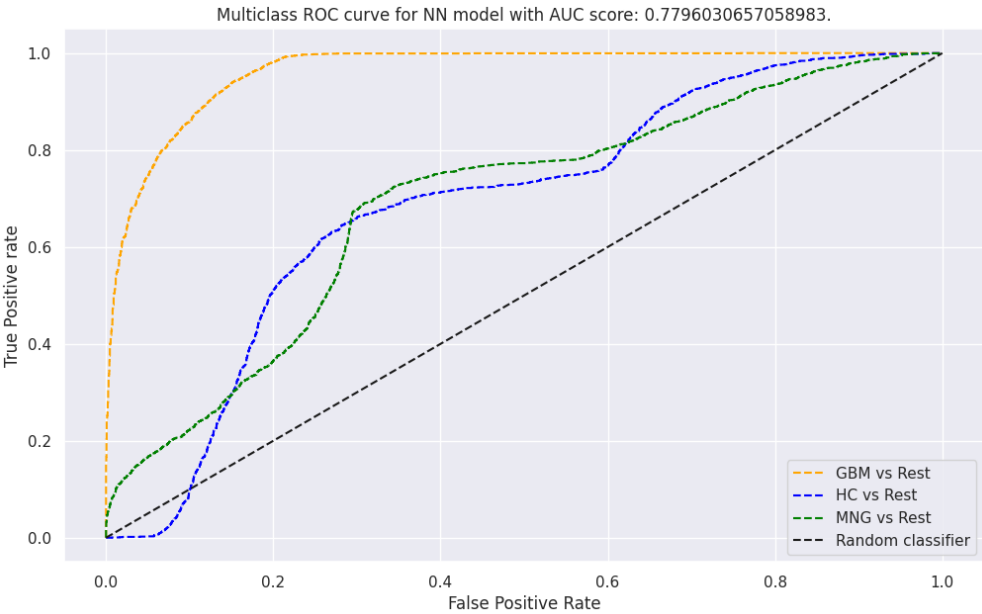


Figure 7.30: ROC Curve for the classification to identify brain tumor-free (BTF) individuals and patients with glioblastoma (GB) or meningioma (MNG) using Neural Networks (NN) for patient-wise classification

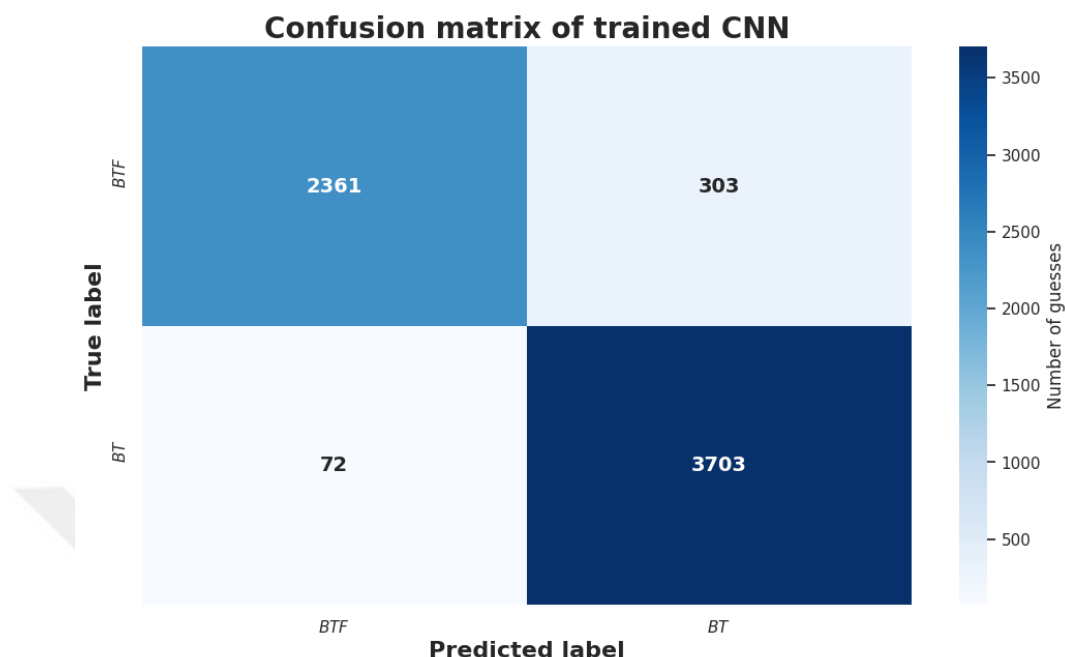


Figure 7.31: Confusion matrix for the first step classification to identify brain tumor free (BTF) individuals and patients with brain tumor (BT) using Convolutional Neural Networks (CNN) for spectra-wise classification

archical approach allows CNNs to learn increasingly abstract features from the data, making them well-suited for analyzing spectral data, which can be treated as one-dimensional image data. Given the large amount of spectral data available, CNNs were ideal for our classification task. Their ability to capture spatial hierarchies in data makes them powerful for recognizing patterns in the SERS spectra.

7.1.7 Downsides of Neural Networks (NNs) and Convolutional Neural Networks (CNNs)

High Computational Cost

- **Training and Inference:** NNs and CNNs require significant computational resources for both training and inference. Training large models, especially deep networks, involves numerous matrix multiplications and gradient calculations, which are computationally intensive and time-consuming [Strubell et al., 2019, Amodei and Hernandez, 2018].

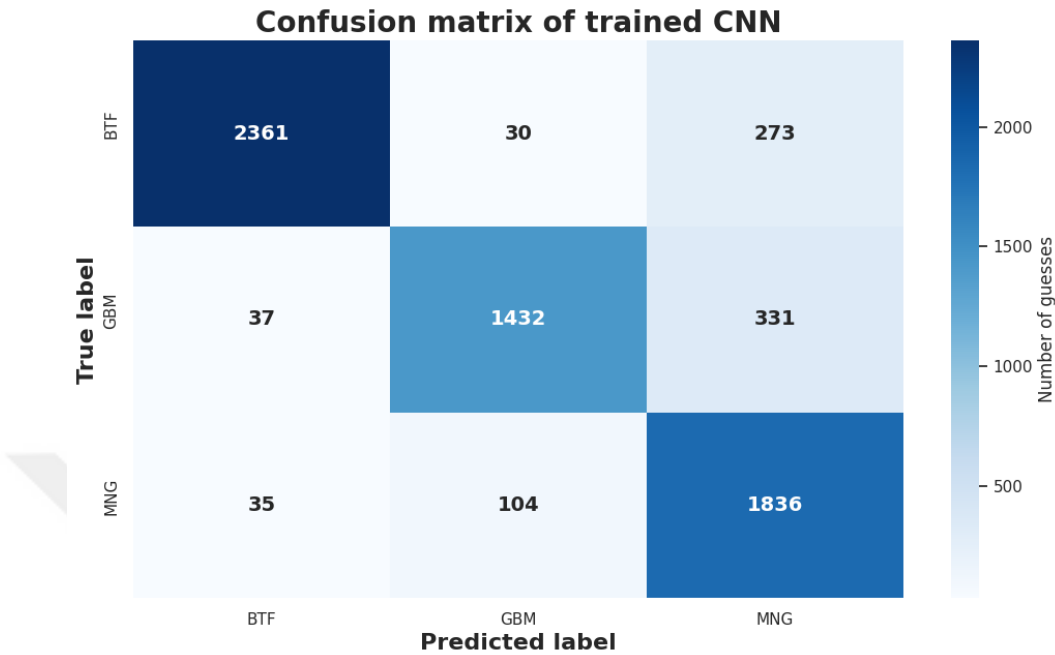


Figure 7.32: Confusion matrix for the second step classification to identify brain tumor free (BTF) individuals and patients with glioblastoma (GB), meningioma (MNG) using Convolutional Neural Networks (CNN) for spectra-wise classification

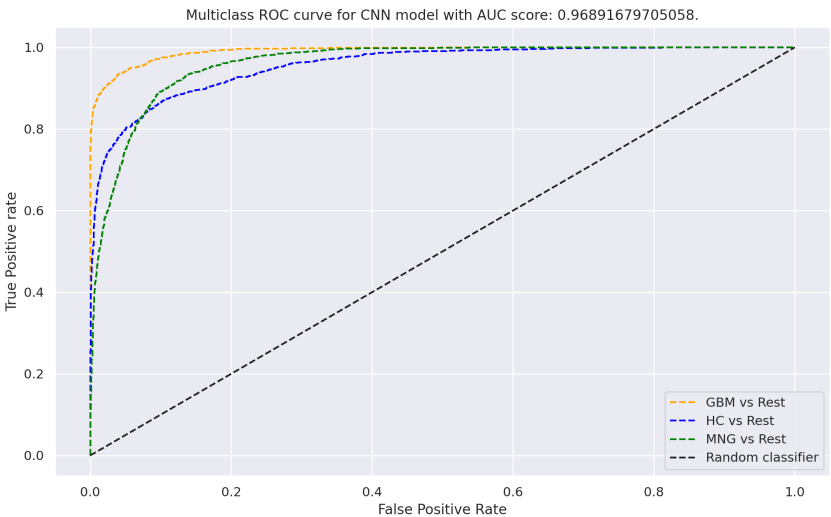


Figure 7.33: ROC Curve for the classification to identify brain tumor-free (BTF) individuals and patients with glioblastoma (GB) or meningioma (MNG) using Convolutional Neural Networks (CNN) for spectra-wise classification

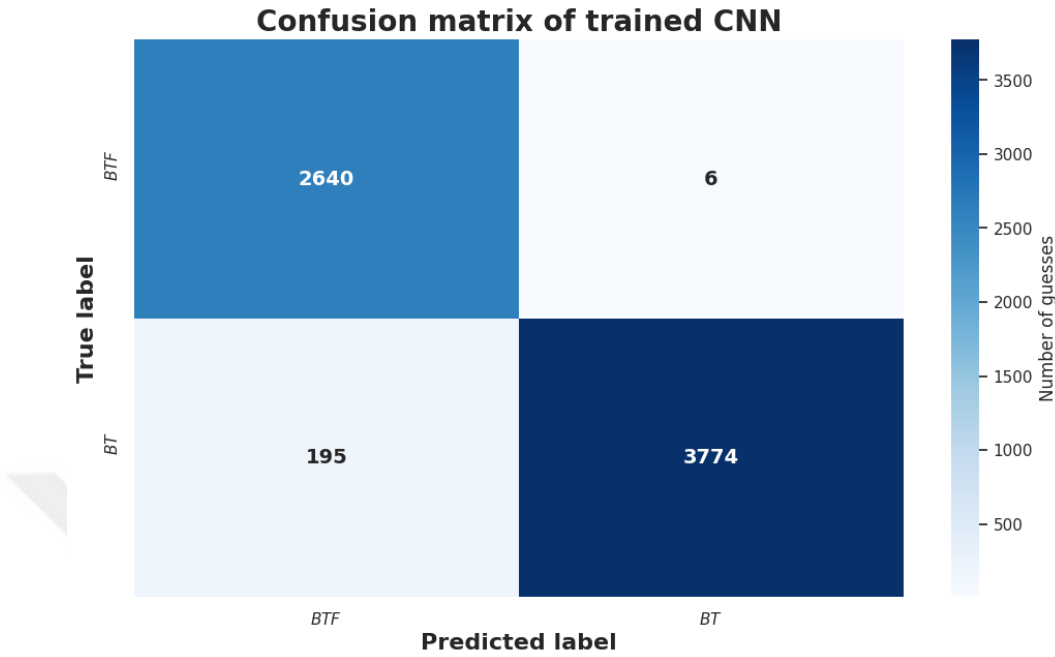


Figure 7.34: Confusion matrix for the first step classification to identify brain tumor free (BTF) individuals and patients with brain tumor (BT) using Convolutional Neural Networks (CNN) for patient-wise classification

