



UNIONE EUROPEA
Fondo Sociale Europeo



Corso di Dottorato in Biomedicina, Neuroscienze e Diagnostica Avanzata

Dipartimento di Biomedicina, Neuroscienze e Diagnostica Avanzata

Settore Scientifico Disciplinare - IINF-05/A

AN OPERATING SYSTEM FOR AI-BASED APPLICATIONS IN RADIOLOGICAL AND ONCOLOGICAL IMAGING

**IL DOTTORE
TIZIANA CURRIERI**

**IL COORDINATORE
FABIO BUCCHIERI**

**IL TUTOR
SALVATORE VITABILE**

**CICLO XXXVII
ANNO CONSEGUIMENTO TITOLO 2025**

Tesi redatta con il sostegno finanziario del programma FSE-REACT-EU, PON “Ricerca e Innovazione” 2014-2020 (D.M. 1061/2021) Azione IV.4 “Dottorati e contratti di ricerca su tematiche dell'innovazione”

Summary

This doctoral thesis presents an integrated and multidisciplinary research path focused on the development, customization, and application of advanced computational tools for radiological workflows, with particular emphasis on open-source software environments and the integration of artificial intelligence. At the core of this work lies the design and implementation of DicomOS, a Linux-based operating system specifically conceived to meet the practical and technological needs of modern radiology departments. DicomOS combines a stable and customizable infrastructure with a suite of graphical and command-line applications, enabling clinicians and researchers to perform tasks such as annotation, anonymization, DICOM-to-image conversion, dataset organization, and longitudinal analysis with greater efficiency and transparency.

The system was designed with the dual aim of improving clinical workflow and supporting research activities. Its architecture facilitates the integration of existing tools and allows the development of new ones through modular interfaces. Particular attention was dedicated to the graphical user interfaces, which were conceived to lower the technical barrier for non-expert users, while maintaining advanced functionality for experienced researchers. The platform also supports command-line interaction, offering full control for scripting and automation in large-scale image processing pipelines.

The second major contribution of this work lies in the theoretical and practical exploration of artificial intelligence methodologies in the context of radiological image analysis. The thesis offers an in-depth review of both shallow and deep learning classifiers, discussing their respective advantages, limitations, and appropriate application contexts. A significant focus was placed on explainable artificial intelligence, which plays a crucial role in increasing trust, safety, and interpretability in clinical applications. Techniques such as SHAP, Grad-CAM, and attention-based methods were critically examined and applied to real-world medical scenarios.

Subsequently, the research addressed the training of Vision Transformer models in constrained data environments. By integrating traditional and diffusion-based data augmentation techniques, robust performance improvements were achieved in the classification of melanoma and mammographic images. These approaches were validated through multiple studies and conference contributions, highlighting the potential of ViT architectures even in low-data regimes typical of medical imaging.

In the final part of the thesis, the focus shifts toward the use of graph-based representations of brain connectivity derived from diffusion-weighted imaging. By extracting structural connectomes and applying both classical machine learning algorithms and graph neural networks, the thesis proposes an original framework for the classification of patients with traumatic brain injury. The analyses consider not only inter-group discrimination (acute, chronic, controls), but also longitudinal recovery trajectories. To ensure transparency and clinical relevance, explainability methods were employed to localize the most influential brain regions contributing to model predictions, offering new insights into the neural substrates of consciousness and recovery.

Taken as a whole, the contributions presented in this thesis advance the field of computational radiology by offering a complete, modular, and explainable framework that bridges the gap between system-level innovation and artificial intelligence methodologies. The proposed tools and methods promote reproducibility, adaptability, and interdisciplinary collaboration, supporting both clinical practice and scientific discovery.

Acknowledgements

These past three years have represented an intense, transformative, and deeply enriching journey, academically and personally. Completing this Ph.D. thesis has been one of the most demanding yet rewarding experiences of my life, and it would not have been possible without the support, guidance, and encouragement of many people who, in different ways, have walked this path with me.

First and foremost, I would like to express my deepest gratitude to Prof. Salvatore Vitabile, who has been not only my supervisor but also a true mentor. His constant support, availability, and scientific insight have guided me through the many complexities of this research. His trust in my work, his patience, and his encouragement have shaped my development as a researcher and made me grow as a person.

I am especially grateful to Ing. Francesco Prinzi, who was the very first person I met upon entering the laboratory. Throughout these years, he has been a friend, a colleague, and a guide—someone I could always rely on. His help, both technical and human, has been invaluable in navigating the various projects and challenges that have marked this journey. I would also mainly like to thank the other members of the Medical Imaging Lab, *Giovanni*, *Alberto*, *Carmelo*, and all the people who have passed through over the years. Each of them has contributed in their way to my experience, enriching me both professionally and personally. The spirit of collaboration, the exchange of ideas, and the informal moments shared have left a lasting impression.

A special thought goes to the people I met during my time in Kraków. That experience has been one of the most beautiful of my life. It broadened my horizons, introduced me to new perspectives, and opened doors to areas of knowledge and life I had never imagined before. I will always carry it with me.

I extend my heartfelt thanks to my entire family. My parents and my sister, who have always believed in me and supported me with love, patience, and encouragement, even from afar. In the last stretch of this journey, I have also been accompanied by the little Elena, who, despite her very young age, brought light and tenderness to my days, reminding me of the value of simple things. To her and to all those close to my heart, I owe the strength and serenity with which I have faced this final chapter.

Finally, I thank all my friends, old and new, near and far, for their presence, their affection, and their ability to make me smile even in the most difficult moments. Their support has reminded me, day after day, that even the most ambitious goals can be pursued when surrounded by people who care.

Contents

Summary	1
Acknowledgements	2
1 Introduction	9
1.1 Research Context	9
1.2 Objectives of the Thesis	9
1.3 Scientific Contributions	10
1.4 Structure of the Thesis	11
2 Background and State of the Art	13
2.1 Operating Systems in the Medical Domain	13
2.2 DICOM Standard and Interoperability Challenges	13
2.3 Limitations of Existing Medical Imaging Software	14
2.4 Artificial Intelligence in Radiology: An Overview	15
2.5 Human-Machine Interfaces in Medical Software Systems	16
2.5.1 Technical Overview of Graphical User Interfaces	16
3 Development and Customisation of DicomOS: Adapting Ubuntu for Medical Imaging	18
3.1 Motivations for the Development of DicomOS	18
3.2 System Setup and Customisation	19
3.2.1 Customisation of the Visual Environment	19
3.2.2 Branding and Identity Customisation	20
3.2.3 Integration of Medical Applications and Usability Enhancements	20
3.3 Integration of Medical Workflows and Command-Line Tools	22
3.3.1 Development of Graphical User Interfaces for Medical Imaging	22
3.3.2 Dicom-Annotation GUI: Technical Implementation and Capabilities	22
3.3.3 Image Converter GUI	24
3.3.4 DICOM Anonymizer GUI	25
3.3.5 DICOM File Search GUI	27
3.3.6 DICOM File Organizer GUI	28
3.3.7 DICOM Command Shell GUI	29
3.3.8 DICOM Server Navigator GUI	30
3.3.9 DWI Longitudinal Analysis GUI	31
3.3.10 Medical Image Editor GUI	33
3.3.11 Deployment of a Web-Based Interface for Melanoma Classification	34
3.4 Development of Command-Line Tools for Programmers	36
3.5 Evaluation of the System	37
4 Artificial Intelligence in Radiology: Background and Framework	44
4.1 Definition of Artificial Intelligence and its Role in Radiology	44
5 Classifiers in Radiological Machine Learning	46
5.1 Overview of Classification in Medical Imaging	46
5.2 Model Development: Training, Validation, and Testing	46
5.3 Shallow Learning Classifiers	47

5.3.1	Logistic Regression	47
5.3.2	Support Vector Machine	48
5.3.3	Tree Ensemble Methods	48
5.3.4	K-Nearest Neighbors	49
5.4	Deep Learning Classifiers	49
5.4.1	Foundations of Deep Neural Networks	51
5.4.2	Convolutional Neural Networks	52
5.4.3	Choosing the Right Classifier in Medical Imaging	52
5.5	Evaluation Metrics	53
6	Explainable Artificial Intelligence: Concepts and Techniques	55
6.1	Global and Local Explanation Strategies	55
6.2	Explainability in Deep Learning and Medical Imaging	55
7	AI Applications in Radiological and Oncological Imaging	57
7.1	Transformer Architectures: Foundations and Relevance in Medical AI	57
7.1.1	Vision Transformers for Medical Image Classification	58
7.2	Classical Data Augmentation in Medical Imaging	60
7.3	Diffuser-Based Data Augmentation for Synthetic Medical Imaging	60
8	Vision Transformer Applications in Oncological Imaging	62
8.1	Dermatoscopic Image Classification for Melanoma Detection	62
8.1.1	Dataset Description: ISIC 2017 Challenge	62
8.1.2	Classical Data Augmentation	63
8.1.3	Methodology	64
8.1.4	Results and Discussion on Melanoma Classification	65
8.2	Transformer-Based Approach to Melanoma Detection: Additional Findings	67
8.2.1	Methodology	67
8.2.2	Pre-processing and Model Pipeline	67
8.2.3	Multi-Head Attention Mechanism	67
8.2.4	Experimental Setup	68
8.2.5	Experimental Results	68
8.2.6	Discussion	71
8.3	ViT-based Classification of Mammogram Images	73
8.3.1	Dataset	73
8.3.2	Image Pre-processing	73
8.3.3	Geometric Data Augmentation	73
8.3.4	Diffusion-Based Augmentation	74
8.3.5	Experimental Setup	75
8.3.6	Results and Discussion on Mammography Classification	75
9	Graph-Based and Explainable Approaches to Traumatic Brain Injury	79
	Classification	79
9.1	Introduction	79
9.2	Dataset	80
9.3	Image Preprocessing and Connectome Construction	81
9.4	The Brain as a Graph: Formal Definition and Theoretical Rationale	83
9.5	Assessment of brain connectivity in traumatic brain injury through Statis-	
	tical Analysis and Explainable AI Techniques	84

9.5.1	Exploration of Graph Metrics in Brain Connectomes	84
9.5.2	SHAP: SHapley Additive exPlanations	85
9.5.3	Machine Learning Framework for TBI Classification	86
9.5.4	Topological Analysis and Group-wise Comparison	87
9.5.5	Statistical Validation	87
9.5.6	Correlation Analysis of Network Features	89
9.5.7	Classification Performance and Feature Interpretability	89
9.6	Exploiting Graph Convolutional Networks for Insightful Classification and Explanation of Traumatic Brain Injury	89
9.6.1	Graph Convolutional Networks (GCNs)	89
9.6.2	Gradient-weighted Class Activation Mapping (Grad-CAM) for Explainable Graph Models	92
9.6.3	Graph Metrics for Node-Level Analysis	93
9.6.4	Graph Convolutional Network Architecture and Training Strategy	97
9.6.5	Interpretability via Grad-CAM: Identifying Salient Brain Regions	97
9.7	Results	99
10	Conclusion and Future Directions	118

List of Figures

1	Boot splash screen: Ubuntu vs DicomOS	20
2	DicomOS interface with custom GUI for clinical use	21
3	DicomOS workflow: integration of graphical and command-line tools	23
4	Creating GUI executables in DicomOS	24
5	DCMAnnotator GUI during annotation	25
6	Image Converter GUI: interface and conversion workflow	26
8	DICOM File Search GUI for metadata-based filtering	27
9	DICOM File Organizer GUI for automated sorting	28
10	DICOM Command Shell GUI with advanced image analysis tools	30
12	DWIAalyze GUI for longitudinal DWI comparison	32
14	Web app for melanoma classification with Flask and ViT	36
7	ImgAnonExtract GUI for anonymization and metadata extraction	39
11	DICOM Server Navigator GUI for remote image access	40
13	DicomViewer GUI with image filtering tools	41
15	Manual page of the DicomOS Convert command-line tool	42
16	Hard vs. soft margin in support vector machines	49
17	Polynomial kernel for data separation	50
18	Random forest example: majority vote for tumour classification	50
19	K-nearest neighbours: classification based on majority vote	51
20	Multilayer perceptron architecture with input, hidden, and output layers	52
21	Convolutional neural network architecture for image classification	53
22	Vision Transformer architecture used in this study	59
23	Diffuser architecture for synthetic image generation	61
24	Example images from the dataset with three classes	63
25	Pre-processing techniques on dermatoscopic image	64
26	Real vs synthetic melanoma image generated with DDPM	65
27	Confusion matrices of ViT classifier with and without data augmentation	65
28	End-to-end workflow implemented in the study	68
29	Accuracy vs. number of Transformer layers	70
30	Training and validation loss curves for best ViT configuration	70
31	Confusion matrix for best ViT-Large configuration	71
32	Pre-processing pipeline for CBIS-DDSM dataset	74
33	Test confusion matrices for the different data augmentation methods	77
34	Training and validation losses for the different data augmentation methods	77
35	Attention map example highlighting malignant lesion	78
36	Processing pipeline from raw DWI to connectome	82
37	Atlas alignment with AAL3v1 for brain parcellation	82
38	Neural connectivity metrics across acute, chronic, and control cohorts	88
39	Pearson correlation matrices of the three classes	90
40	SHAP interpretation of model predictions for TBI classification	91
41	Spearman correlation between node-level connectome metrics	96
42	GCN architecture for subject-level TBI classification	98
43	Feature transformation pipeline through GCN layers	98
44	Graph of brain connectivity in acute patient with Grad-CAM highlights	102
45	Brain network views highlighting Grad-CAM relevant nodes	103
46	Significant brain regions in acute patients	104

47	Significant brain regions in controls patients	104
48	Significant brain regions in chronic patients	105
49	Visualization of significant brain regions in acute DWI data	106
50	Significant brain regions in controls from DWI data	107
51	Significant brain regions in chronic TBI patients	108
52	Significant brain regions in VS patients	109
53	Significant brain regions in MCS patients	110
54	Significant brain regions in recovered chronic TBI patients	111

List of Tables

1	Comparison of DicomOS with other open-source systems	19
2	Summary of DicomOS GUIs for medical imaging	43
3	ISIC dataset summary	63
4	ISIC dataset after augmentation	64
5	Performance of ViT classifier under different data augmentation settings	66
6	Model hyperparameters	69
7	Tested configurations and their classification accuracy	70
8	Model Test Performance on the ISIC 2017 dataset.	72
9	Vision Transformer (ViT-Base) hyperparameters.	75
10	Classification results with different data augmentation techniques	76
11	Accuracy comparison with literature on CBIS-DDSM	78
12	Hyperparameters for each model	87
13	Machine learning results for TBI classification	89
14	Summary of GCN Architecture and Training Configuration	100
15	Classification Performance Metrics of the GCN Model	101
16	Grouping of significant brain regions in acute patients	112
17	Grouping of significant brain regions in control group	113
18	Grouping of significant brain regions in chronic patients	113
19	Significant brain regions in MCS patients	114
20	Significant brain regions in VS patients	114
21	Grouped brain regions in recovered chronic TBI patients	115

1 Introduction

1.1 Research Context

In recent years, medical imaging has become a cornerstone of clinical diagnostics, supporting physicians in the detection and monitoring of numerous diseases. Advanced technologies such as computed tomography (CT), magnetic resonance imaging (MRI), and positron emission tomography (PET) have allowed representations of anatomical structures to be obtained, significantly enhancing diagnostic accuracy. However, this evolution has also led to an exponential increase in the volume of imaging data generated, posing new challenges in data management, storage, and large-scale analysis.

One of the critical aspects of modern medical imaging is the need for specialized software capable of efficiently handling the growing number of images produced daily in hospitals and clinics. Most current information technology infrastructures in healthcare institutions rely on a combination of proprietary and open-source software solutions, often plagued by interoperability issues and compatibility gaps between different platforms and devices. The DICOM (Digital Imaging and Communications in Medicine) standard was introduced to address these challenges, providing a common protocol for the storage, transmission, and display of medical images. However, despite its widespread adoption, DICOM is not always implemented consistently across systems, resulting in inefficiencies and bottlenecks in radiological workflows.

Typically, medical images are processed using general-purpose operating systems such as Windows or macOS. These environments, however, are not designed to address the specialized workflows of radiology and often depend on external software that lacks interoperability and scalability. These systems often require additional third-party software for image management, frequently expensive, proprietary, and difficult to integrate, making scalability and workflow optimization challenging. Furthermore, many of these tools are not open-source, restricting customization and hindering research-driven innovation. To overcome these issues, the development of a dedicated operating system tailored to radiological image management has become increasingly relevant. DicomOS was conceived to meet this need, offering a Linux-based operating system optimized for the visualization, processing, and organization of DICOM images. The choice of a Linux distribution as the foundation for this project is motivated by its stability, security, and the extensive flexibility provided by the open-source ecosystem, which facilitates easy customization and integration with advanced tools. While DicomOS does not natively incorporate artificial intelligence algorithms, it provides a robust and flexible infrastructure for the integration of AI tools and models. The growing use of artificial intelligence (AI) in radiology, particularly for segmentation, classification, and predictive analytics, demands an operating environment that can accommodate and support such advancements without relying solely on external software ecosystems. This thesis presents the development of DicomOS, outlining its architecture and functionality, and assessing its potential impact on radiological practice. It further explores possible future enhancements, including integration with AI-driven tools, paving the way for a new paradigm in computational medical imaging.

1.2 Objectives of the Thesis

The overarching objective of this thesis is to design, develop, and present a complete, Linux-based operating system, DicomOS, tailored specifically to the needs of radiologists and researchers working with medical imaging. DicomOS is envisioned as an open-source,

modular, and lightweight environment that supports both clinical workflows and computational experiments.

More specifically, the thesis aims to:

- Develop a specialized Linux distribution optimized for radiological image management, with native support for DICOM standards.
- Provide a dual-interface system: a graphical environment for clinical users and a shell-based interface for researchers and developers.
- Enable seamless integration of external imaging tools and support extensions based on artificial intelligence applications.
- Promote transparency, customizability, and reproducibility in radiological workflows through an open-source framework.

1.3 Scientific Contributions

During the course of this Ph.D. program, several scientific contributions have been produced in the form of journal articles, conference papers, and peer-reviewed workshop proceedings. These works reflect the multidisciplinary nature of the research, which spans medical image analysis, artificial intelligence, operating systems for radiology, and explainable AI.

The first contribution, co-authored with colleagues from the Medical Imaging Laboratory, presents a comprehensive review of shallow and deep learning classifiers in the context of medical imaging. This article provides a systematic comparison of classical machine learning methods and neural architectures, highlighting their suitability for different diagnostic tasks in radiology.

A second line of research focused on the integration of Vision Transformers (ViT) and data augmentation strategies for dermatoscopic image classification. To overcome limitations associated with small datasets, diffusion-based synthetic data generation techniques were explored. These findings were first presented at the 18th Conference on Computational Intelligence Methods for Bioinformatics & Biostatistics (CIBB) and subsequently published in *Sensors*.

The same approach was extended to mammographic image analysis, where ViT-based classifiers were trained on augmented datasets for breast cancer detection. These results were presented at the Italian Workshop on Neural Networks (WIRN).

Another significant contribution is the development of DicomOS, a Linux-based operating system specifically tailored for radiological workflows and DICOM image management. The preliminary results of this system were published in *Electronics* and showcase the potential of open-source infrastructures in enhancing interoperability and clinical integration.

Finally, recent work has been dedicated to the assessment of brain connectivity in traumatic brain injury patients through graph-based methods and explainable artificial intelligence. A study combining statistical analysis and graph convolutional networks was accepted at the WIRN, aiming to characterize neurodegeneration and consciousness recovery through interpretable machine learning.

The full list of scientific contributions produced during the doctoral program is reported below:

- Prinzi F., **Currieri T.**, Gaglio S., & Vitabile S. (2024). Shallow and deep learning classifiers in medical image analysis. *European Radiology Experimental*, 8(1), 26. DOI: [10.1186/s41747-024-00428-2](https://doi.org/10.1186/s41747-024-00428-2)
- **Currieri T.**, Cicceri G., Cannata S., Cirrincione G., Lovino M., Militello C., Prinzi F., Pasero E., Vitabile S. Leveraging Diffuser Data Augmentation to enhance ViT-based performance on Dermatoscopic Melanoma Images Classification. *Lecture Notes in Computer Science*. In press.
- Cirrincione G., Cannata S., Cicceri G., Prinzi F., **Currieri T.**, Lovino M., ... & Vitabile S. (2023). Transformer-based approach to melanoma detection. *Sensors*, 23(12), 5677. DOI: [10.3390/s23125677](https://doi.org/10.3390/s23125677)
- Cannata S., Cicceri G., Cirrincione G., **Currieri T.**, Lovino M., Militello M., Prinzi F., Pasero E. & Vitabile S. (2023). ViT-based Classification of Mammogram Images: Impact of Data Augmentation Techniques. *The Italian Workshop on Neural Networks*. In press.
- **Currieri, T.**, Gambino, O., Pirrone, R., & Vitabile, S. (2025). DicomOS: A Preliminary Study on a Linux-Based Operating System Tailored for Medical Imaging and Enhanced Interoperability in Radiology Workflows. *Electronics*, 14(2), 330. DOI: [10.3390/electronics14020330](https://doi.org/10.3390/electronics14020330)
- **Currieri T.**, Falco-Roget J., Crimi A., Vitabile S. (2024) Assessment of Brain Connectivity in Traumatic Brain Injury through Statistical Analysis and Explainable AI Techniques. *Italian Workshop on Neural Networks*. Accepted.
- **Currieri T.**, Falco-Roget J., Rostami E., Vitabile S., Crimi A. (2025) Exploiting Graph Convolutional Networks for Insightful Classification and Explanation of Traumatic Brain Injury. *Computer Methods and Programs in Biomedicine*. Under Review.

1.4 Structure of the Thesis

This thesis is structured into ten main chapters, each addressing a specific aspect of the work developed during the doctoral project:

Chapter 1 - introduces the research context, outlines the objectives of the study, presents the main contributions, and provides an overview of the thesis structure.

Chapter 2 - provides a comprehensive overview of existing medical operating systems, the DICOM standard and its related interoperability challenges, and the limitations of currently available software tools for medical image management. It also introduces the increasing role of artificial intelligence in radiology and the relevance of human-machine interfaces in clinical environments.

Chapter 3 - details the development and customization of DicomOS, a Linux-based operating system tailored for radiology. It includes the motivations behind the project, the visual and functional customizations applied to the Ubuntu base, and the integration of medical workflows through both graphical and command-line tools.

Chapter 4 - focuses on the conceptual background and methodological framework underpinning the application of artificial intelligence to radiological data, setting the stage for the subsequent analysis of classification models.

Chapter 5 - presents a thorough overview of machine learning classifiers used in radiology, distinguishing between shallow and deep learning approaches. It discusses model development processes, performance metrics, and criteria for algorithm selection.

Chapter 6 - explores the core concepts and techniques of explainable artificial intelligence (XAI), highlighting their relevance in making AI-based decisions interpretable in medical contexts.

Chapter 7 - describes the application of Vision Transformer architectures to medical image classification tasks, including both dermatoscopic and mammographic datasets. The chapter includes classical and generative data augmentation strategies and presents experimental results.

Chapter 8 - presents an in-depth evaluation of transformer-based models in oncological imaging, detailing additional findings from melanoma detection and mammography analysis using ViT models and diffusion-based augmentation.

Chapter 9 - focuses on the application of graph-based approaches and explainable methods to the study of brain connectivity in traumatic brain injury. It presents both classical machine learning and graph neural network frameworks for patient classification, supported by statistical and topological analyses.

Chapter 10 - concludes the thesis by summarizing the contributions and results, and outlines future research directions for the continued development and integration of DicomOS in clinical workflows.

2 Background and State of the Art

2.1 Operating Systems in the Medical Domain

Operating systems play a foundational role in the management and execution of digital tools across various domains, including healthcare. In medical imaging, the operating system must ensure not only reliability and performance but also compliance with data protection standards and adaptability to complex clinical workflows. Traditional operating systems such as Windows and macOS, while widely adopted, often fall short when it comes to interoperability and integration with specialized imaging tools. These limitations are particularly pronounced in radiology, where cross-platform compatibility, streamlined workflows, and customization are essential. Linux has increasingly emerged as the preferred foundation for domain-specific operating systems due to its open-source nature, modular architecture, and active developer community. Several Linux-based distributions have been tailored to the biomedical field, most notably Debian Med, which offers a curated ecosystem of preinstalled applications for clinical and research purposes. Despite its utility, Debian Med relies heavily on third-party software and lacks native DICOM integration, limiting its effectiveness for radiology-specific workflows [1]. Similarly, open-source initiatives such as MONAI focus primarily on enabling AI model development for medical image analysis [2]. While powerful in terms of deep learning infrastructure, these platforms require significant programming expertise and do not offer cohesive environments for radiologists with limited technical backgrounds. In contrast, DicomOS was developed to bridge the gap between usability and customizability, offering a cohesive operating system with built-in DICOM support, a dual-interface paradigm (GUI for clinicians, command-line for developers), and compatibility with widely used imaging tools such as ITK-SNAP and 3D Slicer [3]. In addition to these existing solutions, other Linux-based projects illustrate the adaptability of open platforms to specialized needs. For instance, NeuroDebian and Lin4Neuro provide integrated environments for neuroimaging analysis, combining popular software packages like FSL, AFNI, and FreeSurfer within a reproducible ecosystem [4, 5]. These examples underscore the broader trend of customizing Linux for domain-specific workflows in healthcare. A major advantage of Linux lies in its ability to unify fragmented imaging workflows into a single environment. By natively integrating essential functionalities, such as image visualization, annotation, and anonymization, into the operating system, platforms like DicomOS aim to minimize the cognitive overhead associated with switching between multiple applications. The fragmentation typical of proprietary radiological ecosystems often results in workflow interruptions, reduced collaboration, and increased potential for error. A dedicated OS that consolidates these tasks into a seamless interface offers significant gains in clinical productivity and consistency. Overall, the evolution of Linux-based medical operating systems reflects the growing demand for flexible, secure, and interoperable platforms in clinical imaging. With the continued rise of AI and the need for real-time analytics, these systems will play an increasingly central role in enabling scalable and efficient medical workflows.

2.2 DICOM Standard and Interoperability Challenges

The DICOM standard is the cornerstone of digital medical imaging and has played a fundamental role in ensuring interoperability between imaging devices and software systems [6]. Introduced in the early 1990s, DICOM defines the structure and format of image files as well as the communication protocols used for transmitting these files across systems.

Its widespread adoption has enabled healthcare institutions to standardize the acquisition, storage, retrieval, and transfer of medical images across a variety of modalities and platforms. Despite the ubiquity of DICOM, the practical implementation of the standard often varies significantly across vendors, leading to a multitude of interoperability challenges. For example, while the DICOM file format and transmission protocols are well-defined, the actual content of DICOM metadata fields can differ across systems. Optional tags, vendor-specific extensions, and inconsistencies in how fields are populated result in heterogeneity that hampers seamless integration between imaging modalities, Picture Archiving and Communication Systems (PACS), and viewing or processing software [7].

One major obstacle is the lack of enforced uniformity in DICOM tag usage. While some fields are required, many are optional or context-specific, leading to a situation in which devices from different manufacturers may produce DICOM files that conform to the standard but are not truly interoperable without vendor-specific adaptations. This issue is particularly problematic in cross-institutional collaborations or research settings, where heterogeneous datasets need to be combined and analyzed systematically. Another challenge lies in the integration of DICOM into modern, AI-driven workflows. Most AI pipelines require preprocessing steps to extract image data from DICOM files and convert them into formats compatible with machine learning (ML) frameworks (e.g., NIFTI, PNG, NumPy arrays). This conversion step is non-trivial and often involves metadata loss, errors in orientation, or failure to handle multi-frame and volumetric data correctly [8]. Furthermore, annotations and segmentations, critical for training and evaluation, are not consistently encoded in standard DICOM files, leading to the use of external formats and a break in the semantic pipeline. Efforts have been made to address some of these shortcomings. The DICOMweb standard, for instance, introduces RESTful web services for image transfer and querying, aimed at facilitating access to DICOM data in cloud-based environments [9]. Similarly, initiatives like the Integrating the Healthcare Enterprise (IHE) Radiology Technical Framework promote harmonization through integration profiles and implementation guidelines. Nevertheless, real-world adoption of these technologies remains limited and fragmented. In this context, the development of systems like DicomOS becomes particularly valuable. By embedding DICOM support natively into the operating system and enabling compatibility with both traditional imaging tools and modern AI frameworks, DicomOS provides a unified environment for navigating these interoperability challenges. It not only supports legacy DICOM operations but also facilitates the extension of medical imaging workflows into data-driven and AI-enhanced domains.

2.3 Limitations of Existing Medical Imaging Software

Despite the rapid development of digital imaging technologies, current software ecosystems in medical imaging remain fragmented, inconsistent, and often poorly integrated into clinical workflows. Traditional PACS and Radiology Information Systems (RIS) are typically vendor-specific, closed-source, and rigid in their architecture, offering limited flexibility and customizability for research and advanced processing.

One of the most prominent limitations is the absence of native integration between essential imaging functionalities. Radiologists and researchers often rely on a combination of standalone tools for image viewing, annotation, anonymization, and export. This disjointed workflow increases cognitive load, introduces redundant operations, and creates room for human error [3]. Even open-source tools such as ITK-SNAP or 3D Slicer, while

powerful in their domain, require separate installation, manual configuration, and are not seamlessly embedded within a unified operating environment [10]. Another drawback is the lack of scripting capabilities and automation in many imaging tools. The inability to batch process datasets or to integrate software into shell-based workflows makes it difficult to handle large volumes of images, an increasingly common requirement in both clinical and research settings, especially when using artificial intelligence methods [11]. Furthermore, commercial platforms are often constrained by expensive licensing models and closed APIs. These factors inhibit extensibility, making it difficult to customize tools or to embed novel AI algorithms into existing workflows. This stands in contrast with modern scientific computing paradigms, where openness, modularity, and reproducibility are essential [12]. DicomOS was conceived to address these challenges. As a Linux-based operating system tailored for radiology, it integrates native DICOM support, graphical and command-line interfaces, and compatibility with third-party software. Unlike other open-source solutions focused solely on deep learning (e.g., MONAI), DicomOS offers an accessible, modular, and self-contained environment that bridges the gap between clinical usability and research-grade processing [3].

2.4 Artificial Intelligence in Radiology: An Overview

In recent years, AI has begun to transform the field of medical imaging, particularly radiology, by introducing new tools for automation, pattern recognition, and decision support. With the rise of deep learning and convolutional neural networks (CNNs), AI systems have demonstrated high accuracy in a range of radiological tasks, including lesion detection, segmentation, image classification, and even disease prognosis prediction [13, 14]. The integration of AI into radiological workflows offers significant advantages: improved diagnostic efficiency, reduced inter-reader variability, and the ability to handle large volumes of data with minimal human intervention. These benefits are particularly relevant in settings with limited resources or high clinical workload, where AI systems can act as second readers or triage assistants [15]. However, the widespread adoption of AI in radiology also presents several challenges. These include the need for high-quality annotated datasets, the difficulty in ensuring model generalization across institutions and imaging modalities, and the lack of interpretability of many deep learning models. The “black-box” nature of AI algorithms often raises concerns among clinicians and regulatory bodies, especially in high-stakes diagnostic contexts. As a result, the field of XAI has emerged as a critical area of research, aiming to provide transparency and accountability in medical decision-making [16]. Additionally, AI tools often require complex software environments for development and deployment, environments that are rarely available on standard radiology workstations. This disconnect between research prototypes and clinical systems remains a key barrier to effectively translating AI innovations into real-world practice. To bridge this gap, operating environments like DicomOS aim to offer an accessible, Linux-based platform that supports both traditional radiological workflows and the integration of AI-powered tools. By combining native DICOM compatibility with extensibility and scripting capabilities, DicomOS enables a new generation of hybrid systems that bring together imaging, automation, and explainability under a unified architecture.

2.5 Human-Machine Interfaces in Medical Software Systems

In the domain of medical informatics, Human-Machine Interfaces (HMIs) play a pivotal role in mediating the interaction between healthcare professionals and digital tools. As the volume and complexity of medical imaging data grow, the need for intuitive and task-specific interfaces becomes increasingly urgent. Traditional command-line tools, while powerful, often impose a steep learning curve and can hinder adoption by clinical personnel who lack technical training. This usability gap has driven the adoption of Graphical User Interfaces (GUIs) as a means to democratize access to advanced functionalities. Effective GUI design in radiology must adhere to domain-specific constraints, including support for fast-paced decision-making, seamless integration with existing data standards (such as DICOM), and minimization of cognitive workload. Research shows that well-designed HMIs improve clinical accuracy, reduce interpretation time, and enhance diagnostic confidence, particularly in high-stakes environments like emergency radiology and critical care [17]. Modern medical HMIs often serve as graphical front-ends to complex computational backends, enabling users to perform tasks such as image annotation, DICOM metadata editing, anonymization, and server-based queries through intuitive visual elements. This layered approach aligns with human-centered design principles, allowing different user roles, radiologists, technologists, researchers, and data scientists, to interact efficiently with the system according to their specific workflows and expertise [18].

In the context of DicomOS, a custom Linux-based operating system tailored for medical imaging, the development of domain-specific GUIs reflects a shift towards greater usability and accessibility in medical software. By embedding lightweight, modular interfaces within a pre-configured imaging platform, DicomOS supports both novice and expert users while preserving compatibility with powerful open-source libraries and pipelines. This approach exemplifies the convergence of usability engineering and medical data science, and provides a foundation for future systems that aim to be both robust and clinician-friendly [19].