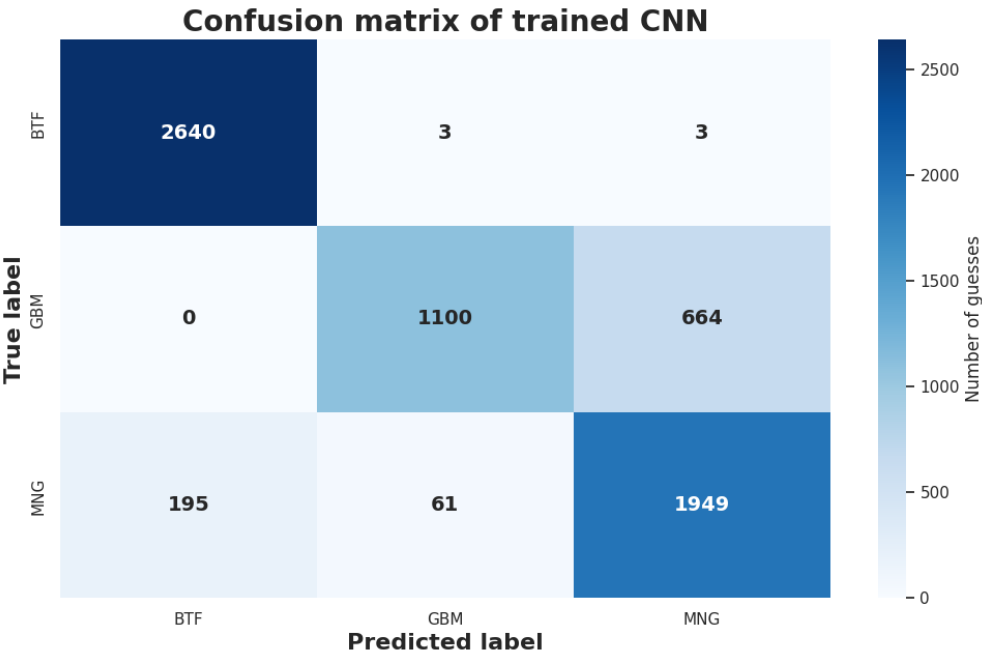


Figure 7.35: Confusion matrix for the second step classification to identify brain tumor free (BTF) individuals and patients with glioblastoma (GB), meningioma (MNG) using Convolutional Neural Networks (CNN) for patient-wise classification

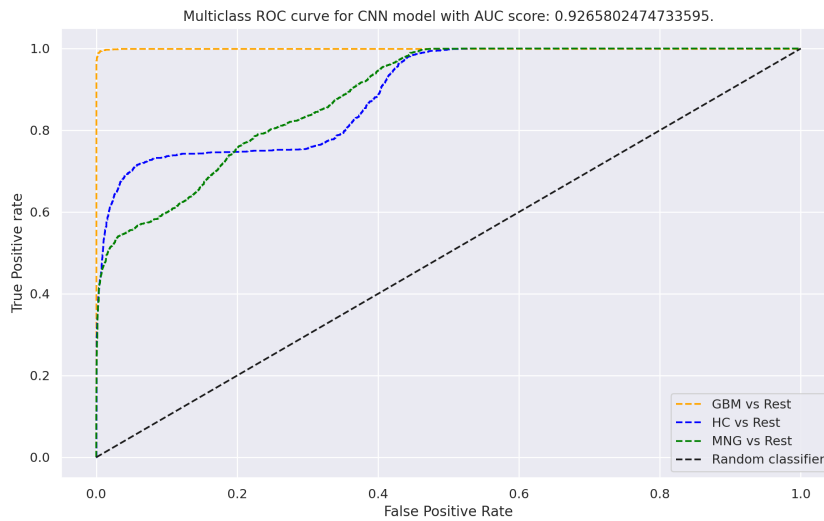


Figure 7.36: ROC Curve for the classification to identify brain tumor-free (BTF) individuals and patients with glioblastoma (GB) or meningioma (MNG) using Convolutional Neural Networks (CNN) for patient-wise classification

- **Hardware Requirements:** Efficient training of NNs and CNNs often necessitates specialized hardware such as GPUs or TPUs, which are expensive and not always accessible for all researchers or organizations [Sze et al., 2017].

Large Amount of Data Needed

- **Data Hungry:** For NNs and CNNs to perform well, they generally require large datasets. This is especially true for CNNs used in image recognition tasks. The need for substantial amounts of labeled data can be a significant barrier, as collecting and annotating data is often labor-intensive and costly [Sun et al., 2017, Hestness et al., 2017].
- **Data Quality:** The performance of these models heavily depends on the quality of the data. Noisy or unbalanced datasets can lead to poor model performance and unreliable predictions [Rolnick et al., 2017].

Extensive Hyperparameter Tuning

- **Hyperparameter Complexity:** NNs and CNNs have many hyperparameters that need to be carefully tuned, such as learning rate, batch size, number of layers, number of neurons per layer, and activation functions. This tuning is crucial for the model to converge to an optimal solution [Bergstra and Bengio, 2012, Hutter et al., 2019].
- **Time-Consuming:** Hyperparameter tuning is often an iterative and time-consuming process, typically involving trial and error. Techniques like grid search, random search, or more sophisticated methods like Bayesian optimization are used, which can still be computationally expensive [Bergstra et al., 2011].

Randomness and Variability

- **Weight Initialization:** The initial weights in NNs and CNNs are randomly initialized, which introduces variability in the training process. Different initializations can lead to different local minima, affecting the final model accuracy [Glorot and Bengio, 2010].
- **Stochastic Nature:** The training process itself is stochastic, involving random shuffling of data and mini-batch updates. As a result, the performance of the model can vary across different training runs even with the same hyperparameters and data [Goodfellow et al., 2016].

Overfitting

- **Complex Models:** Due to their high capacity and flexibility, NNs and CNNs are prone to overfitting, especially when the training dataset is small or not representative of the test data. Overfitting occurs when the model learns the training data too well, including its noise and outliers, leading to poor generalization to new data [Srivastava et al., 2014, Goodfellow et al., 2016].

- **Regularization Techniques:** To combat overfitting, techniques such as dropout, weight regularization, and early stopping are used. However, these techniques add additional complexity and require further tuning [Srivastava et al., 2014].

Interpretability

- **Black Box Nature:** NNs and CNNs are often criticized for being "black boxes" because it is difficult to interpret how they make decisions. This lack of transparency can be problematic in applications where understanding the decision-making process is crucial, such as in healthcare or finance [Lipton, 2018, Rudin, 2019].
- **Explainability Methods:** Various methods, such as SHAP values, LIME, and saliency maps, have been developed to improve interpretability, but they add another layer of complexity and are not always fully reliable [Ribeiro et al., 2016, Lundberg and Lee, 2017].

Training Time

- **Long Training Periods:** Training deep networks can take a long time, ranging from hours to weeks, depending on the size of the dataset and the complexity of the model. This long training time can be a bottleneck in iterative model development and experimentation [Dean et al., 2012].

While Neural Networks and Convolutional Neural Networks have revolutionized many fields with their powerful capabilities, they come with significant downsides that must be carefully managed. High computational costs, large data requirements, extensive hyperparameter tuning, inherent randomness, risk of overfitting, and challenges with interpretability are some of the primary concerns that practitioners must address to leverage these models effectively.

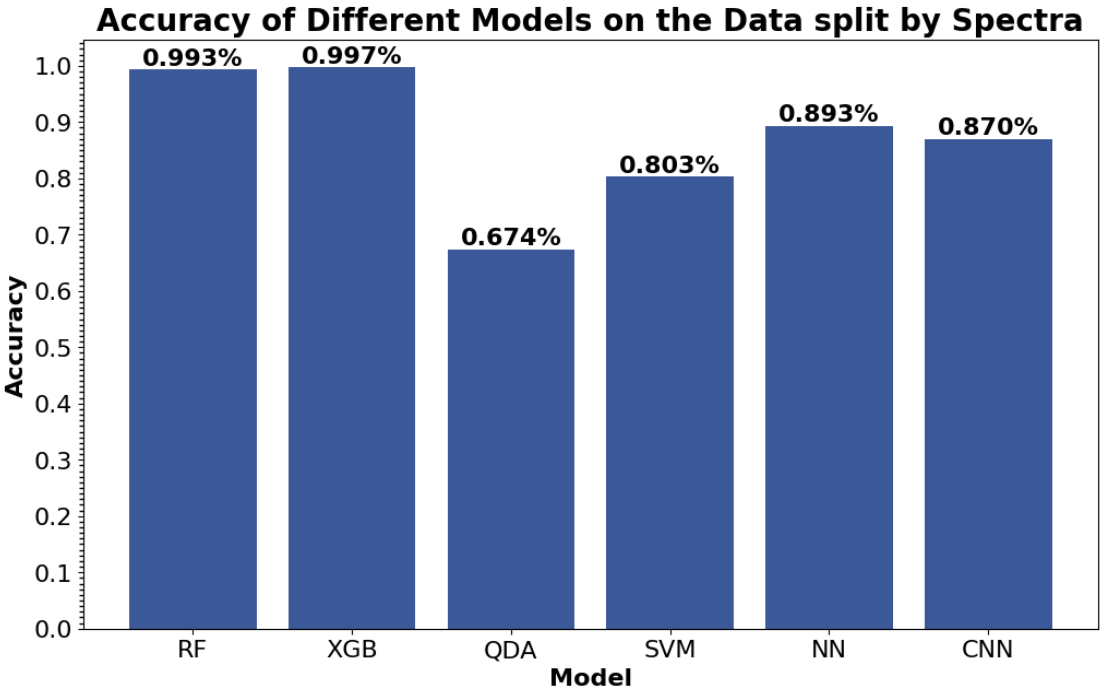


Figure 7.37: Comparison of different artificial intelligence algorithms with spectra-wise classification: We obtained 99.3% prediction accuracy with Random Forest (RF), 99.7% with XG-Boost (XGB), 67.4% with Quadratic Discriminant Analysis (QDA), 80.3% with Support Vector Machine (SVM), 89.3% with Neural Networks (NN), 87% with Convolutional Neural Networks (CNN) classifiers

7.2 Model Training and Evaluation

This section details the training and evaluation process for the AI models used in this study. It covers the composition of the training and test sets, the metrics used for performance evaluation, and the overall outcomes.

7.3 Training Set Composition

Eighty percent of the original dataset (23,638 spectra) was used as the training set for optimizing each model’s parameters. This large training set ensured that the models had sufficient data to learn the complex patterns in the spectra. The remaining twenty percent of the data served as the test set. This division ensured that the models were evaluated on unseen data, providing a realistic measure of

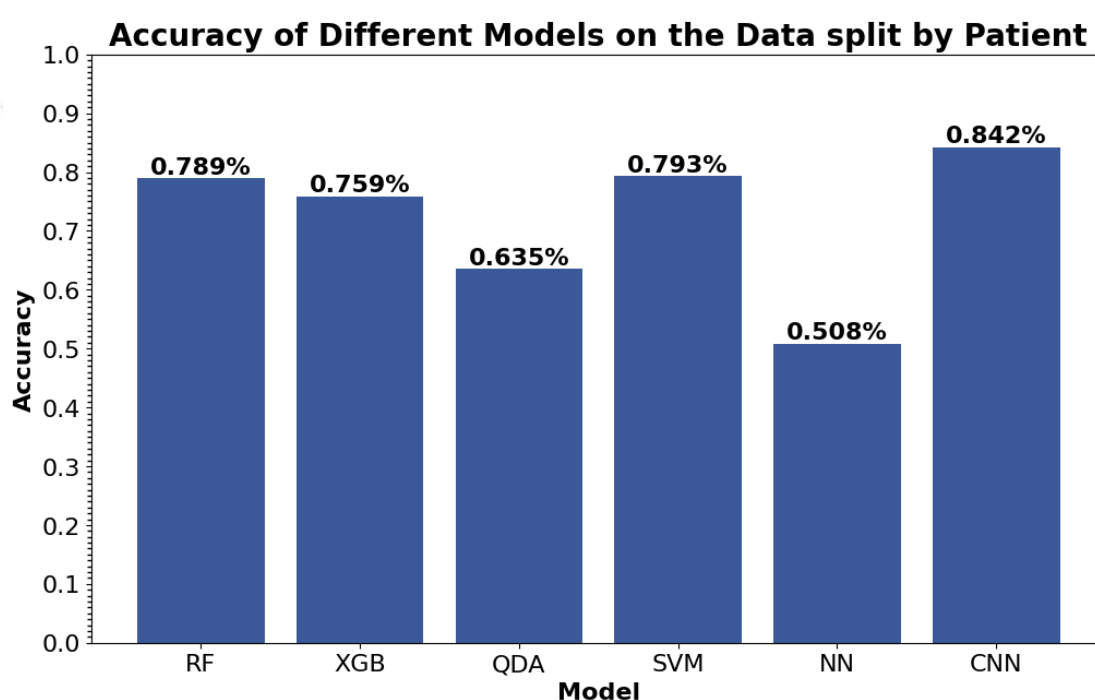


Figure 7.38: Comparison of different artificial intelligence algorithms with spectra-wise classification: We obtained 78.9% prediction accuracy with Random Forest (RF), 75.9% with XG-Boost (XGB), 63.5% with Quadratic Discriminant Analysis (QDA), 79.3% with Support Vector Machine (SVM), 50.8% with Neural Networks (NN), 84.2% with Convolutional Neural Networks (CNN) classifiers

their generalization capabilities.