2.5.1 Technical Overview of Graphical User Interfaces

GUIs are a class of human-computer interfaces that allow users to interact with software systems through graphical elements such as buttons, icons, menus, sliders, and windows. Unlike command-line interfaces (CLIs), which require textual input and familiarity with specific syntax, GUIs abstract the underlying complexity by presenting functions visually, thus improving usability and accessibility, especially in domains such as healthcare, where non-technical users are common. From a technical standpoint, a GUI typically comprises a presentation layer, responsible for rendering graphical elements; a control layer, which manages user input and events (e.g., mouse clicks, keyboard entries); and a logic layer, which interfaces with the backend system to execute the requested operations. These layers are often implemented using model-view-controller (MVC) or model-view-viewmodel (MVVM) architectural patterns, which enhance modularity, testability, and scalability [20, 21].

In medical software systems, GUI design must adhere to strict usability standards and regulatory guidelines (e.g., ISO 9241, IEC 62366), ensuring safety, traceability, and ergonomic consistency. The interface should support complex domain-specific workflows, such as radiological annotation, DICOM header manipulation, image segmentation, or integration with PACS systems, while minimizing cognitive load and error-proneness. As shown in several studies, a well-designed GUI can significantly accelerate diagnostic

procedures and reduce fatigue in high-demand environments such as teleradiology and emergency imaging [22, 19]. Modern medical GUIs are often developed using high-level toolkits such as Qt, GTK, Tkinter (Python), or JavaFX, which provide cross-platform compatibility and native-like rendering. They may also integrate with scripting engines to expose advanced functionalities (e.g., Python bindings for GUI automation), and with standard medical libraries for parsing, visualizing, and manipulating medical images and metadata. The modular approach adopted in systems like DicomOS exemplifies this integration, enabling researchers and clinicians to access powerful tools through intuitive visual workflows without needing to interact with low-level code. We can say that GUIs represent a critical component of contemporary medical imaging software, acting as the user-facing bridge between complex computational pipelines and clinical practice. Their role is not merely aesthetic but functional, impacting diagnostic accuracy, workflow efficiency, and ultimately patient care.

3 Development and Customisation of DicomOS: Adapting Ubuntu for Medical Imaging

3.1 Motivations for the Development of DicomOS

The development of DicomOS was driven by the growing need for an integrated, open-source environment specifically tailored to the demands of radiology. In clinical practice, radiologists are frequently required to manage large volumes of medical images using heterogeneous tools and platforms, often resulting in fragmented workflows and inefficient task execution. Commercial software, though feature-rich, is typically expensive, closed-source, and lacks flexibility for customization or research-oriented use cases. On the other hand, open-source solutions often require substantial technical knowledge and are not optimized for everyday clinical use. To address these limitations, DicomOS was conceived as a Linux-based operating system that consolidates essential functionalities, such as DICOM visualization, image annotation, and metadata management, into a unified platform. Built on Ubuntu 22.04 LTS (Long Term Support), it offers a dual interface: GUIs for clinicians and command-line tools for researchers and developers. This dual-mode interaction ensures that both technical and non-technical users can efficiently operate within the same environment [23]. A key motivation was also the lack of natively integrated DICOM workflows in existing Linux distributions such as Debian Med, which rely heavily on third-party software and lack GUI support for basic radiological operations. In contrast, DicomOS provides a complete ecosystem out-of-the-box, with pre-configured GUI applications for annotation, anonymization, image format conversion, and DICOM file organization. These applications follow best practices in human-computer interaction and were designed to minimize the learning curve, especially for users transitioning from macOS or Windows systems [24]. Moreover, DicomOS supports modular integration with powerful open-source tools such as ITK-SNAP and 3D Slicer. By incorporating these into the OS with launcher scripts and GUI accessibility, users can execute advanced imaging tasks without the need for manual configuration or terminal-based operations [25]. Importantly, the choice of Linux as a base was not only due to its open-source nature, but also because of its community support, high customizability, and long-term stability. This makes DicomOS an ideal foundation for continuous development, enabling institutions or individual users to adapt the system to emerging needs in radiology and medical AI. DicomOS was created to overcome the limitations of both proprietary and fragmented open-source imaging environments by providing a flexible, integrated, and accessible operating system for radiologists, researchers, and developers alike. It bridges the gap between usability and extensibility, supporting both clinical practice and experimentation.

To contextualize the motivations for developing DicomOS, Table I provides a comparative overview with two other well-known open-source systems used in medical imaging: Debian Med and MONAI. The comparison highlights key differentiators in terms of DICOM support, ease of use, tool integration, accessibility, and customisation. While Debian Med offers a broad but technically demanding ecosystem, and MONAI focuses primarily on AI-driven research, DicomOS aims to strike a balance between clinical usability and research flexibility.

Feature	DicomOS	Debian Med	MONAI
DICOM Support	Native integration with visualization, annotation, and data management capabilities	Limited support through third-party applications	No direct DICOM integration, focused on AI-based medical imaging workflows
Ease of Use	User-friendly graphical interface for clinicians and command-line tools for developers	Primarily oriented toward Linux experts	Requires advanced technical expertise for implementation and use
Integrated Tools	ITK-SNAP, 3D Slicer, and custom GUIs	Extensive ecosystem, but lacks automated integration	Powerful tools for AI training, but without GUI functionalities
Accessibility	Under development for ISO image distribution on physical hardware	Available as Debian packages	Open-source framework for Python
Customisation	Highly customisable due to its open-source nature and Linux foundation	Limited to the compatibility of available packages	Extendable via Python modules, focusing on AI workflows rather than clinical needs

Table 1: Comparison of DicomOS with other open-source systems for medical imaging. The table highlights the key features of each system, focusing on clinical functionalities and customisation capabilities.

3.2 System Setup and Customisation

The initial development of DicomOS was based on Ubuntu 22.04 Desktop LTS [23], selected due to its stability, extensive documentation, and compatibility with many open-source medical imaging tools. The system installation took place within a VirtualBox 6.1 [26] virtual machine hosted on a macOS High Sierra workstation equipped with a 2.3 GHz Intel Core i5 processor and 8 GB of DDR3 RAM. This virtualized approach allowed the customization process to proceed safely, offering flexibility and preventing any risk of interference with the host machine. Once the core installation was complete, VirtualBox Guest Additions were installed to optimize interaction between the virtual machine and the host system. These additions enabled improved graphics performance, seamless mouse integration, and clipboard sharing between environments, which collectively enhanced the overall experience during the development and testing phases.

3.2.1 Customisation of the Visual Environment

After setting up the system, considerable effort was dedicated to enhancing the graphical interface of Ubuntu to better accommodate the needs of clinical users, while preserving a simple and accessible design. The Canta GTK theme and Vimix icon set were selected for their clean, professional appearance. These visual enhancements aimed to reduce the learning curve for users less familiar with Linux-based systems, providing a more

intuitive user experience. The choice of the Vimix theme was intentional, as its aesthetic closely resembles the macOS interface, commonly adopted in the medical domain for its user-friendliness [27]. This visual familiarity was intended to support radiologists transitioning from Apple environments. The customization required manual adjustments to the GNOME desktop, where selected themes and extensions were integrated to refine the user interface. Notably, the "Dash to Dock" extension was installed to replace the default GNOME dock with a more functional and accessible alternative. This extension improved navigation and system usability, offering a more comfortable transition for users accustomed to different platforms [24].

3.2.2 Branding and Identity Customisation

In parallel with the visual adjustments, attention was given to building a coherent visual identity for DicomOS. The custom branding began with the boot splash screen, the first element seen by the user at startup. Plymouth, the tool responsible for managing Linux boot animations [28], was configured to replace the default Ubuntu visuals with a splash screen featuring the DicomOS logo and colour palette. Custom themes and configuration files were modified within Plymouth to produce this branded startup experience. As shown in Figure 1, the default Ubuntu splash screen (left) was replaced with a fully customized animation and branding that introduces the DicomOS identity (right), including its distinctive logo and colour scheme. Similarly, the system login screen was adapted to reflect DicomOS's visual identity, ensuring consistency throughout the user journey. Predefined login credentials were added to facilitate immediate access to the system, while still allowing users to create personal accounts for tailored use. Figure 2 displays the DicomOS desktop, illustrating the modernized visual theme, iconography, and custom applications designed to enhance usability in clinical settings. These design choices emphasize the flexibility of the system and its focus on user-centred design principles.

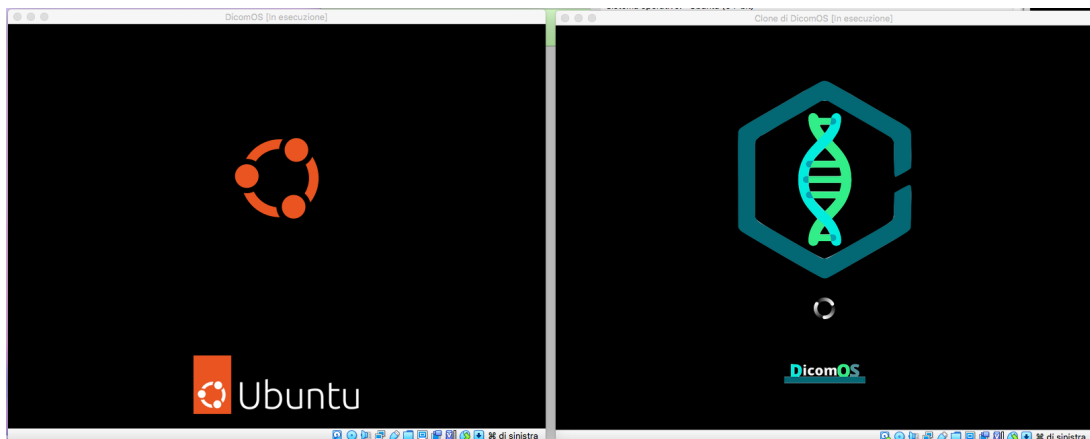


Figure 1: Comparison of boot splash screens: standard Ubuntu (left) versus the customized DicomOS splash screen (right), implemented via Plymouth with a custom logo and colour palette.

3.2.3 Integration of Medical Applications and Usability Enhancements

One of the primary objectives of DicomOS was to embed essential medical imaging applications directly into the operating system. Tools such as ITK-SNAP [29] and 3D Slicer



Figure 2: Screenshot of the DicomOS interface, showing the customised theme, icons, and new Graphical User Interface applications tailored for clinical use.

[25] were included because of their advanced capabilities for medical image segmentation and visualization [30, 31]. Despite their wide use, these tools do not support native double-click execution within a standard Ubuntu setup. To make them more accessible to users with limited technical expertise, additional steps were taken to make them executable like native applications. For 3D Slicer, the integration began with adjusting file permissions to enable extension installation. A custom shell script was then written to simplify execution, saved under `/usr/bin` to ensure terminal-wide availability. Initially limited to a specific version, the script was later generalized to support multiple Slicer versions using a pattern match (e.g., `Slicer*amd*`) for the installation directory. This ensured greater flexibility and future compatibility. Subsequently, a `.desktop` file was created, specifying the executable path, an icon path using the official logo, and categorization metadata. This file was placed in `/usr/share/applications`, making the software launchable through the GUI like any standard system application. This greatly improved accessibility for non-technical users [24]. A similar process was applied to ITK-SNAP to ensure consistency. By following the same pattern of creating executable scripts and launchers, ITK-SNAP was integrated seamlessly into DicomOS. Throughout the integration process, issues such as multi-version support and application stability were addressed with iterative testing and optimization. This testing included startup checks, version compatibility evaluations, and user experience reviews. Since both applications are well established, no changes have been made to their internal logic, an attempt has been made to improve their accessibility within the operating system environment. Additionally, improvements were made to support daily administrative tasks for clinicians and system maintainers. Commands like `upgrade`, `restart`, and `shutdown` were aliased within shell configuration files to simplify routine operations. These shortcuts allowed non-expert users to perform essential system functions efficiently, reducing cognitive load and streamlining usage [32, 24].

3.3 Integration of Medical Workflows and Command-Line Tools

DicomOS was designed to support both clinical usability and developer-level flexibility by providing graphical interfaces for non-technical users and command-line tools for advanced users.

3.3.1 Development of Graphical User Interfaces for Medical Imaging

To make the system usable for clinicians and researchers alike, DicomOS employs a dual-pathway architecture: one with GUIs for clinicians, and another with command-line access for programmers. Figure 3 illustrates this dual structure, where Python scripts link with shell commands and `.desktop` entries to create intuitive, double-clickable GUI applications. These GUIs simplify complex tasks such as DICOM annotation, anonymization, and image conversion for users unfamiliar with coding. Clinicians can operate the tools through menus and icons, while developers can run the same functionality via terminal commands, supported by detailed man pages for each tool. The GUI tools were built using Python 3.9, incorporating libraries like Tkinter [33] for UI development, matplotlib [34] for image rendering, and pydicom [35] for handling DICOM files [6]. These lightweight tools enable medical professionals to perform essential imaging operations without needing technical expertise. The executables were implemented through shell scripts that wrap the Python code, paired with `.desktop` entries that define metadata such as icon paths and categories. These were installed in the `/opt` directory, commonly used for third-party applications, and registered in the system application menu to appear like any native app.

High-resolution icons were added to maintain a polished, professional appearance. Appropriate file permissions were set to ensure usability for all system users. While *PyInstaller* [36] was initially considered for packaging, it failed to resolve all required dependencies. Instead, using shell wrappers allowed better control of the runtime environment. This approach not only improved usability but also ensured modularity and future extensibility. By abstracting the underlying complexity, users can launch applications without needing to interact with the Linux terminal. The method also supports scalability, allowing for the integration of future tools or updates.

Figures 3 and 4 visualize this architecture. The first diagram shows the dual user pathways, and the second breaks down the process of turning Python scripts into clickable GUI applications. In particular in Figure 3, the two sections are connected by a shared Python script layer, which supports GUI and command-line functionalities. This structure enables DicomOS to cater to the needs of both medical and technical users, providing an intuitive GUI for clinicians while offering direct, customisable access for programmers.

3.3.2 Dicom-Annotation GUI: Technical Implementation and Capabilities

Within the DicomOS framework, a core tool developed to support radiological annotation workflows is the **DCMAnnotator**, a graphical interface built to facilitate structured interaction with DICOM and non-DICOM images. This tool empowers radiologists and researchers to annotate regions of interest interactively through an intuitive and lightweight interface, reducing the friction typically encountered when using general-purpose annotation platforms. The graphical interface supports the import of various image formats, including `.dcm`, `.jpg` and `.png`, and leverages Matplotlib for interactive canvas rendering. The back-end logic integrates with *Pydicom* to ensure compatibility with DICOM metadata and pixel arrays. Users can add annotations in the form of lines, rectangles, and

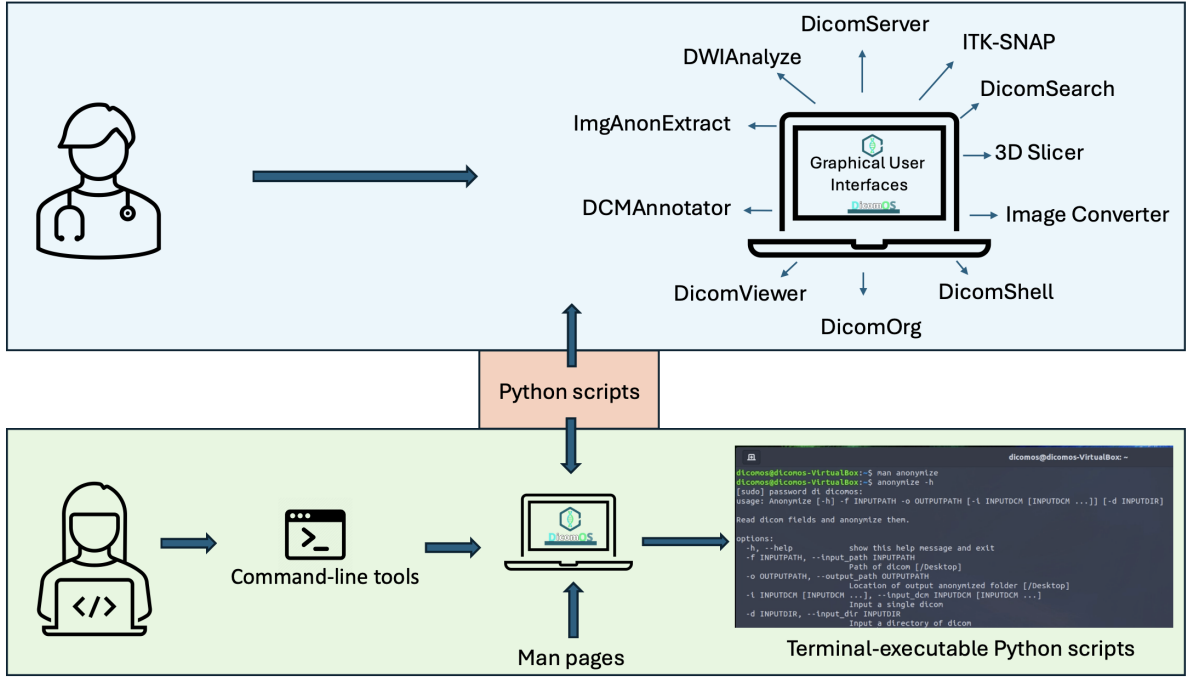


Figure 3: Workflow integration in Dicomos, demonstrating the development of GUI executables for medical professionals and command-line tools for programmers.

textual labels, using either mouse-driven GUI buttons or keyboard shortcuts (**Ctrl+L**, **Ctrl+R**, **Ctrl+T**), which are implemented using Matplotlib’s event listeners. Each annotation is stored as a structured dictionary containing spatial coordinates, object type, and associated text. Importantly, annotations are saved separately from the source image in a .txt metadata file, ensuring that original image integrity is preserved and annotations can be reloaded in future sessions.

Figure 5 provides a screenshot of the active DCMAnnotator session. The interface displays a mammographic image with red-highlighted rectangular annotations and a toolbar of functional buttons located at the bottom. These controls allow the user to switch between annotation modes, load a new image file, and export annotation data.

From a software engineering perspective, the interface adopts an object-oriented design in which each drawing type is implemented as a subclass (“LineDrawer”, “RectDrawer”, “TextDrawer”) of a common “DrawerObject” parent class. These classes manage mouse events, real-time updates through canvas redrawing, and metadata persistence. For instance, rectangle drawing is achieved by tracking four lines with live updates using “canvas.blit()” for performance optimization. Text insertion temporarily disables Matplotlib’s default keymap to allow seamless typing and cursor management within the annotation area. The separation of annotation logic from the core DICOM visualization functionality enables modularity and extensibility. Future developments could include support for structured annotation ontologies (e.g., RadLex), semi-automatic segmentation tools, or direct integration with PACS systems for clinical deployment.

Overall, the DCMAnnotator GUI serves as a practical demonstration of how graphical interfaces embedded within a medical operating system like Dicomos can streamline domain-specific workflows, supporting both clinical and research activities in radiology.

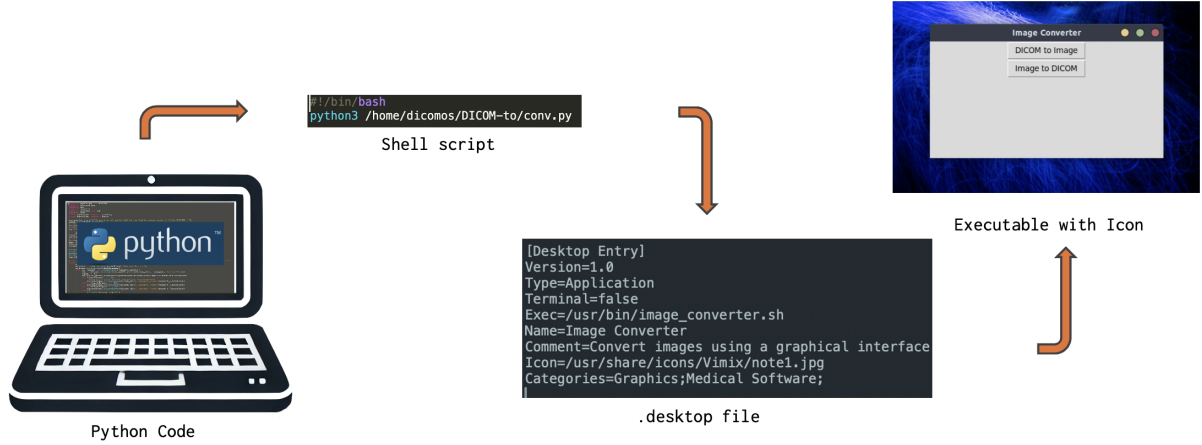


Figure 4: Example process for creating GUI executables in DicomOS, showing Python code execution through a shell script and desktop entry to facilitate easy user access.

3.3.3 Image Converter GUI

The **Image Converter** GUI provides an intuitive and efficient interface for bidirectional conversion between standard image formats (such as PNG and JPEG) and the DICOM format, which is the standard for medical imaging data. The interface (Figure 6) is minimal by design, with two primary buttons allowing the user to select either *DICOM to Image* or *Image to DICOM* operations.

The DICOMtoimage conversion process is based on the `pydicom` and `PIL` libraries. Once a folder or single DICOM file is selected, the software parses its contents, extracts the pixel array, and applies normalization through the `normalize_image()` function. This ensures that pixel intensity values are scaled to the $[0, 255]$ range, suitable for standard image formats. The normalized data is then encoded into a PNG image and saved to the desired output directory. Upon successful conversion, a confirmation dialogue is shown to the user (Figure 6b), providing transparency and usability feedback. In the reverse process, users can select an input image (JPEG or PNG) to convert into a DICOM-compliant object. After image selection, the software converts the image to a grayscale NumPy array, which is then packed into a `pydicom.Dataset` object. A secondary dialogue (Figure 6c) is prompted, requesting essential metadata such as `PatientName`, `StudyDescription`, and `SeriesDescription`. These fields are critical to ensuring compatibility with clinical standards and PACS systems. Once the metadata is entered, the image is saved as a valid DICOM file using implicit VR and little-endian encoding settings. The GUI was designed to be usable by both technical and clinical users, requiring minimal configuration while still offering control over important data attributes. In the backend, robust error handling routines ensure that invalid files are properly detected, and informative messages are shown to the user to guide corrective actions.

This tool is particularly useful in clinical research workflows where interoperability between image sources and DICOM-based systems is required. It also allows radiologists to convert archived DICOM exams into more portable formats for presentations or discussions, and conversely, to re-encode annotated or processed images back into DICOM format while preserving the semantic context through manual metadata entry.

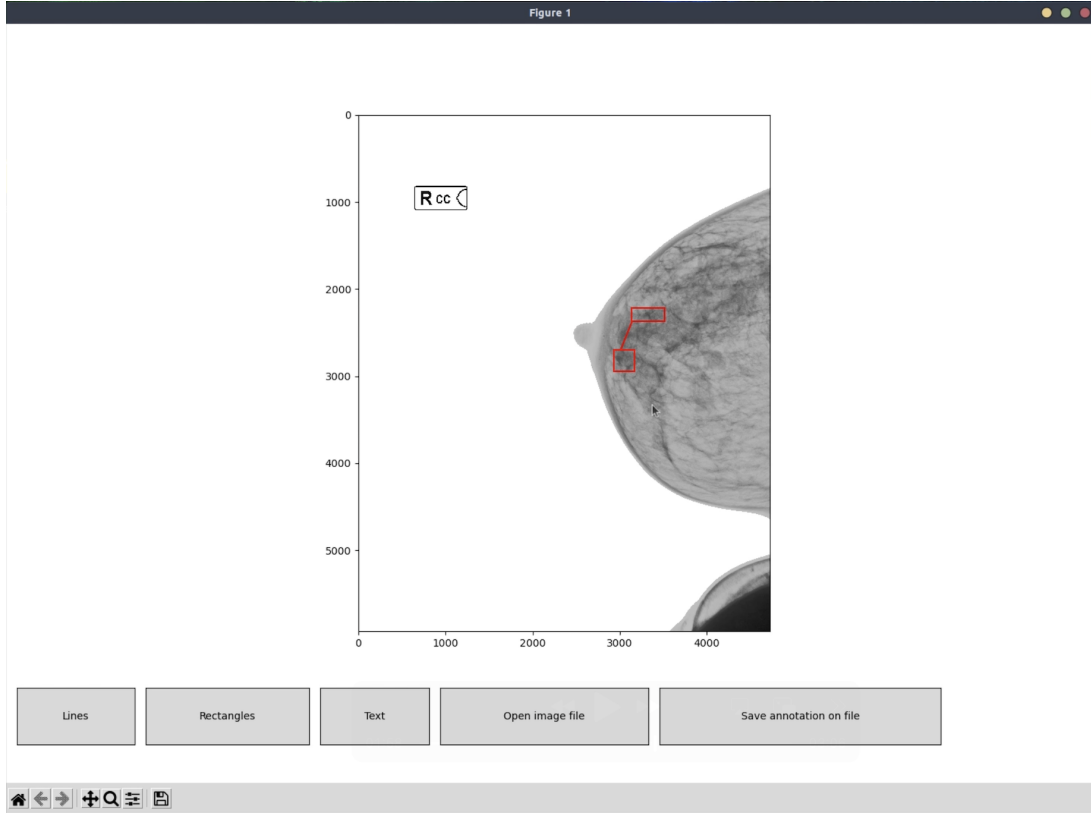


Figure 5: Screenshot of the DCMAnnotator GUI during an annotation session. The annotated DICOM image is shown alongside tools for drawing lines, and rectangles, inserting text, loading image files, and saving annotations.

3.3.4 DICOM Anonymizer GUI

To ensure compliance with privacy regulations and protect sensitive patient data, the DicomOS system integrates a dedicated graphical interface named **ImgAnonExtract**. This tool enables intuitive and highly customizable anonymization of DICOM files and metadata extraction, without requiring any programming knowledge from the user. The interface has been designed with modular sections, reflecting a clear workflow, from file selection to anonymization options, metadata extraction, and result visualization (Figure 7).

The anonymization module allows users to process individual files or entire directories. Through an interactive GUI, users can specify which metadata fields to redact, including Patient ID, Patient Name, Study Date, Birth Date, Institution Name, and Referring Physician Name. These fields are managed via a series of checkboxes, each bound to a specific anonymization flag. When one or more fields are selected, the tool traverses each file and replaces the original values with empty strings, ensuring the data is effectively removed from the DICOM header. If a selected tag is missing from a file, a corresponding warning is issued in the log window, enhancing traceability. The tool supports two anonymization strategies: either overwriting the original file or exporting anonymized versions to a dedicated folder selected by the user. In the latter case, a suffix such as `_anonymised.dcm` is automatically appended to the new filenames. In both modes, the tool generates a detailed log file (`anonymization_log-<timestamp>.txt`) recording all redacted fields and paths of saved files, thereby maintaining audibility and transparency.

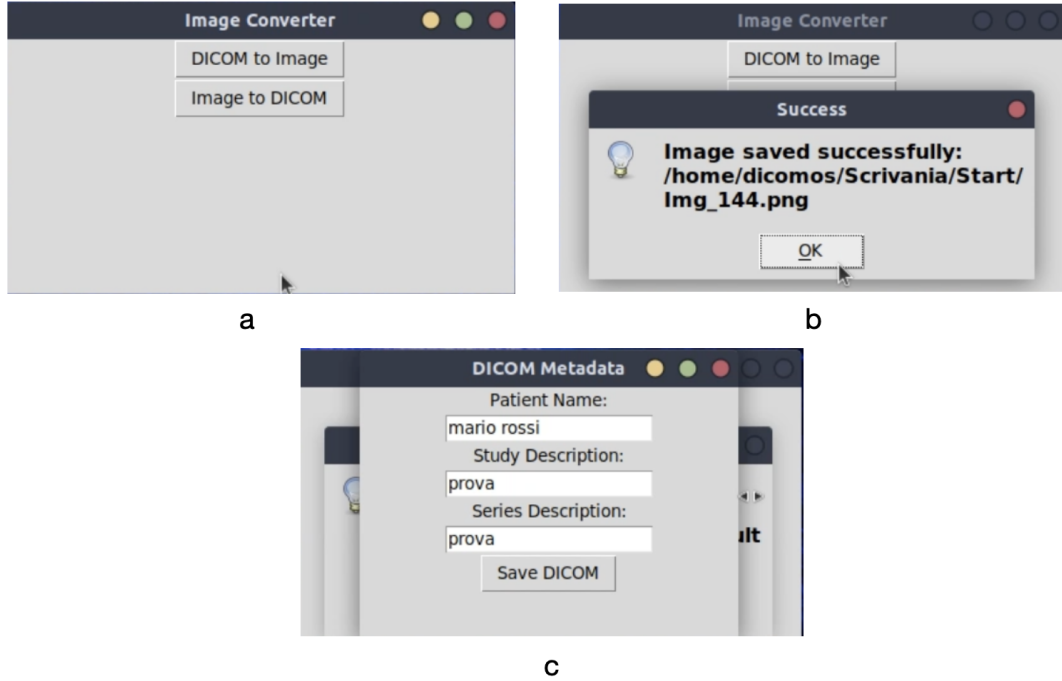


Figure 6: The Image Converter GUI. (a) The main interface for selecting conversion direction. (b) Confirmation dialogue after successful DICOM-to-image conversion. (c) Metadata entry form for converting a standard image into a DICOM file, including fields such as Patient Name, Study Description, and Series Description.

A critical complementary feature is the built-in metadata extraction utility. Users can extract metadata either from individual files or from entire directories. The extracted tags are parsed and stored in a structured list, which is subsequently converted into a tabular format using the `pandas` library. The GUI displays a scrollable log of successful operations and allows the exporting of the metadata into a CSV file. A filter panel lets the user select only specific fields (e.g., Patient ID, Study Date, Institution) to include in the output, reducing redundancy and improving interpretability. Each field value is cleaned and truncated when overly verbose, to optimize visualization and facilitate downstream analyses. A dedicated backup module is also available, enabling the safe preservation of original data before anonymization. Selected files can be duplicated into a user-defined directory through a single button, reducing the risk of data loss due to misconfiguration or human error. Technically, the application leverages the `pydicom` library for parsing and manipulating DICOM headers and relies on `tkinter` for GUI design and interaction. Metadata is dynamically converted into human-readable strings and visualized in real time. Error handling is implemented at each stage to catch I/O failures, corrupted headers, or missing fields. The GUI is also designed to support high-DPI screens and extensive tag tables, providing scrollable interfaces and responsive layout management.

This anonymization and extraction framework is compliant with the FAIR and General Data Protection Regulation (GDPR) principles [37], and is particularly suited for clinical workflows involving secondary use of imaging data in research. By combining privacy, usability, and interoperability, `ImgAnonExtract` enables clinicians and researchers to safely handle medical images while ensuring regulatory compliance.

3.3.5 DICOM File Search GUI

The **DicomSearch** GUI was designed to enhance the navigability of large DICOM archives by providing a flexible, tag-based search mechanism. Built using `tkinter`, `pydicom`, and `matplotlib`, the application allows users to locate specific images based on key DICOM metadata fields such as **PatientID**, **StudyDate**, and **Modality**. This tool is particularly useful in clinical settings and research workflows where the rapid retrieval of specific exams or series is critical. Once a directory is selected, users can initiate a preview operation to explore the contents of the dataset. This preview, powered by the function `list_dicom_files()`, scans the DICOM headers without loading pixel data (via the `stop_before_pixels=True` parameter) to ensure efficiency even in large datasets. The preview displays key metadata for each file, aiding in pre-search assessments.

The core search functionality is handled by the `search_dicom_files()` function, which recursively traverses the selected folder and filters files where the specified metadata field contains the provided search term. Results are dynamically listed in a scrollable text widget, and for each match, an interactive button is generated using `log_text.window_create()`, allowing users to click and display the full-resolution image via `matplotlib`. As shown in Figure 8, the interface layout is divided into logical zones: directory input, search criteria selection, term entry, and an output console where results and previews are listed. This structure was designed for clarity and usability, enabling fast interaction with datasets and facilitating diagnosis, annotation, or AI-based preprocessing.

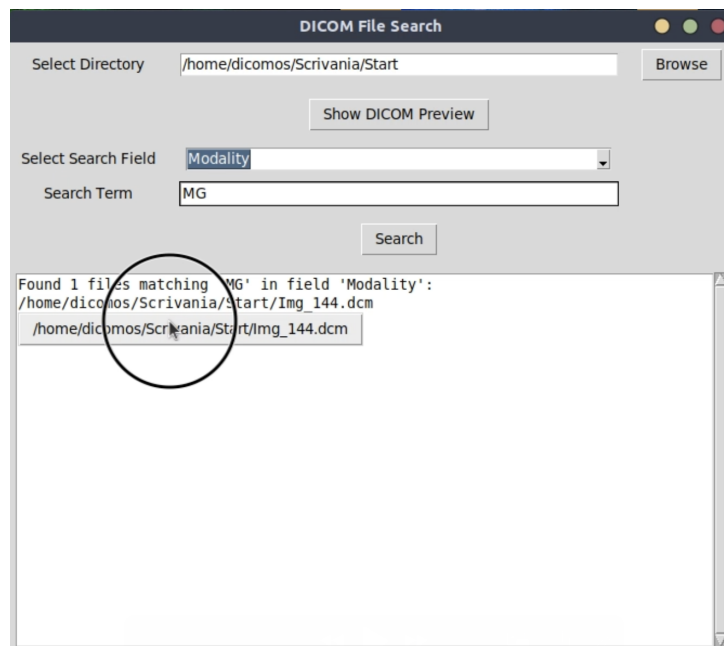


Figure 8: Screenshot of the DICOM File Search GUI. The tool enables users to search within a directory for DICOM files matching a given value in a selected metadata field (e.g., `Modality = "MG"`). Matching results are shown with clickable buttons that launch inline previews of the corresponding image.