7.4 Testing and Performance Evaluation

The performance of the models was evaluated based on their accuracy, defined as the percentage of correctly predicted states on the test set. Accuracy provides a straightforward metric for assessing the overall performance of the models. Other metrics, such as precision, recall, and F1 score, were considered to provide a more comprehensive evaluation of the models’ performance.

7.5 Confusion Matrix Analysis

For each model, a confusion matrix was produced to summarize the counts of true positive, true negative, false positive, and false negative predictions. The confusion matrix provides detailed insights into each model’s accuracy and error rates, highlighting areas where the model performs well and areas where it may need improvement. This analysis helps identify specific challenges in classification and guide further model refinement. For instance, a high number of false positives or false negatives in a particular class can indicate areas where the model may be overfitting or underfitting.

For CNN model, we calculated the specificity and sensitivity values below:

Actual/Predicted	Glioblastoma	Meningioma	Healthy Control
Glioblastoma	1141	623	0
Meningioma	131	2006	68
Healthy Control	3	14	2629

Table 7.1: Confusion Matrix for CNN Patient-wise Classification

Sensitivity (Recall, True Positive Rate)

Sensitivity for Glioblastoma

$$\begin{aligned} \text{Sensitivity}_{\text{Glioblastoma}} &= \frac{\text{TP}_{\text{Glioblastoma}}}{\text{TP}_{\text{Glioblastoma}} + \text{FN}_{\text{Glioblastoma}}} \\ \text{TP}_{\text{Glioblastoma}} &= 1141, \quad \text{FN}_{\text{Glioblastoma}} = 623 + 0 = 623 \\ \text{Sensitivity}_{\text{Glioblastoma}} &= \frac{1141}{1141 + 623} = \frac{1141}{1764} \approx 0.647 \end{aligned}$$

Sensitivity for Meningioma

$$\begin{aligned} \text{Sensitivity}_{\text{Meningioma}} &= \frac{\text{TP}_{\text{Meningioma}}}{\text{TP}_{\text{Meningioma}} + \text{FN}_{\text{Meningioma}}} \\ \text{TP}_{\text{Meningioma}} &= 2006, \quad \text{FN}_{\text{Meningioma}} = 131 + 68 = 199 \\ \text{Sensitivity}_{\text{Meningioma}} &= \frac{2006}{2006 + 199} = \frac{2006}{2205} \approx 0.910 \end{aligned}$$

Sensitivity for Healthy Control

$$\begin{aligned} \text{Sensitivity}_{\text{Healthy Control}} &= \frac{\text{TP}_{\text{Healthy Control}}}{\text{TP}_{\text{Healthy Control}} + \text{FN}_{\text{Healthy Control}}} \\ \text{TP}_{\text{Healthy Control}} &= 2629, \quad \text{FN}_{\text{Healthy Control}} = 3 + 14 = 17 \\ \text{Sensitivity}_{\text{Healthy Control}} &= \frac{2629}{2629 + 17} = \frac{2629}{2646} \approx 0.994 \end{aligned}$$

Specificity (True Negative Rate)

Specificity for Glioblastoma

$$\begin{aligned} \text{Specificity}_{\text{Glioblastoma}} &= \frac{\text{TN}_{\text{Glioblastoma}}}{\text{TN}_{\text{Glioblastoma}} + \text{FP}_{\text{Glioblastoma}}} \\ \text{TN}_{\text{Glioblastoma}} &= 2006 + 68 + 3 + 14 + 2629 = 4720, \quad \text{FP}_{\text{Glioblastoma}} = 131 + 0 = 131 \\ \text{Specificity}_{\text{Glioblastoma}} &= \frac{4720}{4720 + 131} = \frac{4720}{4851} \approx 0.973 \end{aligned}$$

Specificity for Meningioma

$$\begin{aligned} \text{Specificity}_{\text{Meningioma}} &= \frac{\text{TN}_{\text{Meningioma}}}{\text{TN}_{\text{Meningioma}} + \text{FP}_{\text{Meningioma}}} \\ \text{TN}_{\text{Meningioma}} &= 1141 + 0 + 3 + 2629 = 3773, \quad \text{FP}_{\text{Meningioma}} = 623 + 14 = 637 \\ \text{Specificity}_{\text{Meningioma}} &= \frac{3773}{3773 + 637} = \frac{3773}{4410} \approx 0.856 \end{aligned}$$

Specificity for Healthy Control

$$\text{Specificity}_{\text{Healthy Control}} = \frac{\text{TN}_{\text{Healthy Control}}}{\text{TN}_{\text{Healthy Control}} + \text{FP}_{\text{Healthy Control}}}$$

$$\text{TN}_{\text{Healthy Control}} = 1141 + 623 + 131 + 2006 = 3901, \quad \text{FP}_{\text{Healthy Control}} = 68$$

$$\text{Specificity}_{\text{Healthy Control}} = \frac{3901}{3901 + 68} = \frac{3901}{3969} \approx 0.983$$

Metric	Glioblastoma	Meningioma	Healthy Control
Sensitivity	0.647	0.910	0.994
Specificity	0.973	0.856	0.983

Table 7.2: Sensitivity and Specificity for Glioblastoma, Meningioma, and Healthy Control

7.6 Conclusion and Future Work

Using Surface-Enhanced Raman Spectroscopy (SERS) combined with AI models presents a promising approach for non-invasive tumor detection. Pre-processing steps and computational analysis given in the previous chapters provided a robust framework for classifying brain tumor exosome samples. Investigating the landscape of tumor-associated proteins in exosomes from various disease states and control groups provides crucial insights into disease mechanisms and potential therapeutic targets. The methods and analytical techniques used to study the proteomic profiles of exosomes derived from brain tumor patients and healthy individuals focused on tumor-related proteins using advanced mass spectrometry and bioinformatics tools.

Future work will focus on refining these models, exploring additional pre-processing techniques, and expanding the dataset to include a broader range of tumor types and conditions. Additionally, integrating other spectroscopic techniques and multi-modal data analysis could enhance this approach's diagnostic capabilities and clinical applicability. Future research will also expand the sample size, refine analytical techniques, and explore the functional roles of the identified proteins in disease progression.

Chapter 8

CONCLUSION

This chapter summarizes the results of the study, discussing the implications of the findings for glioblastoma (GB) diagnostics, the strengths and limitations of the study, and potential future directions for research. The study aimed to develop a non-invasive diagnostic method for GB using Surface-Enhanced Raman Spectroscopy (SERS) and AI. The results demonstrate the effectiveness of this approach in distinguishing between exosomes from GB, meningioma (MNG), and healthy control (HC) plasma samples.

8.1 Results and Discussion

8.1.1 Interpretation of Results

The study successfully demonstrated the potential of distinguishing between exosomes from GB patient plasma samples, MNG plasma samples, and HC plasma samples using SERS and various machine learning models. The characterization of exosomes via Transmission Electron Microscopy (TEM), Cryo-Electron Microscopy (Cryo-EM), Interferometric Scattering Microscopy (iSCAT), and Nanoparticle Tracking Analysis (NTA) provided detailed morphological and size distribution information. Biological characterization through Bead-Captured Flow Cytometry (BC-FC) and Western Blotting (WB) confirmed the presence and purity of exosomal markers, aligning with MISEV guidelines. SERS measurements revealed distinct spectral peaks for exosomes from healthy individuals and similarities between GB and MNG exosomes, indicating potential molecular overlaps.

Multiple models were employed in this study, including Random Forest (RF), Quadratic Discriminant Analysis (QDA), Support Vector Machine (SVM), Convo-

lutional Neural Network (CNN), and Extreme Gradient Boosting (XGB). Among these, the CNN model achieved the highest accuracy of 87.3%, demonstrating its superior capability in classifying patient states. As a result of patient-wise classification, the accuracy values for each model were as follows:

- Random Forest (RF): 78.9%
- Quadratic Discriminant Analysis (QDA): 63.5%
- Support Vector Machine (SVM): 79.3%
- Convolutional Neural Network (CNN): 87.3%
- Extreme Gradient Boosting (XGB): 75.9%

The sensitivity and specificity values for Brain Tumor (BT) and Brain Tumor Free (BTF) Classification are as follows:

- **Sensitivity:**
 - BTF: 99.77%
 - BT: 95.08%
- **Specificity:**
 - BTF: 95.08%
 - BT: 99.77%

The sensitivity and specificity values for GB, MNG, and HC using the CNN model are as follows:

- **Sensitivity:**
 - Glioblastoma: 62.32%
 - Meningioma: 90.85%

- Healthy Control: 99.89%

- **Specificity:**

- Glioblastoma: 93.04%
- Meningioma: 98.32%
- Healthy Control: 98.26%

The Receiver Operating Characteristic (ROC) curve analysis for patient-wise classification showed an AUC value of 0.92, highlighting the effectiveness of the proposed diagnostic approach and its potential for clinical application in neuro-oncology.

8.1.2 Implications for Glioblastoma Diagnostics

The new diagnostic method employing SERS and AI has significant implications for clinical practice. By providing a non-invasive, rapid, and highly accurate means of detecting GB through exosome analysis, this method can potentially reduce the need for invasive biopsies and unnecessary imaging procedures and allow for earlier detection, therapy monitoring, and treatment. The ability to distinguish between different brain tumor types using exosomal signatures could revolutionize neuro-oncology diagnostics, facilitating timely and personalized therapeutic interventions.

8.1.3 Strengths and Limitations

Strengths of this study include the use of multiple advanced characterization techniques to validate the presence and purity of exosomes and the integration of SERS and AI for high-accuracy diagnostic potential. However, limitations include the small sample size and potential confounding variables introduced by different conditions under which blood samples were collected from patients and healthy controls (in the operation room before surgery for patients and in a walk-in setting for healthy people). The heterogeneity within patient populations and the lack of standardized protocols for exosome isolation and SERS analysis across different laboratories may also affect reproducibility.

8.1.4 *Summary of Key Findings*

The study achieved an accuracy of 87.3% in predicting patient states using the CNN model for SERS and AI-based analysis of exosome spectra. ROC curve analysis for patient-wise classification showed an AUC value of 0.92. These findings highlight the effectiveness of the proposed diagnostic approach and its potential for clinical application in neuro-oncology.

8.1.5 *Impact on Patient Care*

Implementing this novel diagnostic method in clinical practice could significantly improve patient outcomes by providing a rapid, cost-effective, and non-invasive means of detecting and differentiating brain tumors. This approach could facilitate early intervention and more personalized treatment strategies, ultimately enhancing the quality of care for patients with brain tumors.

8.1.6 *Final Remarks*

The integration of exosomes, SERS, and AI presents a promising advancement in the field of neuro-oncology diagnostics. This research underscores the potential of these technologies to transform brain tumor detection and monitoring, offering a new era of precision medicine. Continued research and clinical validation are essential to fully realize the benefits of this innovative diagnostic tool.

8.2 *Discussion*

Liquid biopsy holds promise as a revolutionary solution in the realm of brain tumor detection. This non-invasive technique involves the analysis of small extracellular vesicles (EVs) and their components, including DNA, RNA, proteins, and cargo molecules, offering a comprehensive view of tumor heterogeneity and evolution over time. By examining molecular markers of brain tumors, including genetic mutations, epigenetic alterations, and protein expression profiles in these liquid components, clinicians can obtain valuable insights into the underlying biology of brain tumors.

In this study, using a combination of SERS and neural networks, we successfully demonstrated the potential of distinguishing between exosomes from GB patient plasma samples, MNG plasma samples, and healthy plasma samples. The rationale behind choosing MNG as the control tumor group was due to the need to establish a proof of concept. By choosing a non-glial, benign tumor such as MNG, we aimed to validate our hypothesis that distinct tumors could be differentiated based on their exosomal molecular signatures using SERS. MNG is an ideal starting point due to its molecular dissimilarity to GB, which is a malignant glial tumor. This allows for a foundational understanding before progressing to differentiations between more closely related glial tumors.