This tool also introduces a low-overhead preview mechanism, ensuring usability without requiring full integration with complex PACS viewers. Its simplicity, combined with high utility, makes it a valuable asset within the DicomOS suite, contributing to faster dataset exploration and selection, particularly in the early stages of ML workflows or manual review sessions.

3.3.6 DICOM File Organizer GUI

The **DicomOrg** GUI is a lightweight and practical application developed to simplify the organization of DICOM datasets. Built using Python’s `tkinter` library and leveraging the `pydicom` module, this tool enables users to automatically sort large volumes of DICOM files into structured subdirectories based on key metadata attributes such as *PatientID*, *StudyDate*, or *Modality*. This reorganization enhances both data accessibility and consistency, which are critical when managing heterogeneous imaging repositories across clinical or research environments. Upon launching the interface, the user can select a working directory via the "Browse" button, followed by choosing the metadata criterion from a drop-down menu. Once initiated, the program traverses the entire folder hierarchy, parses DICOM headers without loading image pixel data (to optimize performance), and creates subfolders corresponding to the selected attribute. Each DICOM file is then relocated into its respective subdirectory. This process is fully automated and logged in real time via a scrollable text console at the bottom of the GUI, which provides immediate feedback and error handling (e.g., in cases of unreadable or corrupted files).

Internally, the file management is handled by the `os.walk()` function to recursively locate DICOM files and by `pydicom.dcmread()` for metadata extraction. Each file is moved using `os.rename()`, while the tool remains compliant with data integrity principles by not modifying file contents. As shown in Figure 9, the interface is designed to be intuitive and accessible. The top fields allow directory selection and sorting criterion input, while the central panel displays real-time logs of the reorganization process. This structure offers full control and transparency to the user during execution.

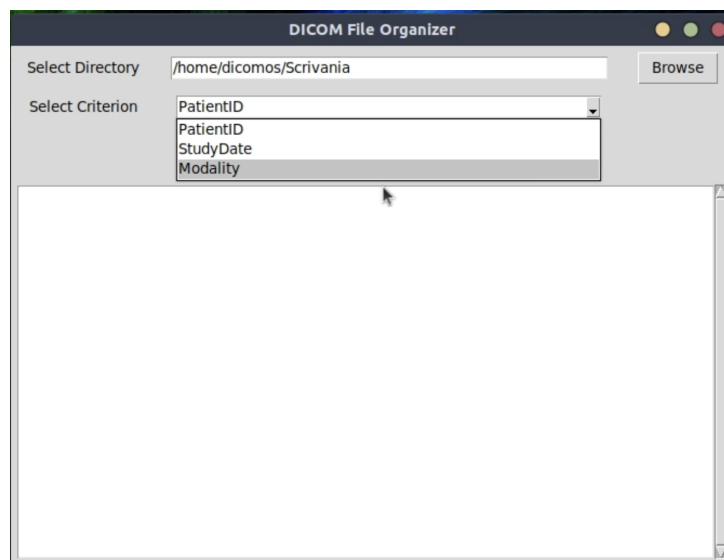


Figure 9: Screenshot of the **DICOM File Organizer** GUI. The user selects a directory and a sorting criterion (e.g., *PatientID*, *StudyDate*, or *Modality*). The tool then organizes all DICOM files into subfolders based on the chosen metadata, streamlining dataset structuring for clinical or research purposes.

This application plays a fundamental role in the DicomOS ecosystem by providing radiologists and researchers with a rapid, customizable solution for dataset preparation, particularly beneficial when integrating DICOM images into annotation workflows or ML pipelines.

3.3.7 DICOM Command Shell GUI

To meet the requirements of users who prefer graphical environments while still needing the power and flexibility of command-line operations, the **DICOM Command Shell (DicomShell)** offers a hybrid interface that bridges both paradigms. Through a single GUI window, users can execute a broad range of commands in DICOM files and directories, using intuitive selection mechanisms and automated logs for complete traceability (Figure 10). Available operations include file listing, metadata inspection, statistical feature extraction, histogram generation, image comparison, and manual annotation. These functionalities are accessible via a drop-down command selector and a single execution button. The logging panel provides real-time feedback, recording outcomes and potential errors. A built-in help system guides users through available operations without requiring knowledge of command-line syntax.

The **list** and **view** commands support efficient navigation through large collections of DICOM files, with the latter displaying extended patient and image metadata, including acquisition parameters and pixel dimensions. For exploratory data analysis, the **analyze** function enables the generation of histograms based on key metadata attributes such as *PatientID*, *Modality*, or *StudyDate*, offering insights into dataset composition. Image-level descriptors can be extracted through the **extract** and **extract.all** commands. These routines compute intensity statistics (mean, standard deviation, min/max), higher-order features (skewness, kurtosis), and entropy-based on localized pixel value distributions. Such metrics are essential for downstream ML pipelines or radiomic profiling and can be automatically saved into structured .txt reports. The **compare** command calculates the Structural Similarity Index (SSIM) [38] between two DICOM images and displays a difference heatmap, which is particularly useful in longitudinal or pre/post-processing evaluations. Manual annotations can be appended using the **annotate** function, allowing users to document image-specific information or clinical observations in an external companion file.

This GUI was designed with modularity in mind, enabling future expansions with minimal effort. Each operation is implemented as a discrete callable module, facilitating integration into broader imaging pipelines or educational environments.

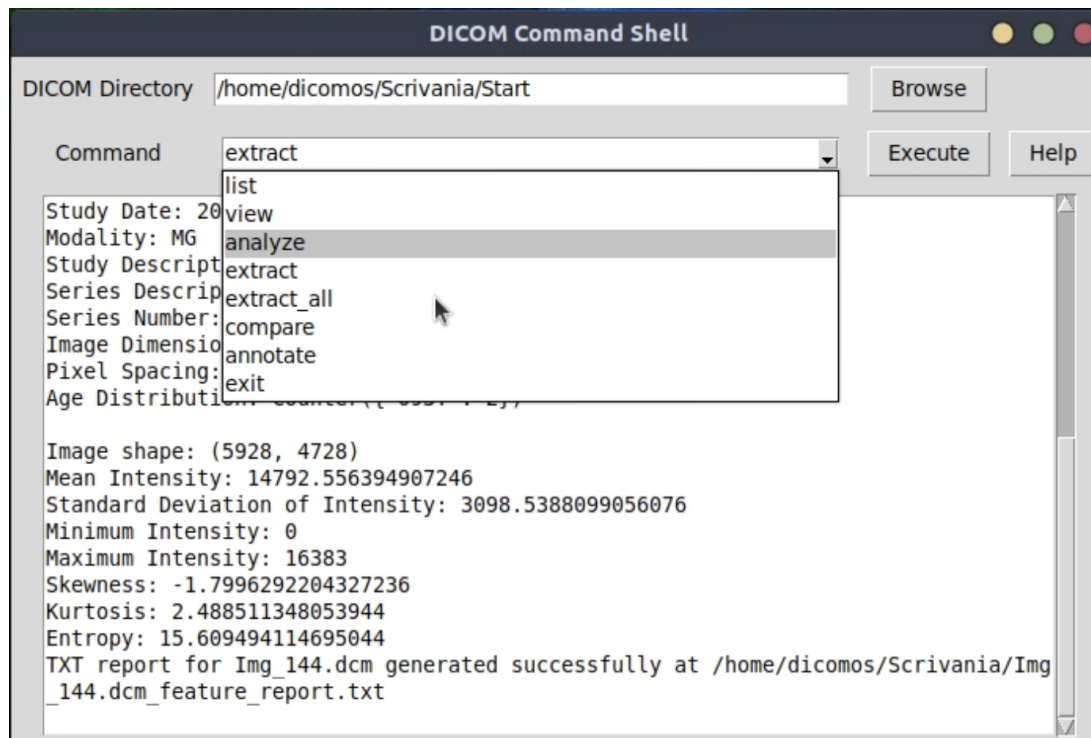


Figure 10: The DICOM Command Shell GUI. Commands such as feature extraction, histogram generation, metadata inspection, and SSIM-based image comparison are available via a dropdown menu. In the displayed example, advanced statistics including skewness, kurtosis, and entropy are computed from a selected mammogram.

3.3.8 DICOM Server Navigator GUI

The **DicomServer Navigator** application was developed to facilitate seamless interaction with remote repositories of medical imaging data. In many clinical or research settings, images are hosted on HTTP-accessible servers and must be retrieved on demand. This GUI component provides users with a complete client-side interface to browse, visualize, and download DICOM content from such remote endpoints. Built with **Tkinter**, **requests**, and **pydicom**, the tool allows users to input a server URL and dynamically list all available files and directories.

Using the function `list_contents_on_server()`, it performs recursive directory crawling, parsing the HTML content of server indexes and correctly handling special characters through URL encoding/decoding mechanisms. The dropdown menu is updated in real-time with both DICOM files and nested folders, allowing the user to navigate between directories or preview images from the current selection. To preview an image, the `fetch_and_display_dicom()` function is called. This routine fetches the DICOM object from the remote server, decodes it, and extracts the pixel array for rendering. If the image contains appropriate DICOM windowing metadata (such as `WindowCenter` and `WindowWidth`), the pixel intensity is normalized accordingly using the `apply_windowing()` function. This results in a grayscale image optimized for diagnostic interpretation, which is then resized and displayed within the GUI. In cases where these parameters are missing, the system estimates them based on pixel statistics. The interface also includes a robust downloading mechanism that handles both individual files and entire directory structures. If a user selects a folder, the system recursively traverses and downloads its contents, preserving the original structure locally. Two helper functions, `download_single_file()` and

`download_entire_folder()`, manage the retrieval process, ensuring resiliency against incomplete transfers and providing real-time progress feedback in the log area. Figure 11 shows an example of the interface in action. In this scenario, a local server instance is queried, a file is selected and visualized, and relevant metadata (e.g., `PatientID`) is logged in the output panel. This tool significantly streamlines workflows involving distributed data sources, enabling clinicians and researchers to remotely access and manage large imaging datasets without leaving the DicomOS environment.

In summary, this GUI provides a practical solution for integrating remote imaging resources into local workflows, supporting flexible access, real-time previewing, and bulk acquisition of DICOM files with minimal technical overhead.

3.3.9 DWI Longitudinal Analysis GUI

To facilitate longitudinal monitoring of neuroanatomical changes in patients affected by traumatic brain injury (TBI), the **DWIAalyze** GUI was developed. This tool provides an end-to-end pipeline for comparing diffusion-weighted imaging (DWI) scans acquired at different clinical stages, typically during the acute phase (immediate post-trauma) and the chronic phase (months later). Its primary function is to allow non-technical users, including clinicians and radiologists, to perform voxel-level comparisons and generate quantitative reports without requiring direct programming intervention. The workflow begins with the manual selection of two input directories: one containing DWI images from the acute phase and the other from the chronic phase [39]. The software adheres to a standardized folder structure, where each patient has nested folders for each time point (e.g., `ses-early` and `ses-late`) and a subfolder named `dwi`. The system recursively searches these paths and attempts to match image pairs based on consistent patient identifiers. If the naming or structure deviates, the GUI provides real-time logging to indicate unmatched or incompatible entries (see Figure 12). Once a valid image pair is identified, the pipeline performs a cleanup step to remove any previously generated intermediate files such as difference maps or registered images. This ensures a clean slate for reproducibility and avoids contamination from prior runs. The core processing workflow involves several stages. First, both acute and chronic images are preprocessed to reduce noise and enhance comparability. This includes Gaussian smoothing, intensity normalization (rescaling values to an 8-bit grayscale format), and optional averaging across temporal volumes if 4D data is provided. These operations are performed using a combination of `nibabel` [40] and `OpenCV` [41], ensuring that both neuroimaging-specific and generic image processing libraries are leveraged. Following preprocessing, rigid image registration is applied using `SimpleITK` [42]. The chronic image is geometrically aligned to the acute image using a centred Euler 3D transform, optimized with a gradient descent algorithm on the mean squared error metric. If resolution or dimensional discrepancies exist between the two images, the pipeline performs automatic resampling to match the spatial characteristics of the reference volume. This step is critical to ensure that subsequent voxel-wise operations are meaningful. Once spatial alignment is achieved, a voxel-wise absolute difference map is computed. This highlights regions of anatomical change, which may correspond to tissue damage, recovery, or evolution of lesions. To quantify these differences, the system calculates statistical descriptors including the mean, standard deviation, and maximum intensity of the difference image. These values are essential for identifying subtle versus extensive changes across time. In addition to raw difference metrics, the pipeline integrates a perceptual similarity measure: the Structural Similarity Index Measure (SSIM). This index evaluates not only pixel intensity differences but also contrast, texture, and

structural patterns, yielding a value between 0 (completely different) and 1 (identical). The SSIM calculation is performed on the entire 3D volume, normalized for intensity, and reported with precision up to two decimal places.

To aid clinical interpretation, a comprehensive PDF report is automatically generated for each patient. This document includes all computed metrics, explanatory text interpreting the significance of values (e.g., high mean difference suggesting significant anatomical change), and graphical visualizations such as the central axial slice of the difference map rendered using a heatmap colour scale. Depending on the magnitude of the changes, the system dynamically adjusts the narrative content of the report to guide the clinician’s interpretation. For example, if the SSIM score is low (≤ 0.5), the report highlights the potential for significant anatomical alterations and recommends clinical follow-up. The GUI also features an interactive progress bar and a scrollable log window, which continuously updates with each processing step, from image matching to report generation. This ensures transparency and traceability throughout the execution of the pipeline.

The modularity and automation of DWIAnalyze make it a powerful tool for longitudinal neuroimaging studies. Its clinical applicability is particularly relevant for monitoring TBI progression, evaluating therapeutic interventions, and detecting patterns of recovery. The interface is also designed to support batch processing, handling multiple patients in a single session, and producing individualized reports for each. Figure 12 displays the interface during execution, showcasing the selected input directories, the processing feedback, and the summary of matched patients. In subsequent chapters, this tool will be applied to longitudinally acquired DWI datasets to investigate structural correlates of consciousness recovery, and its outputs will support the quantitative results discussed in Chapters 9.

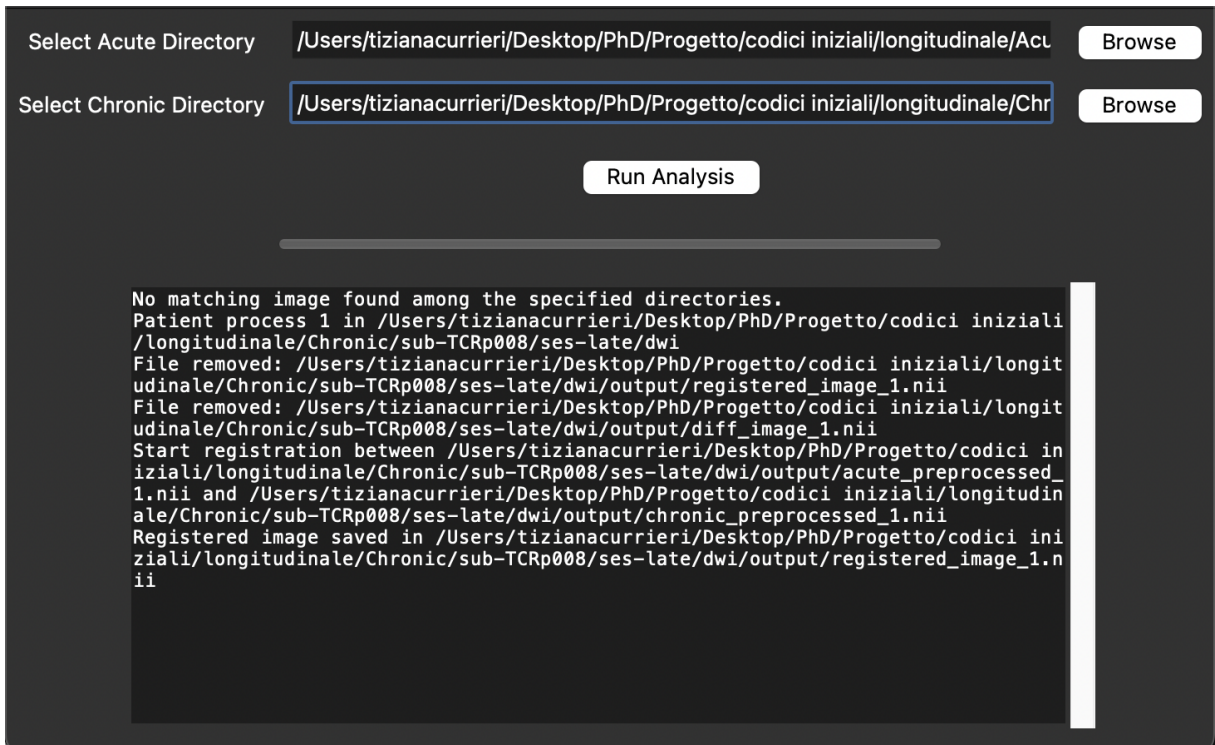


Figure 12: DWIAnalyze GUI during execution. The user selects acute and chronic folders, and the software performs preprocessing, registration, difference computation, SSIM evaluation, and PDF report generation with real-time log output.