The decision to focus on MNG as a control group was also influenced by its abundance. As the most common primary benign tumor, MNG is readily available for study, simplifying the process of obtaining samples. Our successful discrimination between GB and MNG exosomes not only supports the validity of our methodology but also lays the groundwork for the next phase of research.

However, it is essential to acknowledge the limitations of this study. The collection of plasma samples from GB and MNG patients before tumor resection posed a challenge, as these patients were under anesthesia during the sampling process. Healthy control subjects did not receive anesthesia, potentially introducing confounding variables. Additionally, the small sample size may impact the generalizability of these findings.

The ROC curves provided an AUC value of 0.92 for patient-wise classification, which highlights the strength of the model when accounting for patient provenance. This more realistic accuracy value does not appear overfitted and is clinically significant.

In addition, the heterogeneity within the GB and MNG patient populations could introduce variability in SERS exosomal signatures. Factors such as age, gender, and comorbidities may influence the molecular composition of exosomes, potentially complicating the development of a universally applicable diagnostic model. The absence of standardized protocols for exosome isolation and SERS analysis across

different labs may also limit the reproducibility of findings.

Moving forward, our research can be strengthened by increasing the number of patients in our study and expanding the spectral library. This approach will enhance the reliability of our predictive model, providing more accurate results for distinguishing between different tumor types. As the next logical step, our focus should shift towards distinguishing between glial tumors with varying grades. Given the molecular similarities among these tumors, this represents a more intricate challenge that our methodology and neural network model can be developed to address. Furthermore, exploration into potential applications, such as predicting prognosis or monitoring treatment response, can broaden the utility of our tool clinically. Ultimately, the integration of exosomes, SERS, and neural networks holds promise in advancing the field of cancer early detection.

8.3 Future Directions and Perspectives

Expanding the cohort to include a larger number of patients and performing longitudinal analyses with tumor and plasma exosomes will strengthen the reliability of our findings. Incorporating targeting molecules to increase sensitivity in the sensors and conducting clinical validation with expanded patient samples are crucial next steps. These efforts will enhance the applicability and accuracy of this diagnostic method, paving the way for its integration into routine clinical practice.

BIBLIOGRAPHY

- [A. Radbruch and Wiestler, 2021] A. Radbruch, J. F. and Wiestler, M. (2021). Molecular imaging of glioblastoma: Advanced techniques and clinical applications. *Molecular Imaging and Biology*, 23(6):919–928.
- [Ajuh et al., 2000] Ajuh, P. et al. (2000). Cdc5l. *Molecular Cell Biology*, 20(11):3544–3554.
- [Akers et al., 2013] Akers, J. C., Ramakrishnan, V., Yang, I., Hua, W., Mao, Y., Carter, B. S., et al. (2013). A cerebrospinal fluid microrna signature as biomarker for glioblastoma. *Oncotarget*, 4(10):1526.
- [Ali et al., 2020] Ali, M. Y., Oliva, C. R., Noman, A. S. M., et al. (2020). Radioresistance in glioblastoma and the development of radiosensitizers. *Cancers*, 12(9):2511.
- [Alix-Panabières and Pantel, 2020] Alix-Panabières, C. and Pantel, K. (2020). Liquid biopsy: current status and future perspectives. *Molecular oncology*, 14(5):1184–1218.
- [Alizadeh et al., 2000] Alizadeh, A., Eisen, M., Davis, R., Ma, C., Lossos, I., Rosenwald, A., Boldrick, J., Sabet, H., Tran, T., Yu, X., and Powell, J. (2000). Distinct types of diffuse large b-cell lymphoma identified by gene expression profiling. *Nature*, 403:503–511.
- [Alvarez-Erviti et al., 2011] Alvarez-Erviti, L. et al. (2011). Delivery of sirna to the mouse brain by systemic injection of targeted exosomes. *Nature Biotechnology*, 29(4):341–345.

- [Alzubaidi et al., 2021] Alzubaidi, L., Zhang, J., Humaidi, A. J., et al. (2021). Review of deep learning: Concepts, cnn architectures, challenges, applications, future directions. *Journal of Big Data*, 8:1–74.
- [Amodei and Hernandez, 2018] Amodei, D. and Hernandez, D. (2018). Ai and compute. *OpenAI Blog*, 1:1–5.
- [Argraves et al., 1990] Argraves, W. S. et al. (1990). Fibulin is an extracellular matrix and plasma glycoprotein with repeated domain structure. *Journal of Cell Biology*.
- [Argraves et al., 2003] Argraves, W. S. et al. (2003). Fibulin-1. *Matrix Biology*, 22(4):305–317.
- [Armstrong and Wildman, 2018] Armstrong, D. and Wildman, H. (2018). Extracellular vesicles as a next-generation diagnostic tool. *Nature Reviews Clinical Oncology*, 15(5):276–284.
- [Basso et al., 2013] Basso, M. et al. (2013). Mutant sod1-containing exosomes secreted from astrocytes trigger neuronal death. *Cell Death Differentiation*, 20(4):398–407.
- [Belderbos et al., 2014] Belderbos, S., Pouwels, P. J. W., van den Bergh, W. M., Vogt, A., Dankbaar, J. W., van Veen, E., Dijkstra, M. J., van der Kolk, A. G., van Putten, M. J. A. M., Dippel, D. W. J., and Smits, M. (2014). Magnetic resonance spectroscopy of cerebrospinal fluid and serum in patients with traumatic brain injury: a comparison with proton mrs of brain tissue. *Journal of Cerebral Blood Flow & Metabolism*, 34(5):838–845.
- [Bergstra et al., 2011] Bergstra, J., Bardenet, R., Bengio, Y., and Kégl, B. (2011). Algorithms for hyper-parameter optimization. In *Proceedings of the 24th International Conference on Neural Information Processing Systems*, pages 2546–2554.

- [Bergstra and Bengio, 2012] Bergstra, J. and Bengio, Y. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, 13(Feb):281–305.
- [Berinstein et al., 2012] Berinstein, D. M. et al. (2012). The role of *serpina3* in cancer progression. *American Journal of Cancer Research*.
- [Best et al., 2015] Best, M. G. et al. (2015). Liquid biopsies to monitor cancer progression and treatment. *Nature Reviews Clinical Oncology*, 12(10):532–542.
- [Bettegowda et al., 2014] Bettegowda, C. et al. (2014). Detection of circulating tumor dna in early- and late-stage human malignancies. *Science translational medicine*, 6(224):224ra24–224ra24.
- [Bhakdi et al., 1989] Bhakdi, S. et al. (1989). The complement system in tumor immunology. *Biochimica et Biophysica Acta*.
- [Bhoola et al., 1992] Bhoola, K. D. et al. (1992). Kallikrein and kinin receptors as targets for new drug development. *Pharmacological Reviews*.
- [Bilgin et al., 2020] Bilgin, B., Türkmen, B., Baysal, K., Solaroğlu, , and Onbaşlı, M. C. (2020). Genetic algorithm-driven design of sers-active surfaces for early detection of diseases. *Proceedings of SPIE*, 11236:PWB.
- [Blainey, 2013] Blainey, P. (2013). The future is now: single-cell genomics of bacteria and archaea. *FEMS Microbiology Reviews*, 37(3):407–427.
- [Blom, 2002] Blom, A. M. (2002). C4b-binding protein: a crucial regulator of the innate immunity system. *Scandinavian Journal of Immunology*.
- [Botto, 2001] Botto, M. (2001). Links between complement deficiency and apoptosis. *Arthritis Research*.

- [Brandsma and van den Bent, 2009] Brandsma, D. and van den Bent, M. J. (2009). Pseudoprogression and pseudoresponse in the treatment of gliomas. *Current opinion in neurology*, 22(6):633–638.
- [Brinkerhoff et al., 2019] Brinkerhoff, H., Kang, H., Liu, Z., Aksimentiev, A., and Dekker, C. (2019). Real-time nanopore sequencing for precision oncology: a case study in glioblastoma multiforme. *Biosensors and Bioelectronics*, 142:111502.
- [Butler et al., 2016] Butler, H. J. et al. (2016). Using raman spectroscopy to characterize biological materials. *Nat Protoc*, 11(4):664–87.
- [Calderwood et al., 1999] Calderwood, D. A. et al. (1999). The talin head domain binds to integrin beta subunit cytoplasmic tails and regulates integrin activation. *Journal of Biological Chemistry*.
- [Carmicheal et al., 2019] Carmicheal, J., Hayashi, C., Shin, H., Luchini, A., Fabris, L., and Nath, N. (2019). Label-free characterization of exosome via surface enhanced raman spectroscopy for the early detection of pancreatic cancer. *Nanomedicine: Nanotechnology, Biology and Medicine*, 16:88–96.
- [Cervellin and Lippi, 2012] Cervellin, G. and Lippi, G. (2012). Coagulation factor xii. *Thrombosis Research*, 130(1):e6–e7.
- [Cesarman et al., 1994] Cesarman, G. M. et al. (1994). Plasminogen in cancer progression. *Journal of Clinical Investigation*, 93(4):1593–1601.
- [Cheung et al., 2021] Cheung, T. H. et al. (2021). Validation of liquid biopsy assays: an essential step towards implementation in clinical practice. *Clinical chemistry*, 67(3):454–466.
- [Cocucci et al., 2009] Cocucci, E., Racchetti, G., and Meldolesi, J. (2009). Exosomes and other extracellular vesicles: the emerging role of exosomes in cell-cell communication. *Traffic*, 10(3):268–277.

- [Collins and Jackson, 2021] Collins, A. M. and Jackson, K. J. L. (2021). Immunoglobulin genes and cancer. *The Journal of Immunology*, 207(1):1–10.
- [Colman, 2006] Colman, R. W. (2006). Kininogen and kallikrein in blood coagulation and inflammation. *Thrombosis Research*, 122(4):S9–S14.
- [Colombo et al., 2014] Colombo, M. et al. (2014). Analysis of exosome functions: a new frontier in research. *Nature Reviews Drug Discovery*, 13(4):258–270.
- [Critchley, 2008] Critchley, D. R. (2008). Talin-1 and talin-2: proteins that connect integrins to the cytoskeleton. *Biochimica et Biophysica Acta*.
- [Crowley et al., 2020] Crowley, E. et al. (2020). Liquid biopsy: monitoring cancer-genetics in the blood. *Nature Reviews Clinical Oncology*, 10(8):472–484.
- [Cuevas et al., 2007] Cuevas, B. D. et al. (2007). Roles of mitogen-activated protein kinase kinase 4 in cancer. *Clinical Cancer Research*.
- [Dahlbäck, 1995] Dahlbäck, B. (1995). The protein c anticoagulant system: inherited defects as basis for venous thrombosis. *Thrombosis Research*.
- [Dassati et al., 2014] Dassati, S., Waldner, A., and Schweigreiter, R. (2014). Apolipoprotein d and its role in neurodegenerative disorders. *Neuroscience Letters*, 575:1–5.
- [De Mattos-Arruda et al., 2015] De Mattos-Arruda, L. et al. (2015). Brain metastasis in breast cancer: clinical and molecular characteristics and direction for future research. *Nature Reviews Clinical Oncology*, 12(6):337–352.
- [Dean et al., 2012] Dean, J., Corrado, G. S., Monga, R., et al. (2012). Large scale distributed deep networks. *Advances in Neural Information Processing Systems*, 25.

- [Degen et al., 2003] Degen, J. L. et al. (2003). Prothrombin and fibrinogen in tissue repair and tumor progression. *Seminars in Thrombosis and Hemostasis*.
- [Delgado-López and Corrales-García, 2016] Delgado-López, P. D. and Corrales-García, E. M. (2016). Survival in glioblastoma: a review on the impact of treatment modalities. *Clinical and Translational Oncology*, 18(11):1062–1071.
- [DeMattos-Arruda et al., 2011] DeMattos-Arruda, L., Cortes, J., Diaz-Rubio, E., Pinedo, H. M., and Tabernero, J. (2011). Cancer cell lines as tumor models in the preclinical evaluation of anticancer drugs. *Clinical and Translational Oncology*, 13(4):253–259.
- [Dhermain and Marazano, 2022] Dhermain, F. G. and Marazano, L. (2022). Current advances in radiotherapy for glioblastoma. *Frontiers in Oncology*, 12:865234.
- [Diamandis, 2015] Diamandis, E. P. (2015). Next-generation diagnostics: moving from laboratory research to clinical practice. *Clinical chemistry*, 61(1):139–141.
- [Dingari et al., 2013] Dingari, N. C., Barman, I., Kang, J. U., Galindo, L. H., Dasari, R. R., and Feld, M. S. (2013). Development and comparative assessment of raman spectroscopic classification algorithms for lesion discrimination in stereotactic breast biopsies with microcalcifications. *Journal of Biophotonics*, 6(5):371–381.
- [Dominguez and Holmes, 2011] Dominguez, R. and Holmes, K. C. (2011). Actin structure and function. *Annual Review of Biophysics*.
- [Doyle and Wang, 2019] Doyle, L. M. and Wang, M. Z. (2019). Overview of extracellular vesicles, their origin, composition, purpose, and methods for exosome isolation and analysis. *Cells*, 8(7):727.
- [Dzau et al., 2001] Dzau, V. J. et al. (2001). The relevance of tissue angiotensin-converting enzyme: manifestations in mechanistic and endpoint data. *American Journal of Cardiology*.

- [Ellis and Goodacre, 2006] Ellis, D. I. and Goodacre, R. (2006). Metabolic fingerprinting in disease diagnosis: biomedical applications of infrared and raman spectroscopy. *Analyst*, 131(8):875–85.
- [Endo et al., 2007] Endo, Y. et al. (2007). Role of ficolin in innate immunity and its molecular basis. *Immunobiology*.
- [Esteva et al., 2017] Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., and Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639):115–118.
- [et al., 2021a] et al., D. N. L. (2021a). Advanced diagnostic techniques for glioblastoma. *Neuro-Oncology*, 23(8):1231–1251.
- [et al., 2021b] et al., D. N. L. (2021b). Challenges in glioblastoma diagnosis: molecular testing and histopathological assessment. *Neuro-Oncology*, 23(8):1231–1251.
- [et al., 2021c] et al., D. N. L. (2021c). Histopathological evaluation in glioblastoma diagnosis. *Acta Neuropathologica*, 141(4):803–820.
- [et al., 2021d] et al., D. N. L. (2021d). Improvements in glioblastoma diagnostic imaging and surgical techniques. *Neuro-Oncology*, 23(8):1231–1251.
- [et al., 2021e] et al., D. N. L. (2021e). Molecular characteristics and aggressiveness of glioblastoma. *Acta Neuropathologica*, 141(4):803–820.
- [et al., 2021f] et al., D. N. L. (2021f). Prognosis of idh-wildtype glioblastoma. *Neuro-Oncology*, 23(8):1231–1251.
- [Everitt et al., 2011] Everitt, B., Landau, S., Leese, M., and Stahl, D. (2011). *Cluster Analysis*. Wiley.

- [Ferraro et al., 2003] Ferraro, G. et al. (2003). Detection of tumor markers using exosomes: implications for early cancer diagnosis. *Nature Reviews Cancer*, 3(7):545–555.
- [Fiandaca et al., 2015] Fiandaca, M. S. et al. (2015). Identification of preclinical alzheimer’s disease by a profile of pathogenic proteins in neurally derived blood exosomes: a case-control study. *Alzheimer’s Dementia*, 11(6):600–607.
- [Frühbeis et al., 2013] Frühbeis, C. et al. (2013). Neurotransmitter-triggered transfer of exosomes mediates oligodendrocyte-neuron communication. *PLoS Biology*, 11(7):e1001604.
- [Gaboriaud et al., 2002] Gaboriaud, C. et al. (2002). C1q and mannose-binding lectin interaction with viruses. *Immunobiology*.
- [Galliano et al., 2006] Galliano, M. F. et al. (2006). Itih family. *Biochimie*, 88(1):1209–1221.
- [Ganforina et al., 2008] Ganforina, M. D. et al. (2008). Apolipoprotein d is involved in the mechanisms regulating protection from oxidative stress. *Journal of Cell Science*.
- [Genovese et al., 2010] Genovese, G. et al. (2010). Apolipoprotein l1 variants and kidney disease. *New England Journal of Medicine*, 363(1):388–398.
- [Georgiev et al., 2024] Georgiev, D., Pedersen, S. V., Xie, R., Fernández-Galiana, A., Stevens, M. M., and Barahona, M. (2024). Ramanspy: An open-source python package for integrative raman spectroscopy data analysis. *Analytical Chemistry*, 96(21):8492–8500.
- [Giannakopoulos et al., 2014] Giannakopoulos, B. et al. (2014). The role of the complement system in antiphospholipid syndrome. *Journal of Autoimmunity*.

- [Gigli et al., 1991] Gigli, I. et al. (1991). Interactions of c4b-binding protein with cells and proteins of the complement system. *Immunobiology*.
- [Ginsberg, 2006] Ginsberg, H. N. (2006). Apolipoprotein c-iii and its regulation. *Journal of Lipid Research*, 47(1):152–160.
- [Glorot and Bengio, 2010] Glorot, X. and Bengio, Y. (2010). Understanding the difficulty of training deep feedforward neural networks. *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, pages 249–256.
- [Goetzl et al., 2016] Goetzl, E. J. et al. (2016). Altered lysosomal proteins in neural-derived plasma exosomes in preclinical alzheimer disease. *Neurology*, 85(1):40–47.
- [Gonzalez Castro and Wesseling, 2021] Gonzalez Castro, L. N. and Wesseling, P. (2021). The cimpact-now updates and their significance to current neuro-oncology practice. *Neuro-Oncology Practice*, 8(1):4–10.
- [Goodfellow et al., 2016] Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep Learning*. MIT Press.
- [Gordon et al., 1977] Gordon, T. et al. (1977). High density lipoprotein as a protective factor against coronary heart disease. *American Journal of Medicine*.
- [Greenberg and Shuman, 1982] Greenberg, C. S. and Shuman, M. A. (1982). The zymogen forms of coagulation factor xiii bind specifically to fibrinogen. *Journal of Biological Chemistry*.
- [Greene and McElvaney, 2009] Greene, C. M. and McElvaney, N. G. (2009). Proteases and antiproteases in chronic neutrophilic lung disease—relevance to drug discovery. *British Journal of Pharmacology*.
- [Griffin et al., 2001] Griffin, J. H. et al. (2001). Thrombin receptors and their role in signal transduction. *Thrombosis and Haemostasis*.

- [Grochans et al., 2022] Grochans, S., Cybulska, A. M., Simińska, D., et al. (2022). Epidemiology of glioblastoma multiforme—literature review. *Cancers*, 14(10):2412.
- [Grover et al., 2000] Grover, J. et al. (2000). Lumican-null mice exhibit abnormal collagen fibril morphology and altered skin ultrastructure. *Journal of Biological Chemistry*.
- [Guo et al., 2019] Guo, Y., Wang, Y., Zheng, L., Yang, C., and Chen, R. (2019). Electrochemical impedance spectroscopy biosensor for detection of exosomes. *ACS Sensors*, 4(11):3005–3012.
- [Gyorgy et al., 2011] Gyorgy, B., Szabo, T. G., Turiak, L., Wright, M., Herczeg, P., Ledeczi, Z., Kittel, A., Polgar, A., Toth, K., Dyo, L., et al. (2011). Detection and isolation of cell-derived microparticles and exosomes by high-throughput flow cytometry. *Journal of Thrombosis and Haemostasis*, 9(10):1969–1979.
- [Haber and Velculescu, 2017] Haber, D. A. and Velculescu, V. E. (2017). Liquid biopsy: graphical abstract. *Nature*, 579(7802):35–45.
- [Hackeng et al., 2006] Hackeng, T. M. et al. (2006). Protein s stimulates inhibition of the tissue factor pathway by tissue factor pathway inhibitor. *Proceedings of the National Academy of Sciences USA*.
- [Hale et al., 2009] Hale, L. P. et al. (2009). The role of zinc-alpha2-glycoprotein in the pathogenesis of cachexia. *Current Opinion in Supportive and Palliative Care*.
- [Hanif et al., 2017] Hanif, F., Muzaffar, K., Perveen, K., Malhi, S. M., and Simjee, S. U. (2017). Glioblastoma multiforme: a review of its epidemiology and pathogenesis through clinical presentation and treatment. *Asian Pacific Journal of Cancer Prevention*, 18(1):3–9.
- [Hanna et al., 2022] Hanna, K., Krzoska, E., Shaaban, A. M., Muirhead, D., Abu-Eid, R., and Speirs, V. (2022). Raman spectroscopy: Current applications in

- breast cancer diagnosis, challenges and future prospects. *British Journal of Cancer*, 126(8):1125–1139.
- [Hanson et al., 2016] Hanson, B. W. et al. (2016). Exosomes: emerging as a key player in cancer. *Journal of clinical investigation*, 126(5):1605–1612.
- [Hassan et al., 2008] Hassan, M. I. et al. (2008). Zinc alpha 2-glycoprotein: a multidisciplinary protein. *Molecular Cancer Research*.
- [He et al., 2018] He, M., Crow, J., Roth, M., and Zeng, Y. (2018). High-throughput isolation of malignant pleural effusion-derived extracellular vesicles using a microfluidic chip. *Lab on a Chip*, 18(22):3360–3370.
- [Heitzer et al., 2015] Heitzer, E., Ulz, P., and Geigl, J. B. (2015). Circulating tumor dna as a liquid biopsy for cancer. *Clinical Chemistry*, 61(1):112–123.
- [Hestness et al., 2017] Hestness, J., Narang, S., Ardalani, N. P., et al. (2017). Deep learning scaling is predictable, empirically. *arXiv preprint arXiv:1712.00409*.
- [Hetz et al., 2011] Hetz, C. et al. (2011). Ern1 in cancer. *Nature Reviews Cancer*, 11(1):749–760.
- [Holers, 2014] Holers, V. M. (2014). Complement and cancer. *Nature Reviews Cancer*, 14(1):25–37.
- [Homola, 2008] Homola, J. (2008). Surface plasmon resonance sensors for detection of chemical and biological species. *Chemical Reviews*, 108(2):462–493.
- [Horbinski et al., 2019] Horbinski, C., Ligon, K. L., Brastianos, P., et al. (2019). The medical necessity of advanced molecular testing in the diagnosis and treatment of brain tumor patients. *Neuro-oncology*, 21(12):1498–1508.
- [Hu et al., 2012] Hu, X. et al. (2012). Apolipoprotein f. *Lipids in Health and Disease*, 11(1):1–9.

- [Hutter et al., 2019] Hutter, F., Kotthoff, L., and Vanschoren, J. (2019). Automated machine learning: Methods, systems, challenges. *Springer Nature*.
- [Hvidberg et al., 2005] Hvidberg, V. et al. (2005). Identification of the receptor scavenging hemopexin-heme complexes. *Blood*.
- [Inohara et al., 1999] Inohara, H. et al. (1999). Galectin-3 in human oral cancer cell growth and metastasis. *International Journal of Oncology*.
- [Jackson et al., 2019] Jackson, C. M., Choi, J., and Lim, M. (2019). Mechanisms of immunotherapy resistance: lessons from glioblastoma. *Nature Immunology*, 20(9):1100–1109.
- [Jain, 2018] Jain, K. (2018). A critical overview of targeted therapies for glioblastoma. *Frontiers in Oncology*, 8:419.
- [Jang et al., 2004] Jang, Y. et al. (2004). Apolipoprotein c-iii expression is upregulated by inflammatory cytokines in human hepatocytes. *Metabolism*.
- [Jiang et al., 2019] Jiang, L., Li, X., Zhang, Y., Liu, S., Tang, Y., and Liu, J. (2019). Automatic classification of raman spectra of atherosclerotic plaque with different stability using machine learning. In *Proceedings of the 2019 International Conference on Artificial Intelligence and Advanced Manufacturing*, pages 116–119. Association for Computing Machinery.
- [Jiang et al., 2020] Jiang, N., Zhuo, X., and Wang, J. (2020). Localized surface plasmon resonance biosensors: An introduction. *Sensors and Actuators B: Chemical*, 304:127279.
- [Jin et al., 2022] Jin, Y., Xing, J., Xu, K., Liu, D., and Zhuo, Y. (2022). Exosomes in the tumor microenvironment: Promoting cancer progression. *Frontiers in Immunology*, 13:1025218.