3.3.10 Medical Image Editor GUI

The **DicomViewer** module provides a robust and versatile graphical interface for the interactive manipulation and preprocessing of medical and photographic images. Designed primarily for clinical and research contexts, this tool is fully compatible with a wide range of image formats, including DICOM, JPEG, PNG, and BMP, and is capable of handling both colour and grayscale images with equal efficiency. Its architecture is rooted in the integration of OpenCV for image processing and Tkinter for GUI development, offering high responsiveness and cross-platform portability.

The interface is structured into multiple intuitive menus, each grouping related functionalities to facilitate usability even for non-technical users. The application supports both single-image editing and batch processing across entire directories, making it suitable for both exploratory analysis and automated preprocessing pipelines (see Figure [13](#)).

Image Enhancement and Filtering. The tool enables fine-grained control over image appearance through adjustments to contrast, brightness, and saturation. Users can apply sharpening filters based on Laplacian and high-boost convolution kernels, smoothing filters such as mean and Gaussian blur, and advanced contrast enhancement techniques including histogram equalization and CLAHE (Contrast Limited Adaptive Histogram Equalization). These functionalities are especially valuable in clinical domains where subtle visual details must be enhanced without distorting the overall anatomy or lesion morphology.

Geometric and Morphological Operations. The application supports geometric transformations, including horizontal and vertical flipping, enabling standardized orientations for visual inspection and model training. Morphological operators such as erosion and dilation are provided to facilitate structural refinements, contour simplification, and artefact removal, operations commonly used in skin lesion analysis and mammographic preprocessing workflows.

Segmentation and Edge Detection. For preliminary segmentation tasks, the tool includes thresholding techniques and two well-established edge detection methods: Sobel and Canny. These are implemented with customizable parameters and allow the extraction of structural outlines or region boundaries, useful in both manual annotation and preprocessing stages for ML pipelines.

Noise Simulation and Augmentation. The module also incorporates tools for noise injection and denoising, which serve two purposes: simulating low-quality acquisition scenarios for robustness testing and performing data augmentation for training models. Users can add Gaussian noise or apply Gaussian blurring to individual or multiple images, supporting variability in training datasets.

Cropping and ROI Selection. The cropping tool allows users to interactively select a region of interest (ROI) using mouse input. This ROI is then displayed in a separate window for confirmation and can be saved as a new image. This feature is essential for tasks involving lesion isolation, artifact exclusion, or focused training on subregions of medical images.

Undo, Restore, and History Tracking. All modifications are tracked through a built-in revision system. Each transformation is recorded in a modification history, allowing users to undo previous operations or restore the original image at any point during the

session. This functionality is particularly important in clinical settings where traceability and non-destructive editing are essential.

Multi-file Support and Batch Processing. In addition to editing single images, the tool supports the loading of entire folders. Users can apply any filter or transformation to all images within a directory and specify an output folder for the modified files. This batch-processing capability dramatically reduces manual effort and ensures consistency across image datasets.

Color Management and Display. Special attention has been given to the correct handling of color channels and image encoding. Grayscale images are converted to RGB where necessary, ensuring compatibility with the full range of transformations. Furthermore, the canvas within the GUI dynamically resizes according to the image dimensions, allowing high-resolution previews without performance degradation.

Role in Experimental Workflows. The `DicomViewer` played a pivotal role in the design and execution of the experimental methodologies described in later chapters of this thesis. All preprocessing steps applied to the dermoscopic and mammographic datasets were developed and tested within this framework. For instance, CLAHE was used to enhance contrast in dermoscopic images prior to lesion classification, while morphological operators and edge detectors were employed to refine mammographic structures. The ability to interactively validate preprocessing steps in real time was critical in identifying optimal parameters for data preparation.

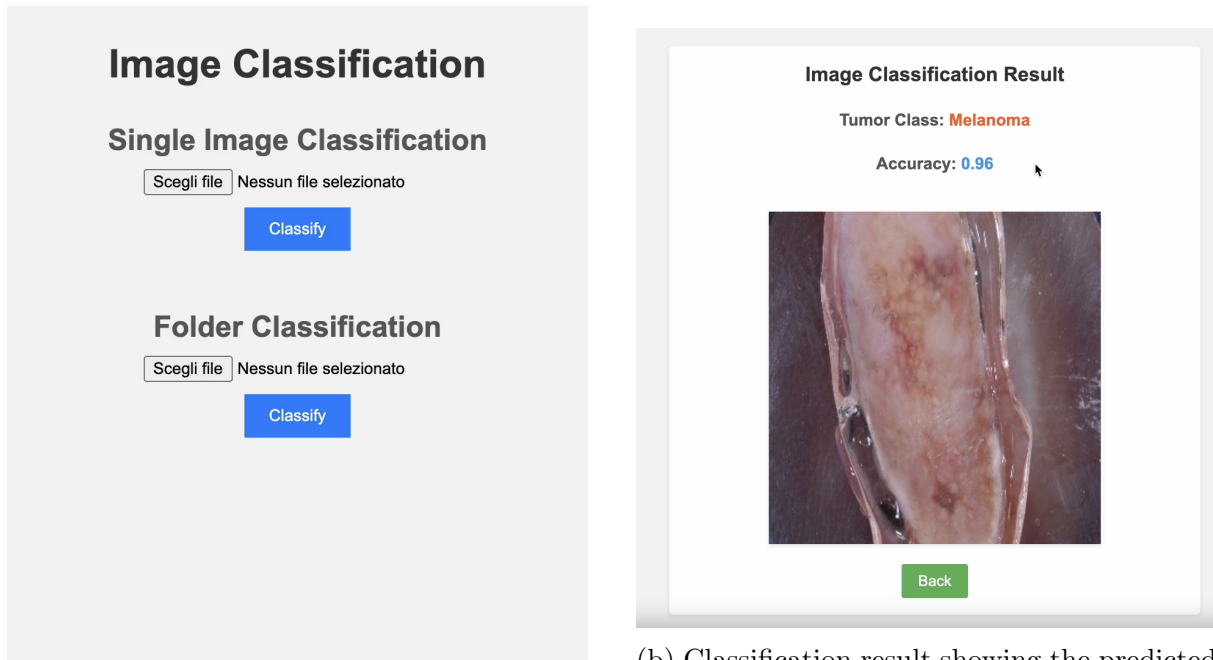
Accessibility and Clinical Integration. The system was specifically designed to be usable by clinicians without programming skills. All functionality is accessible through a visual menu-based interface, avoiding the need for scripting or command-line input. As such, the application is well-suited for deployment in hospital environments or collaborative research settings where multidisciplinary teams require shared tools.

3.3.11 Deployment of a Web-Based Interface for Melanoma Classification

Following the development of several graphical tools within `DicomOS` aimed at facilitating user interaction with medical image data, a standalone web application was also implemented to demonstrate how deep learning models can be integrated into accessible, cross-platform interfaces for clinical support. This application specifically targets the classification of dermatological lesions, with a focus on melanoma detection, and was designed to make the underlying artificial intelligence models usable without requiring any programming knowledge or specialized software installation. The web application was built using Flask, a lightweight web framework for Python, which allows the rapid deployment of interactive and data-driven interfaces. The implemented system provides two distinct functionalities: single-image classification and batch (folder-based) classification. Both options are accessible through a simple user interface built with HTML and CSS, and they are designed to guide the user through image upload, classification, and visualization of results in a few intuitive steps. From a technical perspective, the application loads a Vision Transformer (ViT) model previously fine-tuned for melanoma classification. The model architecture and training procedure are detailed in [8.2](#), and the specific checkpoint used here corresponds to the best-performing weights obtained during training, saved at epoch 31. The model was retrieved and loaded using Hugging Face’s transformers library, ensuring compatibility with the trained ViT variant and allowing full

reproducibility of the results. When a user uploads an image, the server converts the file into a NumPy array using OpenCV, applies a resizing operation to 384×384 pixels (the required input dimension for ViT), and transforms the array into a tensor using PyTorch's `ToTensor()` utility. The RGB-converted tensor is then passed to the model, which returns raw logits. These are processed using a softmax function to obtain class probabilities, and the highest-scoring class is returned as the prediction. The three supported output categories, Melanoma, Nevus and Seborrheic Keratosis, were defined during training and reflect the labels of the ISIC 2017 dataset used for model development.

The resulting prediction, along with the corresponding confidence score, is presented to the user on a new dynamically generated web page. In the case of folder classification, all images are processed in sequence, and the outputs are shown in tabular form with thumbnails, predicted class labels, and associated confidence values. A downloadable text file containing all results is also generated automatically. The application's logic includes error handling, memory management for large batches, and automatic cleanup of uploaded images to maintain storage efficiency. Furthermore, the structure of the project allows for future deployment on platforms such as Heroku or AWS, enabling broader access in clinical or research contexts. Importantly, all visual outputs and classification results are retained in the project's static folder to ensure traceability and reproducibility, in line with FAIR principles. From a user experience perspective, the application was designed to ensure that even non-expert users can interact with complex AI models in a transparent and informative way. The interface includes visual feedback in the form of image previews, clearly highlighted class labels and colour-coded confidence scores. The classification workflow is thus transformed from a black-box process into a guided and understandable diagnostic aid. By leveraging the same ViT model already integrated into the analytical workflows described earlier in the thesis, the web application serves as a bridge between computational research and real-world application, offering a concrete example of how medical AI can be delivered directly to end users. Figures [14a](#) and [14b](#) illustrate two views of the interface: the upload section and the result page showing predicted labels and accuracies. By combining simplicity of use with model transparency, this web-based solution serves as a tangible example of deploying explainable AI in a real-world medical imaging context.



(a) Web interface for selecting a single image or an entire folder for classification.

(b) Classification result showing the predicted tumour class and the associated confidence score.

Figure 14: Screenshots of the web application for melanoma classification developed using Flask and ViT. (a) The interface allows single or batch input; (b) the result view displays the predicted class with a confidence score and image preview.

3.4 Development of Command-Line Tools for Programmers

Understanding the operational needs of more technical users, DicomOS includes a dedicated command-line toolkit named **Dicom-to**. This package includes utilities like **Convert**, **Extract**, **ConvertExtract**, and **Anonymize**, tailored for batch operations and automation.

The **Convert** tool converts DICOM files to standard image formats (JPEG, PNG, BMP) and vice versa. It supports batch processing, user-defined input/output directories, and optional arguments for format selection. A built-in `--help` flag provides inline documentation.

The **Extract** utility retrieves metadata from DICOM files and saves it to CSV, supporting further clinical or research analysis. Metadata fields to extract can be customized via a configuration file located in `~/DICOM_to`.

ConvertExtract combines the functionality of both tools, enabling simultaneous image conversion and metadata export, saving time in integrated workflows.

Anonymize securely removes identifiable metadata from DICOM files. Users can specify input/output locations and customize redacted fields. Like the others, it includes a `--help` option and robust error handling.

These tools are implemented in Python using `pydicom` for DICOM handling and `Pillow` for image manipulation. They are compatible with Unix-like systems and include man pages installed in `/usr/local/share/man/man1`. Each man page outlines usage syntax, descriptions, options, examples, and authorship details, aligning with Linux documentation standards [43].

An example of use (see Figure 15) includes:

```
Convert -f /Desktop -o /Documents -option jpg -i Img.x.dcm
```

This command converts `Img_x.dcm` from the desktop into JPEG format, storing it in the Documents folder under a subfolder `Final`.

Another example:

```
Extract -f /Documents -o /Desktop -d Directory
```

This extracts metadata from DICOM files in `Directory` under Documents and writes the output to `Patient_Detail.csv` on the desktop.

Combining both:

```
ConvertExtract -f /Desktop -o /Documents -option png -d Directory
```

This pipeline handles both conversion and metadata extraction in one step.

The man pages significantly enhance accessibility for developers and system administrators, allowing seamless incorporation of these tools into custom scripts or automated processing pipelines. This documentation ensures clarity, promotes consistency, and supports long-term maintenance.

3.5 Evaluation of the System

The implementation and adaptation of DicomOS led to the creation of a customized Linux-based platform specifically designed for medical imaging professionals. A key achievement was the seamless integration of core radiological applications into a unified, user-oriented environment that supports daily clinical workflows. By visually adapting the Ubuntu desktop through themes and icon sets reminiscent of macOS, DicomOS delivers an interface that eases the transition for users familiar with Apple systems. This visual alignment helps reduce the onboarding time for radiologists switching to a Linux-based system, contributing to a more satisfying and productive user experience [27]. Advanced medical imaging software such as ITK-SNAP and 3D Slicer was fully embedded within the system through the development of execution scripts and GUI launchers. These enhancements allow users to launch the applications with a simple double-click, eliminating the need for terminal commands and facilitating smoother operation. This streamlined access improves productivity by enabling the immediate use of critical segmentation and visualisation tools in radiological analysis. Additionally, a tailored suite of graphical tools was created to address specific imaging-related tasks. Tools like DCMAnnotator, Image Converter, and ImgAnonExtract provided intuitive solutions for image annotation, format conversion, and metadata anonymization. These tools, described in Table 2, were built to follow established principles of usability, integrating clear visual elements, interactive menus, and keyboard shortcuts. As a result, users can execute common imaging operations with greater speed and fewer errors. For users with technical backgrounds, DicomOS also includes a robust set of command-line utilities, such as **Convert**, **Extract**, and **Anonymize** which support automation and bulk processing of datasets. These tools are fully documented with Unix-style man pages to ensure consistent usage across different systems and user levels. Their design allows developers and system administrators to incorporate them into scripts and pipelines, enabling flexible integration with external imaging systems. The synergy between GUI-based applications and command-line utilities equips DicomOS to cater to a wide range of user profiles. Clinical personnel benefit from intuitive interfaces, while technical users can leverage scripting capabilities for advanced customisation. This hybrid approach ensures that DicomOS functions as a practical, extensible, and user-friendly platform for medical image management. Ultimately,

the creation of a specialized operating system bridges the gap between the complexity of medical imaging software and the usability expectations of healthcare professionals. DicomOS enhances the efficiency and quality of radiological workflows by embedding critical tools and ensuring ease of access.

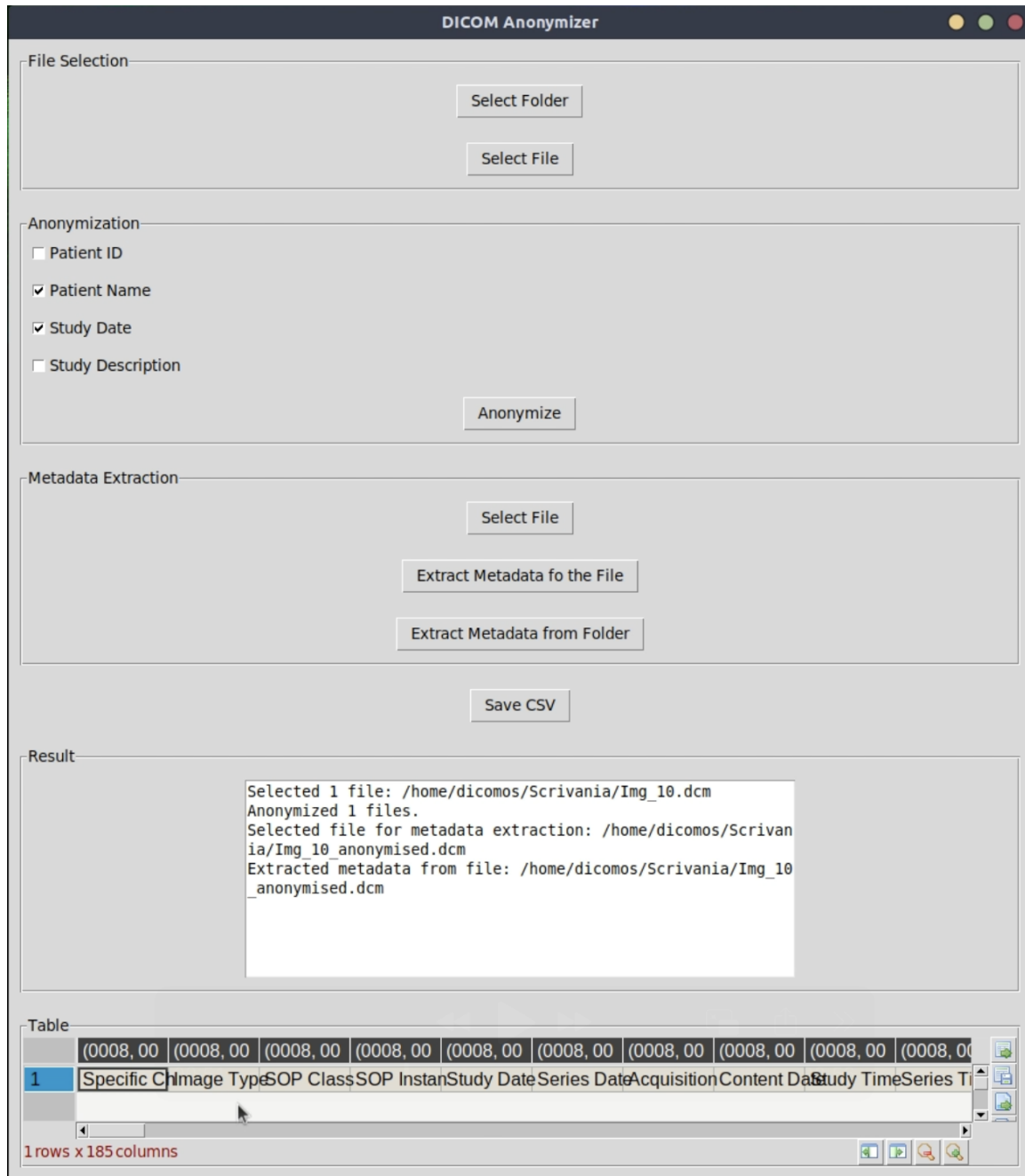


Figure 7: Graphical interface of **ImgAnonExtract**, showing anonymization options, file selection, metadata extraction, and real-time logs.