- [Jones et al., 2002] Jones, S. E. et al. (2002). Clusterin (clu): a diverse protein implicated in many biological pathways. *International Journal of Biochemistry and Cell Biology*.
- [Jones and Jomary, 2002] Jones, S. E. and Jomary, C. (2002). Clusterin. *The International Journal of Biochemistry & Cell Biology*, 34(5):427–431.
- [Jordan and Mitchell, 2015] Jordan, M. I. and Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245):255–260.
- [Kalluri and LeBleu, 2020] Kalluri, R. and LeBleu, V. S. (2020). Exosomes in cancer: keys to a paradigm shift in the understanding of tumor biology. *Cell*, 171(6):1227–1241.
- [Kalsheker, 1996] Kalsheker, N. A. (1996). Alpha 1-antichymotrypsin. *International Journal of Biochemistry and Cell Biology*.
- [Kaufman and Rousseeuw, 1990] Kaufman, L. and Rousseeuw, P. (1990). *Finding Groups in Data: An Introduction to Cluster Analysis*. John Wiley Sons.
- [Kaur et al., 2016] Kaur, A. et al. (2016). Human c1q induces apoptosis in an immature dendritic cell line through regulation of akt pathway, and p21 gene expression. *Frontiers in Immunology*.
- [Keller and Pantel, 2022] Keller, L. and Pantel, K. (2022). The role of circulating tumor dna in early cancer detection. *Nature Reviews Cancer*, 22(2):91–102.
- [Khan et al., 2016] Khan, T. A. et al. (2016). Immunoglobulin kappa light chain. *Journal of Immunology*, 196(4):1655–1662.
- [Khorchid and Ikura, 2002] Khorchid, A. and Ikura, M. (2002). How calmodulin binds its targets: sequence versus shape. *Journal of Biological Chemistry*.

- [Kishore et al., 2014] Kishore, U. et al. (2014). Complement: structure, functions, evolution, and viral molecular mimicry. *Immunology Research*.
- [Klamminger et al., 2021] Klamminger, G. G., Gérardy, J.-J., Jelke, F., et al. (2021). Application of raman spectroscopy for detection of histologically distinct areas in formalin-fixed paraffin-embedded glioblastoma. *Neuro-Oncology Advances*, 3(1):vdab077.
- [Kneipp et al., 2002] Kneipp, K., Moskovits, M., and Kneipp, H. (2002). Surface-enhanced raman scattering: Physics and applications. *Springer*, 103(2):13–45.
- [Kneipp et al., 1997] Kneipp, K., Wang, Y., Kneipp, H., Perelman, L. T., Itzkan, I., Dasari, R. R., and Feld, M. S. (1997). Single molecule detection using surface-enhanced raman scattering (sers). *Physical Review Letters*, 78(9):1667.
- [Kowal et al., 2014] Kowal, J., Tkach, M., and Théry, C. (2014). Biogenesis and secretion of exosomes. *Current opinion in cell biology*, 29:116–125.
- [Lane et al., 2015] Lane, R. E., Korbie, D., Hill, M. M., and Trau, M. (2015). Microvesicles and exosomes: A new horizon of biomedical research. *Proteomics*, 15(17):187–192.
- [Langer et al., 2019] Langer, J., Jimenez de Aberasturi, D., Aizpurua, J., et al. (2019). Present and future of surface-enhanced raman scattering. *ACS Nano*, 14(1):28–117.
- [Langfelder and Horvath, 2008] Langfelder, P. and Horvath, S. (2008). Wgcna: an r package for weighted correlation network analysis. *BMC Bioinformatics*, 9:559.
- [Langlois and Delanghe, 1996] Langlois, M. R. and Delanghe, J. R. (1996). Biological and clinical significance of haptoglobin polymorphism in humans. *Clinical Chemistry*.

- [Leavesley et al., 2013] Leavesley, D. I. et al. (2013). Vitronectin—master controller or micromanager? *IUBMB Life*.
- [LeCun et al., 2015] LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. *Nature*, 521(7553):436–444.
- [Lefranc, 1997] Lefranc, M. P. (1997). Unique database numbering system for immunogenetic analysis. *Immunology Today*.
- [Lefranc, 2001] Lefranc, M. P. (2001). *The immunoglobulin facts book*. Academic Press.
- [Li et al., 2003] Li, W. F. et al. (2003). Paraoxonase 1. *Current Medicinal Chemistry*, 10(14):1677–1694.
- [Li et al., 2021] Li, X., Li, S., He, Z., Zhao, R., Guo, L., and Zhang, F. (2021). Surface plasmon resonance (spr) biosensor for label-free detection of exosomes from breast cancer cells. *Biosensors and Bioelectronics*, 164:112292.
- [Lim, 2001] Lim, Y. K. (2001). Haptoglobin polymorphism and immune response. *Biomedicine Pharmacotherapy*.
- [Lipton, 2018] Lipton, Z. C. (2018). The mythos of model interpretability: In machine learning, the concept of interpretability is both important and slippery. *Queue*, 16(3):31–57.
- [Litjens et al., 2017] Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., van der Laak, J. A., van Ginneken, B., and Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42:60–88.
- [Liu et al., 2017] Liu, F., Vermesh, O., Mani, V., Ge, T. J., Madsen, S. J., Sabour, A., Hsu, E.-C., Gowrishankar, G., Kanada, M., Jokerst, J. V., et al. (2017). The exosome total isolation chip. *ACS nano*, 11(11):10712–10723.

- [Liu et al., 2019] Liu, Y., Jin, Y., Wang, J., and Lin, Z. (2019). Raman spectroscopy for biomolecular detection in cancers. *Oncotarget*, 10(20):2082.
- [Louis et al., 2021a] Louis, D. N., Perry, A., Wesseling, P., Brat, D. J., Cree, I. A., Figarella-Branger, D., Hawkins, C., Ng, H. K., Pfister, S. M., Reifenberger, G., Soffietti, R., von Deimling, A., and Ellison, D. W. (2021a). The 2021 who classification of tumors of the central nervous system: a summary. *Neuro-Oncology*, 23(8):1231–1251.
- [Louis et al., 2021b] Louis, D. N., Perry, A., Wesseling, P., Brat, D. J., Cree, I. A., Figarella-Branger, D., Hawkins, C., Ng, H. K., Pfister, S. M., Reifenberger, G., Soffietti, R., von Deimling, A., and Ellison, D. W. (2021b). Glioblastoma, idh-wild type. pages 52–56.
- [Louis et al., 2021c] Louis, D. N., Perry, A., Wesseling, P., et al. (2021c). The 2021 who classification of tumors of the central nervous system: a summary. *Neuro-oncology*, 23(8):1231–1251.
- [Louis et al., 2021d] Louis, D. N., Perry, A., Wesseling, P., et al. (2021d). The diagnostic and molecular criteria of the 2021 who classification of central nervous system tumors: a practical approach. *Acta Neuropathologica*, 141:803–820.
- [Luchansky and Bailey, 2012] Luchansky, M. S. and Bailey, R. C. (2012). Surface acoustic wave biosensor for detection of microrna and other biomarkers. *Analytical Chemistry*, 84(2):793–800.
- [Lundberg and Lee, 2017] Lundberg, S. M. and Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30:4765–4774.
- [Ma et al., 2022] Ma, Y., Li, W., Yin, Y., Wang, X., and Gao, C. (2022). Optical biosensors for real-time detection of circulating tumor cells: A review. *TrAC Trends in Analytical Chemistry*, 148:116522.

- [Macpherson and Slack, 2007] Macpherson, A. J. and Slack, E. (2007). The functional interactions of commensal bacteria with intestinal secretory iga. *Current Opinion in Gastroenterology*.
- [Maeda et al., 1994] Maeda, N. et al. (1994). Increased atherosclerosis in apolipoprotein c-iii transgenic mice. *Journal of Clinical Investigation*.
- [Maguire et al., 2015] Maguire, A., Lyng, F. M., Geraghty, J., Casey, A., and Byrne, H. J. (2015). Competitive evaluation of data mining algorithms for use in classification of leukocyte subtypes with raman microspectroscopy. *Analyst*, 140(8):2473–2481.
- [Mandel et al., 2024] Mandel, P. et al. (2024). Advancements in liquid biopsy techniques for monitoring brain tumors. *Journal of Clinical Oncology*, 42(4):220–230.
- [Mann et al., 1988] Mann, K. G. et al. (1988). Cofactor proteins in the assembly and expression of blood clotting enzyme complexes. *Annual Review of Biochemistry*.
- [Mao and Schwarzbauer, 2005] Mao, Y. and Schwarzbauer, J. E. (2005). Fibronectin fibrillogenesis, a cell-mediated matrix assembly process. *Matrix Biology*.
- [Matsuda et al., 1998] Matsuda, F. et al. (1998). The complete nucleotide sequence of the human immunoglobulin heavy chain variable region locus. *Journal of Experimental Medicine*.
- [Matsushita, 2012] Matsushita, M. (2012). Ficolin-3 in the complement system. *Immunobiology*, 217(11):1196–1201.
- [Matsushita et al., 2001] Matsushita, M. et al. (2001). Ficolins and the lectin complement pathway. *Immunology Reviews*.
- [Matsuura et al., 1994] Matsuura, E. et al. (1994). Apolipoprotein h. *Journal of Clinical Investigation*, 94(3):978–985.

- [McKinney, 2010] McKinney, W. (2010). Data structures for statistical computing in python. *Proceedings of the 9th Python in Science Conference*, 445:51–56.
- [Melo et al., 2015] Melo, S. A. et al. (2015). Cancer exosomes perform cell-independent microrna biogenesis and promote tumorigenesis. *Cancer cell*, 26(5):707–721.
- [Merker et al., 2018] Merker, J. D. et al. (2018). Circulating tumor dna analysis in patients with cancer: American society of clinical oncology and college of american pathologists joint review. *Journal of Clinical Oncology*, 36(16):1631–1641.
- [Michael et al., 2019] Michael, A. P., Watson, V. L., Ryan, D., Delfino, K. R., Bekker, S. V., and Cozzens, J. W. (2019). Effects of 5-ala dose on resection of glioblastoma. *J Neurooncol*, 141(3):523–531.
- [Michalet et al., 2005] Michalet, X., Pinaud, F. F., Bentolila, L. A., Tsay, J. M., Doose, S., Li, J. J., Sundaresan, G., Wu, A. M., Gambhir, S. S., and Weiss, S. (2005). Quantum dots for live cells, in vivo imaging, and diagnostics. *Science*, 307(5709):538–544.
- [Miller and Perry, 2017] Miller, C. R. and Perry, A. (2017). Brain cancer: biology, genetics, and cell biology. *Acta neuropathologica*, 134(3):349–370.
- [Miyashita and Reed, 1995] Miyashita, T. and Reed, J. C. (1995). Bnip5. *Cell*, 80(2):293–299.
- [Mohamed et al., 2024] Mohamed, A. A., Sargent, E., Williams, C., Karve, Z., Nair, K., and Lucke-Wold, B. (2024). Advancements in neurosurgical intraoperative histology. *Tomography*, 10(5):693–704.
- [Morgan, 2016] Morgan, B. P. (2016). The membrane attack complex as an inflammatory trigger. *Immunobiology*.

- [Muller Bark et al., 2020] Muller Bark, J., Kulasinghe, A., Chua, B., et al. (2020). Circulating biomarkers in patients with glioblastoma. *British Journal of Cancer*, 122(3):295–305.
- [Muszbek et al., 1999] Muszbek, L. et al. (1999). Factor xiii: structure and function. *Thrombosis Research*.
- [Nie and Emery, 1997] Nie, S. and Emery, S. R. (1997). Probing single molecules and single nanoparticles by surface-enhanced raman scattering. *Science*, 275(5303):1102–1106.
- [Nieder et al., 2005] Nieder, C., Grosu, A. L., Astner, S., and Molls, M. (2005). Treatment of unresectable glioblastoma multiforme. *Anticancer Research*, 25(6C):4605–4610.
- [Nikitovic et al., 2008] Nikitovic, D. et al. (2008). Lumican, a small leucine-rich proteoglycan. *IUBMB Life*.
- [Noris et al., 2013] Noris, M. et al. (2013). Complement c1s and cancer. *Nature Reviews Nephrology*, 9(1):423–433.
- [O’Brien et al., 2013] O’Brien, K. et al. (2013). Glioblastoma-derived exosomes promote tumorigenesis through different mechanisms. *Acta Neuropathologica*, 126(5):703–720.
- [Olofsson et al., 1999] Olofsson, S. O. et al. (1999). Apolipoprotein b: a clinically important apolipoprotein which assembles atherogenic lipoproteins and promotes the development of atherosclerosis. *Journal of Internal Medicine*.
- [Olsson et al., 2004] Olsson, A. K. et al. (2004). Histidine-rich glycoprotein, a novel regulator of angiogenesis. *Journal of Thrombosis and Haemostasis*.

- [O'Neill and Lee, 2021] O'Neill, B. P. and Lee, E. Q. (2021). Update on the 2021 who classification of tumors of the central nervous system: What the neuroradiologist needs to know. *Radiographics*, 41(3):764–784.
- [Ostrom et al., 2021] Ostrom, Q. T., Cioffi, G., Waite, K., et al. (2021). Cbtrus statistical report: Primary brain and other central nervous system tumors diagnosed in the united states in 2014–2018. *Neuro-oncology*, 23(12):III1–III105.
- [Otsuji et al., 2024] Otsuji, R. et al. (2024). Liquid biopsy for glioma using cell-free dna in cerebrospinal fluid. *Cancers (Basel)*.
- [Palumbo et al., 2000] Palumbo, J. S. et al. (2000). Fibrinogen is an important determinant of the metastatic potential of circulating tumor cells. *Blood*.
- [Pang et al., 2016] Pang, X. et al. (2016). Coagulation factor x. *International Journal of Molecular Sciences*, 17(1):112.
- [Pankov and Yamada, 2002] Pankov, R. and Yamada, K. M. (2002). Fibronectin at a glance. *Journal of Cell Science*.
- [Pantel and Alix-Panabières, 2019] Pantel, K. and Alix-Panabières, C. (2019). Circulating tumor cells in cancer patients: challenges and perspectives. *Trends in Molecular Medicine*, 25(5):395–416.
- [Park et al., 2017] Park, J., Hwang, M., Choi, Y., Kang, M., Bang, J. H., Kim, H., Lee, S. H., and Choo, J. (2017). Exosome classification by pattern analysis of surface-enhanced raman spectroscopy data for lung cancer diagnosis. *Analytical Chemistry*, 89(12):6695–6701.
- [Patnaik and Moll, 2008] Patnaik, M. M. and Moll, S. (2008). Inherited antithrombin deficiency: a review. *Haemophilia*.
- [Paul et al., 2006] Paul, M. et al. (2006). Regulation of angiotensinogen gene expression. *Hypertension*.

- [Peinado et al., 2012] Peinado, H. et al. (2012). Exosomes: a mechanism of cellular communication. *Cancer Cell*, 21(4):618–630.
- [Philipp Balermipas and Hegi, 2021] Philipp Balermipas, K. E. W. and Hegi, M. E. (2021). Clinical advances in the treatment of glioblastoma: A review. *Cancer*, 127(11):1947–1955.
- [Pollard and Cooper, 2009] Pollard, T. D. and Cooper, J. A. (2009). Actin, a central player in cell shape and movement. *Science*.
- [Poon et al., 2011] Poon, I. K. et al. (2011). Histidine-rich glycoprotein: the swiss army knife of mammalian plasma. *Blood*.
- [Poon et al., 2002] Poon, S. et al. (2002). Clusterin is an extracellular chaperone that influences amyloid formation. *Nature Reviews Molecular Cell Biology*.
- [Preissner, 1991a] Preissner, K. T. (1991a). Structure and biological role of vitronectin. *Annual Review of Cell Biology*.
- [Preissner, 1991b] Preissner, K. T. (1991b). Vitronectin. *The FASEB Journal*, 5(8):2215–2223.
- [Puget et al., 2015] Puget, S. et al. (2015). Biopsy in a series of 130 pediatric diffuse intrinsic pontine gliomas. *Child’s Nervous System*, 31(10):1773–1780.
- [Quadro et al., 2005] Quadro, L. et al. (2005). Role of retinol and retinol-binding protein in the metabolism of retinoids. *Nutrition Reviews*.
- [Quek and Hill, 2017] Quek, C. and Hill, A. F. (2017). Pathogenic role of exosomes in neurodegenerative diseases. *Journal of Neuroscience Research*, 95(10):1941–1959.

- [Raposo and Stoorvogel, 2013] Raposo, G. and Stoorvogel, W. (2013). Exosomes: new vesicles for familiar functions. *Nature Reviews Molecular Cell Biology*, 14(2):135–147.
- [Ribeiro et al., 2016] Ribeiro, M. T., Singh, S., and Guestrin, C. (2016). "why should i trust you?": Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1135–1144.
- [Ricklin et al., 2010] Ricklin, D. et al. (2010). Complement c3. *Nature Reviews Immunology*, 10(1):49–62.
- [Roberts and Monroe, 2010] Roberts, H. R. and Monroe, D. M. (2010). Escaping the thrombin trap: insights into the hemostatic function of factor xi. *Blood*.
- [Rolnick et al., 2017] Rolnick, D., Veit, A., Belongie, S., and Shavit, N. (2017). Deep learning is robust to massive label noise. *arXiv preprint arXiv:1705.10694*.
- [Roubey, 1994] Roubey, R. A. (1994). Autoantibodies to phospholipid-binding plasma proteins: a new view of lupus anticoagulants and other 'antiphospholipid' autoantibodies. *Blood*.
- [Rouet et al., 2005] Rouet, L. et al. (2005). Inter-alpha-trypsin inhibitor family. *Biochimie*, 87(1):249–263.
- [Rudin, 2019] Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5):206–215.
- [Rye and Barter, 2014] Rye, K. A. and Barter, P. J. (2014). Cardioprotective functions of hdl. *Journal of Lipid Research*.

- [Saar-Tsechansky and Provost, 2007] Saar-Tsechansky, M. and Provost, F. (2007). Handling missing values when applying classification models. *Journal of Machine Learning Research*, 8:1623–1657.
- [Sahu et al., 2013] Sahu, R., Madhuri, R., Prakash, S., Reddy, M., Deepa, P., and Chandra, B. (2013). Raman spectroscopy of blood plasma: implications for breast cancer detection. *Vibrational Spectroscopy*, 68:195–203.
- [Salier, 1996] Salier, J. P. (1996). Inter-alpha-trypsin inhibitor: emergence of a family within the kunitz-type protease inhibitor superfamily. *Trends in Biochemical Sciences*.
- [Sanggaard et al., 2005] Sanggaard, K. W. et al. (2005). The tsg-6 and ii interaction promotes a procoagulant phenotype in breast cancer cells. *Journal of Biological Chemistry*.
- [Saraiva, 2001] Saraiva, M. J. (2001). Transthyretin amyloidosis. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, 1537(1):11–22.
- [Sasaki et al., 2007] Sasaki, T. et al. (2007). Galectin-3 and galectin-3-binding protein in tumor progression and metastasis. *Glycoconjugate Journal*.
- [Sattlecker et al., 2010] Sattlecker, M., Bessant, C., Smith, J. E., and Stone, N. (2010). Investigation of support vector machines and raman spectroscopy for lymph node diagnostics. *Analyst*, 135(5):895–901.
- [Schroeder and Cavacini, 2010a] Schroeder, H. W. and Cavacini, L. (2010a). Immunoglobulins. *Journal of Allergy and Clinical Immunology*, 125(2):S41–S52.
- [Schroeder and Cavacini, 2010b] Schroeder, H. W. J. and Cavacini, L. (2010b). Structure and function of immunoglobulins. *Journal of Allergy and Clinical Immunology*.

- [Schumann et al., 1990] Schumann, R. R. et al. (1990). Lipopolysaccharide-binding protein. *Science*, 249(4975):1429–1431.
- [Schwarzenbach et al., 2014] Schwarzenbach, H., Hoon, D. S., and Pantel, K. (2014). Cell-free nucleic acids as biomarkers in cancer patients. *Nature Reviews Cancer*, 11(6):426–437.
- [Seifert et al., 2020] Seifert, S., Kasper, R., Dorn, A., Schnabel, U., Panne, U., Riedel, J., Stiebing, C., Krafft, C., and Popp, J. (2020). Application of random forest based approaches to surface-enhanced raman scattering data. *Scientific Reports*, 10(1):5436.
- [Shao et al., 2021] Shao, H., Im, H., Castro, C. M., Breakefield, X. O., Weissleder, R., and Lee, H. (2021). Mass spectrometry-based proteomic and metabolomic profiling of exosomes derived from glioblastoma multiforme. *Nature Communications*, 12(1):1–12.
- [Shariat-Madar et al., 2002] Shariat-Madar, Z. et al. (2002). Structure and function of plasma prekallikrein and prokallikrein. *Thrombosis and Haemostasis*.
- [Shehzad et al., 2021] Shehzad, A., Islam, S. U., Shahzad, R., Khan, S., and Lee, Y. S. (2021). Extracellular vesicles in cancer diagnostics and therapeutics. *Pharmacology Therapeutics*, 223:107806.
- [Shin and Kwon, 2013] Shin, D. and Kwon, Y. (2013). Protein networks uncovered: a systems approach to probing cell mechanisms. *Molecular BioSystems*, 9(7):1432–1441.
- [Shin et al., 2018] Shin, H., Jeong, H., Park, J., Hong, S., and Choi, Y. (2018). Correlation between cancerous exosomes and protein markers based on surface-enhanced raman spectroscopy (sers) and principal component analysis (pca). *ACS Sensors*, 3(12):2637–2643.

- [Shipp et al., 2019] Shipp, D. W., Sinjab, F., and Notingher, I. (2019). Raman spectroscopy for cancer diagnosis. *Nature Reviews Cancer*, 17(11):743–755.
- [Silantyev et al., 2019] Silantyev, A. S., Falzone, L., Libra, M., et al. (2019). Current and future trends on diagnosis and prognosis of glioblastoma: from molecular biology to proteomics. *Cells*, 8(8):863.
- [Silverman et al., 2001] Silverman, E. K. et al. (2001). Alpha1-antitrypsin deficiency: a model for the protease-antiprotease hypothesis of emphysema. *Chest*.
- [Simons and Raposo, 2011] Simons, M. and Raposo, G. (2011). Exosomes: vesicular carriers for intercellular communication. *Current opinion in cell biology*, 23(4):631–640.
- [Simpson-Haidaris and Rybarczyk, 2001] Simpson-Haidaris, P. J. and Rybarczyk, B. (2001). Histidine-rich glycoprotein. *Trends in Cardiovascular Medicine*, 11(3-4):123–128.
- [Siravegna et al., 2017] Siravegna, G., Marsoni, S., Siena, S., and Bardelli, A. (2017). Integrating liquid biopsies into the management of cancer. *Nature Reviews Clinical Oncology*, 14(9):531–548.
- [Sitter et al., 2009] Sitter, B., Bathen, T. F., Singstad, T. E., Fjøsne, H. E., Lundgren, S., Halgunset, J., and Gribbestad, I. S. (2009). Comparison of hr mas mr spectroscopic profiles of breast cancer tissue with clinical parameters. *NMR in Biomedicine*, 22(6):555–560.
- [Skog et al., 2008] Skog, J. et al. (2008). Glioblastoma microvesicles transport rna and proteins that promote tumour growth and provide diagnostic biomarkers. *Nature cell biology*, 10(12):1470–1476.
- [Soria et al., 2017] Soria, F. N. et al. (2017). Exosomes: new actors in the biology of dendritic cells. *European Journal of Immunology*, 47(1):33–45.

- [Sottoriva et al., 2013] Sottoriva, A. et al. (2013). Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. *Proceedings of the National Academy of Sciences*, 110(10):4009–4014.
- [Sproviero et al., 2018] Sproviero, D. et al. (2018). Pathological proteins are transported via exosomes from neuroglia to neurons in amyotrophic lateral sclerosis. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, 1864(1):3605–3612.
- [Srivastava et al., 2014] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15(56):1929–1958.
- [Stoyanov et al., 2022] Stoyanov, G. S., Dzhenkov, D. L., Ghenev, P. I., et al. (2022). Glioblastoma stem cells and their heterogeneity. *Biomolecules*, 12(2):272.
- [Strubell et al., 2019] Strubell, E., Ganesh, A., and McCallum, A. (2019). Energy and policy considerations for deep learning in nlp. *arXiv preprint arXiv:1906.02243*.
- [Stummer et al., 2013] Stummer, W. et al. (2013). 5-aminolevulinic acid-derived tumor fluorescence: The diagnostic accuracy of visible fluorescence qualities as corroborated by spectrometry and histology and postoperative imaging. *Neurosurgery*, 74(3):310–320.
- [Sun et al., 2017] Sun, C., Shrivastava, A., Singh, S., and Gupta, A. (2017). Revisiting unreasonable effectiveness of data in deep learning era. *arXiv preprint arXiv:1707.02968*.
- [Sun et al., 2010a] Sun, D. et al. (2010a). Exosomes: a new horizon in the pathogenesis, diagnosis and therapy of brain diseases. *Brain Research*, 1443:56–64.
- [Sun et al., 2010b] Sun, L. et al. (2010b). Role of serum amyloid a protein in inflammation and cancer. *Biochemical Pharmacology*.