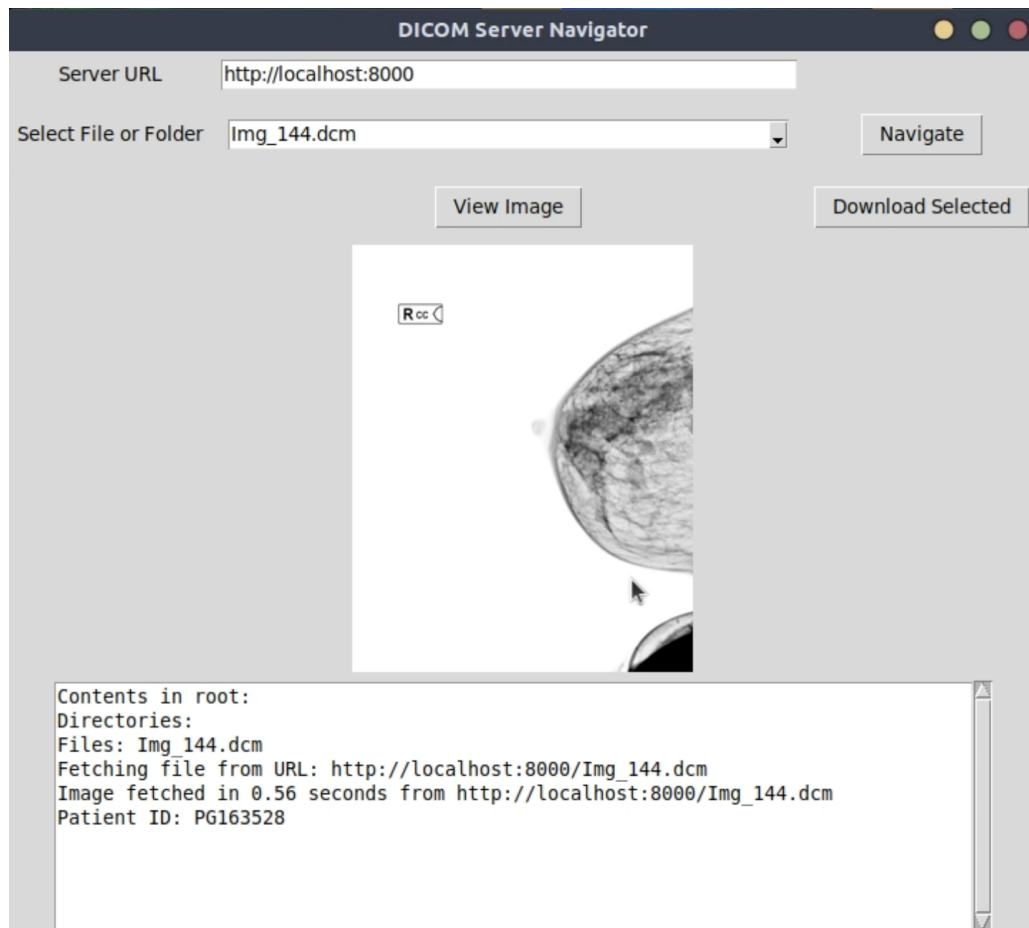


Figure 11: Example usage of the DICOM Server Navigator GUI. A DICOM image is retrieved from a local HTTP server, visualized within the GUI using windowing techniques, and associated metadata such as **PatientID** is logged. The tool supports directory navigation, on-demand visualization, and recursive downloads.

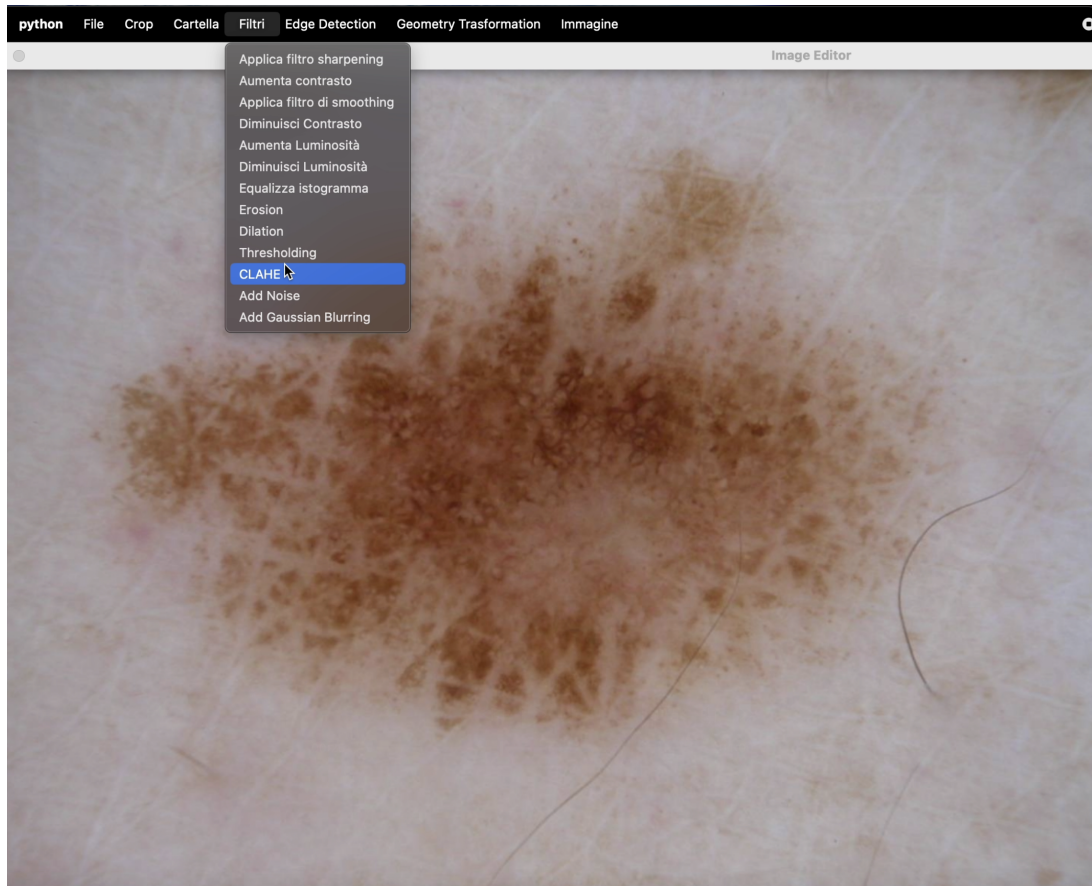


Figure 13: The DicomViewer GUI with filter operations. The CLAHE contrast enhancement technique is being selected to improve local contrast in a dermoscopic image. The toolbar also includes options for histogram equalization, thresholding, morphological operations, edge detection, and more.

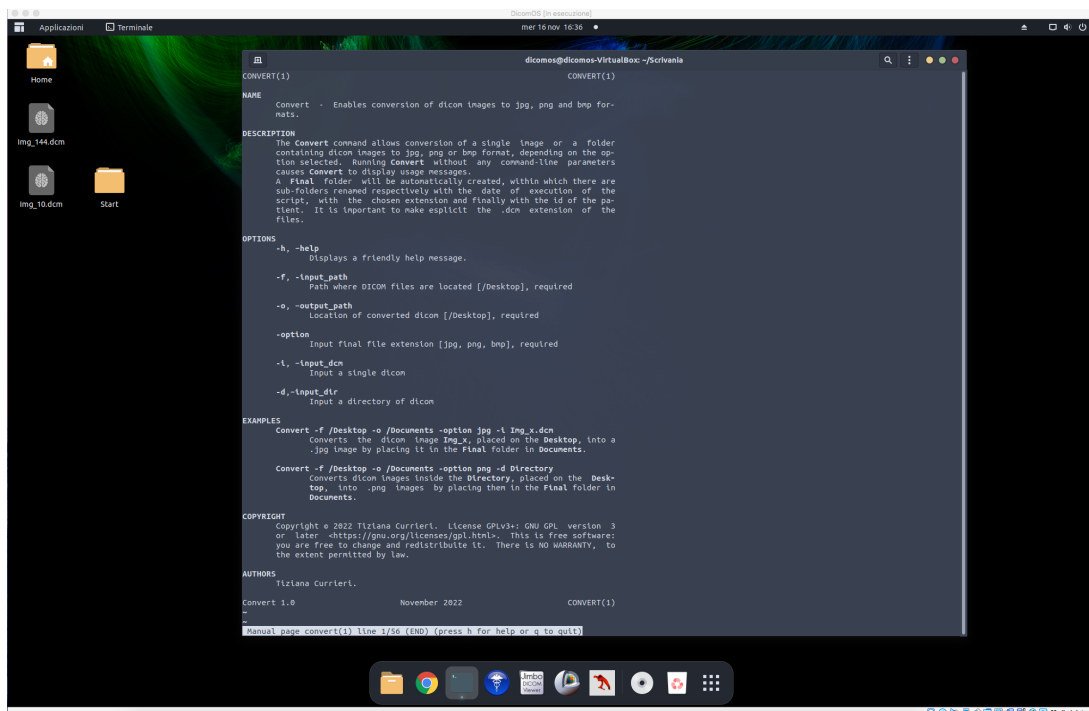


Figure 15: Manual page of the custom **Convert** command developed within DicomOS. This CLI tool enables batch or single DICOM image conversion to standard formats such as JPEG, PNG, and BMP. The manual describes command-line options, including input and output paths, format selection, and usage examples, thereby enhancing usability and reproducibility in medical image processing workflows.

Table 2: Summary of GUIs Integrated into DicomOS for Medical Imaging Tasks

GUI Application	Function
ITK-SNAP	Advanced medical image segmentation and visualization tool, often used for annotating structures in 3D medical images.
3D Slicer	Provides powerful tools for visualization and analysis, including segmentation, registration, and quantitative imaging.
DCMAnnotator	Allows users to annotate medical images with lines, rectangles, and text, saving annotations separately to preserve original image data.
Image Converter	Converts DICOM files to JPEG, PNG, and other formats, and vice versa, supporting metadata input for DICOM conversions.
ImgAnonExtract	Anonymizes selected metadata fields in DICOM files, with options to overwrite or save anonymized copies and create backups.
DicomSearch	Enables the search and preview of DICOM files based on metadata fields, facilitating quick location of relevant images.
DicomOrg	Organizes DICOM files into subfolders by criteria such as PatientID or StudyDate, streamlining data management.
DicomShell	Provides a GUI for command-line operations like listing, viewing, analyzing, extracting features, and comparing DICOM files.
DicomServer	Allows browsing and downloading files from a remote DICOM server, including preview and secure transfer of selected files.
DWIAalyze	Compares acute and chronic DWI scans for a patient, performing image registration, difference computation, and report generation.
DicomViewer	Offers a range of image processing functions (e.g., contrast adjustment, filtering, edge detection) for various medical image formats.

4 Artificial Intelligence in Radiology: Background and Framework

AI has rapidly evolved into a pivotal component of modern radiological practice, offering the capability to augment human expertise in interpreting complex medical images. AI systems are computational frameworks that emulate cognitive processes such as perception, learning, and reasoning, thereby enabling data-driven analysis and decision-making. Within radiology, AI contributes to improved diagnostic precision, reduced inter-reader variability, and enhanced workflow efficiency. As the scope of AI expands, two primary methodological paradigms have emerged: shallow learning and deep learning. These paradigms differ fundamentally in terms of model architecture, feature processing, data requirements, and interpretability. Their application in radiology is not merely a matter of technical implementation but also intersects with broader concerns such as clinical trust, transparency, and regulatory compliance. A concurrent and critical issue in this context is explainability. Explainable AI (XAI) aims to render the predictions and internal mechanics of AI systems interpretable by clinicians, fostering trust and accountability. In high-stakes environments like radiology, where medical decisions have direct consequences for patient care, explainability is not only desirable but essential.

4.1 Definition of Artificial Intelligence and its Role in Radiology

While much of the current discourse on AI in radiology focuses on image interpretation and diagnostic accuracy, the broader systemic and infrastructural implications of AI integration deserve equal attention. AI in radiology should not be seen exclusively as a tool for automating classification or segmentation tasks; rather, it represents a paradigm shift in how medical imaging data are processed, contextualized, and transformed into clinical value. One emerging dimension is the application of AI to non-visual components of radiological workflows. Natural Language Processing (NLP), for example, is increasingly utilized to extract structured data from narrative radiology reports, facilitating large-scale retrospective analyses and enabling an automatic population of databases for research and audit purposes [44]. This contributes not only to operational efficiency but also to the democratization of access to imaging-derived knowledge, especially in data-driven health systems. Another important yet under-discussed application of AI lies in quality assurance. Algorithms are now being developed to detect suboptimal acquisitions in real-time, flagging motion artefacts, underexposed regions, or protocol deviations as they occur [45]. This proactive monitoring helps reduce unnecessary repeat scans and optimizes the use of imaging resources. In advanced implementations, AI systems can also suggest personalized acquisition parameters based on patient-specific features, aligning radiological imaging with the principles of precision medicine. AI is also increasingly embedded in educational platforms. Simulators powered by generative models can create synthetic cases to train radiology residents on rare pathologies or atypical presentations, offering exposure that might not otherwise be possible in clinical practice [46]. These systems support adaptive learning strategies, adjusting difficulty based on the learner's progress and response patterns. Such tools can play a pivotal role in lifelong learning, helping to bridge the gap between technological evolution and professional development. From a governance standpoint, the integration of AI introduces novel ethical, regulatory, and legal questions. The need for explainability extends beyond model transparency and touches upon institutional responsibility, clinical accountability, and the management of algorithm-

mic uncertainty [47]. Questions around who is liable when an AI-driven recommendation leads to harm remain unresolved in many jurisdictions, emphasizing the importance of interdisciplinary dialogue between technologists, clinicians, and policymakers. Lastly, the incorporation of AI into radiological practice is not uniformly experienced across global contexts. While high-resource settings are leading the implementation of sophisticated AI solutions, many low- and middle-income countries still face barriers related to data availability, computational infrastructure, and regulatory clarity [48]. Bridging this digital divide requires not only technological innovation but also strategic efforts in capacity building, open-access model sharing, and context-aware algorithm development.

Taken together, these perspectives underscore that AI in radiology should not be reduced to its diagnostic capabilities alone. Its role spans operational, educational, ethical, and equity dimensions that must be considered in any comprehensive discussion about its integration into modern healthcare systems.

5 Classifiers in Radiological Machine Learning

5.1 Overview of Classification in Medical Imaging

The application of ML to radiological imaging represents one of the most prominent areas of research in medical artificial intelligence. A wide range of clinical tasks, such as breast cancer screening [49], cardiovascular disease classification [50], and prediction of treatment outcomes [51], has been successfully addressed using ML-based approaches [52, 53]. At its core, ML refers to computational methods capable of learning patterns from data to make predictions or classifications without explicit rule-based programming. A fundamental subdivision within ML distinguishes between shallow learning (SL) and deep learning (DL). SL includes traditional algorithms developed prior to the emergence of deep architectures and encompasses methods such as logistic regression, support vector machines (SVM), decision trees (DT), and k-nearest neighbours (KNN) [54]. These techniques are typically characterized by a single transformation layer between input and output. Conversely, DL encompasses all deep architectures, such as convolutional neural networks (CNN), recurrent neural networks (RNN), and transformers, that involve multiple processing layers and are capable of learning hierarchical features directly from data [13].