- [Sven Schipmann and Stummer, 2022] Sven Schipmann, M. N. and Stummer, W. (2022). Recent advances in fluorescence-guided surgery for glioblastoma treatment. *Journal of Neurological Surgery Part A: Central European Neurosurgery*, 83(3):187–198.
- [Sze et al., 2017] Sze, V., Chen, Y.-H., Yang, T.-J., and Emer, J. S. (2017). Efficient processing of deep neural networks: A tutorial and survey. *Proceedings of the IEEE*, 105(12):2295–2329.
- [T. Garzon-Muvdi and Chaichana, 2017] T. Garzon-Muvdi, C. Kut, X. L. and Chaichana, K. L. (2017). Intraoperative imaging techniques for glioma surgery. *Future Oncol*, 13(19):1731–1745.
- [Tait et al., 1991] Tait, R. C. et al. (1991). Antithrombin deficiency. *British Journal of Haematology*.
- [Taylor and Gercel-Taylor, 2008] Taylor, D. D. and Gercel-Taylor, D. (2008). Exosomal microrna: a diagnostic biomarker for cancer. *Journal of clinical investigation*, 117(11):3366–3372.
- [Thakur et al., 2014] Thakur, B. K. et al. (2014). Double-stranded dna in exosomes: a novel biomarker in cancer detection. *Cell research*, 24(6):766–769.
- [Théry et al., 2002] Théry, C. et al. (2002). Exosomes: secreted vesicles and inter-cellular communications. *F1000Research*, 3:15.
- [Thomas et al., 2003] Thomas, E. A. et al. (2003). Apolipoprotein d. *Neurobiology of Aging*, 24(2):209–219.
- [Théry et al., 2018] Théry, C., Witwer, K. W., Aikawa, E., Alcaraz, M. J., Anderson, J. D., Andriantsitohaina, R., Antoniou, A., Arab, T., Archer, F., Atkin-Smith, G. K., et al. (2018). Minimal information for studies of extracellular vesicles 2018 (misev2018): a position statement of the international society for

- extracellular vesicles and update of the misev2014 guidelines. *Journal of extracellular vesicles*, 7(1):1535750.
- [Tiller et al., 2008] Tiller, T. et al. (2008). Ighv gene usage. *Journal of Experimental Medicine*, 205(10):2403–2414.
- [Timpl et al., 1984] Timpl, R. et al. (1984). Fibulin, a major co-purified component of nidogen and laminin purified from the mouse tumor engelbreth-holm-swarm. *European Journal of Biochemistry*.
- [Tkach and Théry, 2016] Tkach, M. and Théry, C. (2016). Evolving role of exosomes in cancer microenvironment. *Nature reviews cancer*, 16(11):760–774.
- [Tolosano et al., 2002] Tolosano, E. et al. (2002). Hemopexin: structure, function, and regulation. *DNA and Cell Biology*.
- [Treps et al., 2017] Treps, L., Perret, R., Edmond, S., Ricard, D., and Gavard, J. (2017). Glioblastoma stem-like cells secrete exosomal vegf to induce fenestrations in brain endothelial cells in vitro. *Oncogene*, 36(31):3900–3912.
- [Uhlar and Whitehead, 1999] Uhlar, C. M. and Whitehead, A. S. (1999). Serum amyloid a, the major vertebrate acute-phase reactant. *European Journal of Biochemistry*.
- [Ullah et al., 2018] Ullah, R., Khan, S., Ali, H., Shah, S., Ayaz, M., and Ullah, I. (2018). Raman spectroscopy combined with a support vector machine for differentiating between feeding male and female infants mother’s milk. *Biomedical Optics Express*, 9(2):844–851.
- [Vaisar et al., 2007] Vaisar, T. et al. (2007). Apolipoprotein a-i. *Journal of Lipid Research*, 48(2):437–448.

- [Valadi et al., 2007] Valadi, H. et al. (2007). Exosome-mediated transfer of mrnas and micrnas is a novel mechanism of genetic exchange between cells. *Nature cell biology*, 9(6):654–659.
- [van Dijken et al., 2017] van Dijken, B. R. J. et al. (2017). Diagnostic accuracy of magnetic resonance imaging techniques for treatment response evaluation in patients with high-grade glioma, a systematic review and meta-analysis. *European Radiology*, 27(10):4129–4144.
- [van Niel et al., 2018] van Niel, G., D’Angelo, G., and Raposo, G. (2018). Extracellular vesicles: exosomes and microvesicles, integrators of homeostasis. *Physiological reviews*, 98(1):19–51.
- [Vanni et al., 2009] Vanni, H. et al. (2009). Cigarette smoking induces overexpression of a fat-depleting gene *azgp1* in the human. *Chest*.
- [Verderio et al., 2012] Verderio, C. et al. (2012). Exosomes release is regulated by a calcium-dependent mechanism in cortical neurons. *Journal of Biological Chemistry*, 287(1):3801–3811.
- [Vo-Dinh et al., 1999] Vo-Dinh, T., Yan, F., and Stokes, D. L. (1999). Surface-enhanced raman spectroscopy for chemical and biological detection. *Sensors and Actuators B: Chemical*, 62(1):2–9.
- [Vuillermoz et al., 2004] Vuillermoz, B. et al. (2004). Lumican. *FEBS Letters*, 569(1-3):201–205.
- [Wan et al., 2017] Wan, J. C., Massie, C., Garcia-Corbacho, J., Mouliere, F., Brenton, J. D., Caldas, C., Pacey, S., Baird, R., and Rosenfeld, N. (2017). Liquid biopsies come of age: towards implementation of circulating tumour dna. *Nature Reviews Cancer*, 17(4):223–238.

- [Wang et al., 2022] Wang, H., Xie, S., Liu, H., Li, Z., He, W., Chen, W., and Liu, J. (2022). Application of nanopore sequencing technology in the detection of glioblastoma multiforme-associated mutations. *Journal of Neuro-Oncology*, 158(1):99–108.
- [Wang et al., 2013] Wang, X. et al. (2013). Lrg1 promotes angiogenesis by modulating endothelial tgf- signaling. *Journal of Clinical Investigation*.
- [Wang et al., 2012] Wang, Y. et al. (2012). Haptoglobin. *The American Journal of Pathology*, 180(4):1514–1523.
- [Waskom, 2021] Waskom, M. (2021). *Seaborn: Statistical Data Visualization*. O’Reilly Media.
- [Weisel, 2005] Weisel, J. W. (2005). Fibrinogen and fibrin. *Advances in Protein Chemistry*.
- [Weller et al., 2021] Weller, M., Wick, W., Aldape, K., et al. (2021). Molecular diagnostics and treatment of glioblastoma. *Nature Reviews Neurology*, 17(10):589–601.
- [Welsh et al., 2024] Welsh, J. A., Goberdhan, D. C., O’Driscoll, L., Buzas, E. I., Blenkiron, C., Bussolati, B., Cai, H., Di Vizio, D., Driedonks, T. A., Erdbrügger, U., et al. (2024). Minimal information for studies of extracellular vesicles (misev2023): From basic to advanced approaches. *Journal of extracellular vesicles*, 13(2):e12404.
- [Wen et al., 2020] Wen, P. Y. et al. (2020). Glioblastoma in adults: a society for neuro-oncology (sno) and european society of neuro-oncology (eano) consensus review on current management and future directions. *Neuro-Oncology*, 22(8):1073–1113.

- [Wilkinson and Schreiber, 2009] Wilkinson, M. and Schreiber, M. (2009). Exact and approximate distributions for the cluster analysis of protein sequence alignments. *BMC Bioinformatics*, 10:302.
- [Wolf, 2011] Wolf, B. (2011). Biotinidase deficiency. *Nature Reviews Endocrinology*, 7(1):56–63.
- [Woof and Kerr, 2006] Woof, J. M. and Kerr, M. A. (2006). The function of immunoglobulin a in immunity. *Journal of Pathology*.
- [Xu et al., 2019] Xu, R., Greening, D. W., Rai, A., Ji, H., and Simpson, R. J. (2019). Exosome detection and characterization using surface plasmon resonance techniques: A review. *Analytical and Bioanalytical Chemistry*, 411(23):5895–5906.
- [Xu et al., 2018] Xu, R., Rai, A., Chen, M., Suwakulsiri, W., Greening, D. W., and Simpson, R. J. (2018). Extracellular vesicles in cancer—implications for future improvements in cancer care. *Nature Reviews Clinical Oncology*, 15(10):617–638.
- [Yáñez-Mó et al., 2015] Yáñez-Mó, M. et al. (2015). The biologic properties of extracellular vesicles and their physiological functions. *Journal of Extracellular Vesicles*, 4:27066.
- [Yang et al., 2005] Yang, Q. et al. (2005). Serum retinol binding protein 4 contributes to insulin resistance in obesity and type 2 diabetes. *Nature*.
- [Young, 1990] Young, S. G. (1990). Recent progress in understanding apolipoprotein b. *Circulation*.
- [Zanetti, 2005] Zanetti, M. (2005). Immunoglobulin m. *Nature Reviews Immunology*, 5(4):281–292.
- [Zhang et al., 2015a] Zhang, H. et al. (2015a). Lrg1: a potential prognostic biomarker in non-small cell lung cancer. *American Journal of Cancer Research*.

- [Zhang et al., 2012] Zhang, J. et al. (2012). Prothrombin and cancer. *Seminars in Thrombosis and Hemostasis*, 38(6):567–573.
- [Zhang et al., 2020] Zhang, W., Jiang, Y., Wang, X., Liu, L., and Wang, W. (2020). Electrochemical detection of exosomes based on aptamer recognition and eis signal amplification. *Biosensors and Bioelectronics*, 165:112373.
- [Zhang et al., 2015b] Zhang, Y. et al. (2015b). Exosomes: biological functions and diagnostic potential in cancer. *Journal of translational medicine*, 13(1):1–12.
- [Zhang et al., 2013] Zhang, Y., Hong, H., and Cai, W. (2013). Label-free raman spectroscopy provides early indication of metastatic bone. *The Analyst*, 138(14):4070–4075.
- [Zhang et al., 2019] Zhang, Y., Mi, X., Tan, X., and Xiang, R. (2019). Recent progress on liquid biopsy analysis using surface-enhanced raman spectroscopy. *Theranostics*, 9(2):491.
- [Zhou et al., 2019] Zhou, Y. et al. (2019). Optical biopsy identification and grading of gliomas using label-free visible resonance raman spectroscopy. *J Biomed Opt*, 24(9):1–12.
- [Zhu et al., 2017] Zhu, H., Zhang, C., Shi, J., Zhou, L., Shi, L., Du, M., Wang, Z., Wang, X., Xiong, Y., and Li, H. (2017). Photonic crystal biosensors: the ultimate platform for rapid and sensitive detection of cancer markers. *Biosensors and Bioelectronics*, 94:207–217.
- [Zhu et al., 2019] Zhu, L., Wang, K., Cui, J., Liu, H., Bu, X., Ma, H., Wang, W., Hu, G., Chen, X., Liu, X., et al. (2019). Microfluidics for the detection of exosomes: a review. *ACS Applied Bio Materials*, 2(8):3741–3759.
- [Zhuang et al., 2011] Zhuang, X. et al. (2011). Exosome-mediated sirna delivery to suppress oncogene expression in glioma. *Nature Nanotechnology*, 6(1):15–23.

- [Zong et al., 2018] Zong, C., Xu, M., Xu, L., Wei, T., Ma, X., Zheng, X., Hu, R., and Ren, B. (2018). Sers-activated platforms for cancer immunotherapy: Simultaneous tracking of multiple immune functions. *Accounts of Chemical Research*, 51(12):2964–2972.