In medical image analysis, most classification tasks fall under the category of supervised learning, in which the goal is to predict a categorical label based on input data. These labels may represent clinical conditions, such as the presence or absence of a lesion, or the effectiveness of a therapy. The success of supervised learning models depends on the availability of high-quality annotated data, typically derived from expert assessment or gold-standard diagnostic procedures such as histopathological examinations. To train a classifier, input data must be converted into a numerical representation that captures salient and discriminative features. This step, known as feature extraction, is critical to the success of the model and can be performed either through handcrafted techniques [55, 56], such as radiomic analysis of intensity, shape, or texture, or through automatic feature learning using deep neural networks [57]. The choice of extraction strategy influences the subsequent selection of the classifier, as SL methods require precomputed features, whereas DL methods integrate both feature learning and classification within a single framework. Developing high-performing models involves a series of technical steps that require careful tuning and robust methodology. These include not only feature extraction but also feature selection, hyperparameter optimization, and rigorous validation protocols. Improper handling of any of these stages can significantly undermine the model’s generalization ability and clinical utility [58]. The next section outlines the key phases involved in model development, including training, validation, and testing.

5.2 Model Development: Training, Validation, and Testing

A core component of developing a reliable classifier in medical imaging is the rigorous separation of the available dataset into distinct subsets for training, validation, and testing. These phases are essential for evaluating the model’s ability to generalize beyond the data it was trained on, particularly when applied to unseen clinical cases.

The training set is used to learn a decision function that maps input features to output classes. In the case of neural networks, for instance, this involves adjusting internal weights to minimize a predefined loss function. Once an initial model is obtained, a separate validation set is employed to assess its performance and to tune its hyperparameters,

configurable parameters that influence the learning dynamics, such as the number of hidden layers, learning rate, or kernel functions. Unlike trainable parameters (e.g., weights), hyperparameters are not learned during training and must be selected through systematic experimentation. Hyperparameter tuning is often carried out using grid search, random search, or more advanced automated techniques [59, 60]. For each configuration, the model is evaluated on the validation set, and the one yielding the best performance is selected for final testing. The test set, composed of data unseen during both training and validation, serves to evaluate the generalization capacity of the final model. This step is crucial to ensure that the algorithm performs reliably on real-world clinical data and is not overfitted to the training distribution [61]. For large datasets, typically comprising several thousand samples, a three-way split (e.g., 70%-15%-15% or 70%-10%-20%) is common, although no fixed standard applies. Importantly, this partitioning is often performed in a stratified manner to maintain a balanced representation of all classes within each subset. When dealing with more limited datasets (i.e., several hundred samples), a two-way split (training and testing) is often preferred, with cross-validation strategies applied within the training set to optimize and evaluate the model. Among these, k-fold cross-validation is widely adopted: the data set is divided into k equal-sized folds, and training is iteratively performed on k-1 folds while validating on the remaining fold. In scenarios involving extremely small datasets (fewer than 100 samples), the Leave-One-Out (LOO) cross-validation method is frequently employed [62, 63]. In this strategy, the model is trained on all samples except one, which is used for testing. The process is repeated until every sample has been used as the test instance. While LOO can be informative in data-scarce contexts, it tends to increase the risk of overfitting compared to k-fold cross-validation [64], which is therefore preferred when the dataset size permits [65, 66].

Furthermore, recent guidelines such as CLEAR [67] and CLAIM [68] have clarified the nomenclature of data partitioning, emphasizing that the validation set is used to optimize the model during development, while the test set serves exclusively for final evaluation. Misuse or misunderstanding of these distinctions can undermine the reliability and reproducibility of ML studies in medical imaging.

5.3 Shallow Learning Classifiers

SL models, often referred to as traditional ML approaches, are characterized by their relatively simple architectures and limited representational depth. Unlike DL models, which consist of multiple hierarchical layers for feature abstraction, shallow classifiers typically operate with one or few processing stages. Their effectiveness relies heavily on the quality of the input features, which are often manually engineered using domain knowledge. In radiology, these features are commonly derived from hand-made radiomic descriptors, which include intensity, texture, and shape-based attributes extracted from regions of interest (ROI). The following subsections describe the most frequently used shallow classifiers in radiological applications.

5.3.1 Logistic Regression

Logistic regression (LR), also referred to as logit regression, is one of the most widely adopted linear classification models in clinical practice [69]. It estimates the probability of a sample belonging to a specific class by applying the logistic function, also known as the sigmoid function, to a linear combination of the input features. This function compresses

the output into a continuous range between 0 and 1, enabling binary classification. The classification decision is typically based on a threshold of 0.5, above which the sample is assigned to the positive class.

The training process relies on maximum likelihood estimation (MLE) to minimize the discrepancy between the predicted probabilities and the true labels. One of the major advantages of LR is its interpretability and simplicity, as it does not require complex hyperparameter tuning. These properties make LR particularly suitable for use in medical imaging studies where model transparency is essential [70, 71, 72].

5.3.2 Support Vector Machine

Support Vector Machines (SVMs) are powerful classifiers that aim to identify the optimal hyperplane that separates data points of different classes with the maximum margin [73]. This hyperplane is defined by the so-called support vectors, critical data points that lie closest to the decision boundary. The distance between the support vectors and the hyperplane is known as the margin, and maximizing this margin improves the model's generalizability.

In practice, data are often not linearly separable in their original feature space. To address this, SVMs employ kernel functions that map the data into a higher-dimensional space where a linear separation is feasible. Common kernel types include polynomial, radial basis function (RBF), and sigmoid kernels, each influencing the geometry of the transformed feature space [74]. The model's flexibility is regulated by the regularization parameter C , which balances the trade-off between maximizing the margin and minimizing classification errors. A lower value of C promotes a wider margin with more misclassifications (better generalization), while a higher value of C leads to fewer misclassifications in training at the risk of overfitting. Figure [16, 17] illustrate these concepts.

5.3.3 Tree Ensemble Methods

Tree ensemble (TE) models are a class of ML techniques that aggregate the predictions of multiple decision trees to improve overall performance and robustness [75]. By combining several weak learners, ensemble methods mitigate individual tree variance and enhance generalization. In medical image analysis, random forests (RF) and gradient boosting (GB) are the most prominent ensemble strategies due to their superior performance on small to medium-sized datasets [76, 77].

5.3.3.1 Random Forest Random forests construct multiple decision trees by sampling the training data with replacement, a process known as bootstrapping, and selecting a random subset of features for each split [78]. Each tree operates independently and casts a vote for the predicted class, with the final output determined by majority voting. This method reduces overfitting by promoting tree diversity and has proven to be resilient to noise and missing values. Key hyperparameters include the number of trees in the forest and the number of features considered at each split. A common heuristic sets the latter as the square root of the total number of features [79, 80]. Random forests are particularly valued in clinical contexts for their balance between performance and interpretability. Figure [18] may be used to show a schematic representation.

5.3.3.2 Gradient Boosting Gradient boosting (GB) builds decision trees sequentially, with each new tree correcting the errors made by the ensemble of previously trained

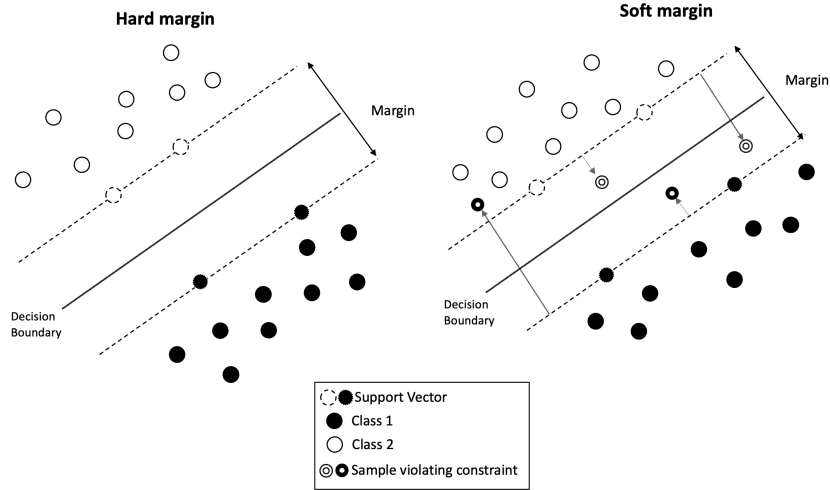


Figure 16: Graphical representation of hard and soft margin of a support vector machine. With the soft margin, some misclassifications (double circles) are allowed

trees [81]. This iterative refinement allows the model to focus more on difficult-to-classify instances, making it particularly effective for handling complex patterns. Unlike random forests, where trees are built in parallel, GB trains each tree based on the residuals of the previous ones. This approach is known as boosting and generally leads to higher predictive accuracy. The process is governed by parameters such as the learning rate, tree depth, and number of estimators. A highly efficient implementation of this method is provided by the XGBoost library, which incorporates additional regularization techniques to further enhance performance and computational speed [82, 83].

5.3.4 K-Nearest Neighbors

The K-Nearest Neighbors (KNN) algorithm is a non-parametric method that classifies new instances based on the majority label among their K closest neighbours in the feature space [84]. The algorithm relies on a distance metric, such as Euclidean, Manhattan, or Hamming distance, to determine proximity between data points [85]. When a new observation is introduced, the algorithm computes distances to all points in the training set and assigns the class most frequent among the nearest K examples. The performance of KNN depends on the selection of three key hyperparameters: the number of neighbours K, the choice of distance metric, and the weighting function applied to each neighbour's vote. While computationally simple, KNN can become inefficient with large datasets or high-dimensional feature spaces. Nonetheless, it remains a popular baseline due to its ease of implementation and intuitive nature. Figure 19 is recommended to illustrate the decision-making process.

5.4 Deep Learning Classifiers

DL represents a specialized subset within the broader domain of ML, distinguished by its architectural depth and capacity to model complex hierarchical data representations [86]. Inspired by the structure of the human brain, deep neural networks (DNNs) are composed of multiple interconnected layers of artificial neurons. This layered configuration enables the progressive extraction of features from data, advancing from simple patterns to highly abstract concepts, commonly referred to as feature hierarchies. DL algorithms

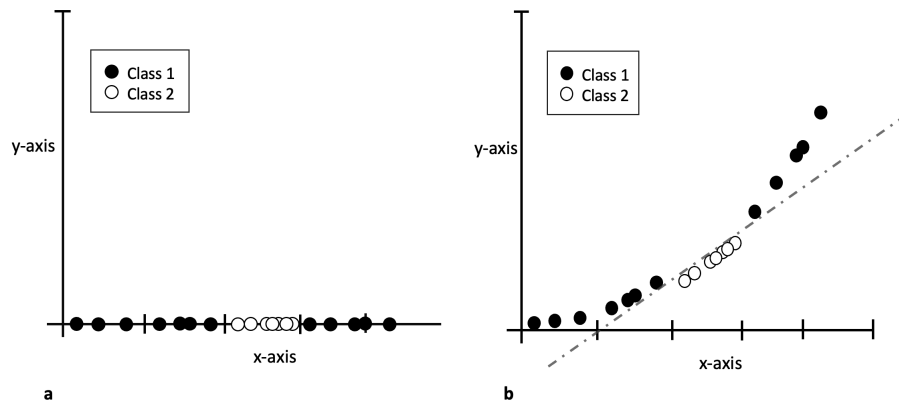


Figure 17: (a) The data on the x-axis are the original non-separable data. (b) Application of a second-degree polynomial function to make the two classes separable

Classification Problem: Tumor Diagnosis

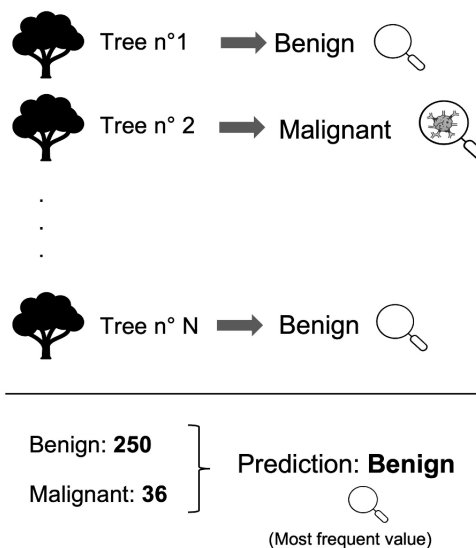


Figure 18: Application of the random forest algorithm. Each decision tree in the forest calculates its own prediction: 250 indicate that the tumour is benign and 36 malignant. It is shown that the final result is the most frequent prediction made by the entire forest (benign tumour)

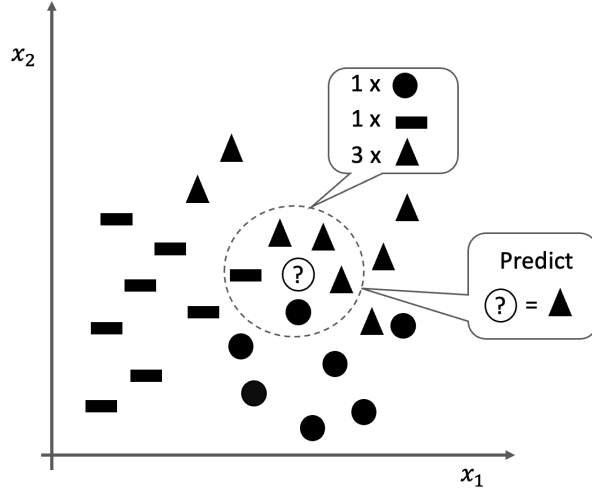


Figure 19: Representation of how the K-nearest neighbours algorithm works. Considering the new point to classify (?), the category is assigned based on the five nearest neighbours (K=5). In this case 3 triangles versus 1 circle versus 1 rectangle

are especially well-suited for medical imaging tasks due to their ability to automatically learn relevant features from raw data, thereby eliminating the need for manual feature engineering. Among the various DL approaches, CNNs and ViTs have emerged as the most prominent in the field of radiology. These models integrate both feature extraction and classification within a unified architecture, making them particularly effective for analyzing complex and high-dimensional medical images.

5.4.1 Foundations of Deep Neural Networks

The foundational concept of the artificial neuron dates back to the early perceptron model introduced by McCulloch and Pitts in 1943 [87]. This model, which could only process linearly separable data, featured a simple architecture consisting of an input and an output layer. The limitations of this approach were subsequently addressed by introducing intermediate layers, known as hidden layers, resulting in the development of the multilayer perceptron (MLP) (Figure 20). MLPs are composed of fully connected neurons, each computing a weighted sum of its inputs and passing the result through a non-linear activation function.

During training, the model parameters (weights) are updated through a process known as backpropagation, in which gradients of a loss function are computed and propagated backwards through the network. Optimization algorithms such as stochastic gradient descent (SGD) or its variants are then used to update the weights to minimize the prediction error. The configuration of a neural network, including the number of layers, choice of activation functions, learning rate, and optimization algorithm, constitutes its set of hyperparameters [88, 89]. These must be tuned according to the specific nature of the task and the available data. While it may seem intuitive that deeper networks yield better performance due to their greater representational capacity, the Universal Approximation Theorem asserts that even a neural network with a single hidden layer is capable of approximating any continuous function under certain conditions [90, 91]. Nevertheless, deeper networks often offer practical advantages in capturing hierarchical patterns in complex datasets.

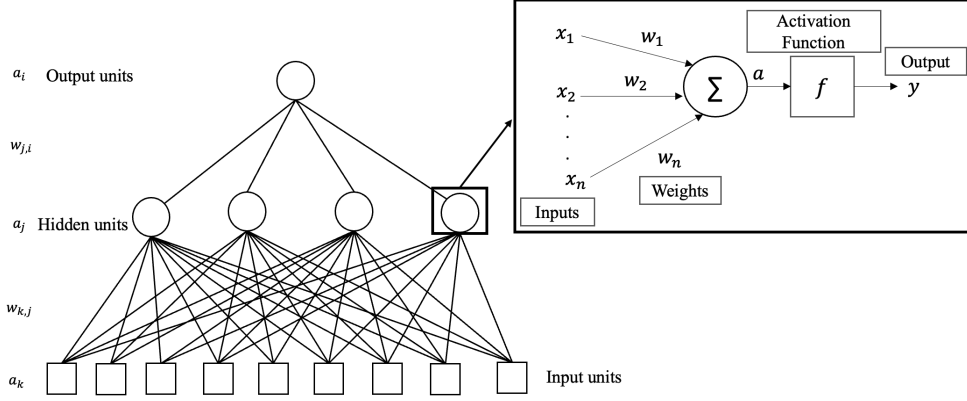


Figure 20: Representation of the multilayer perceptron, composed of one input layer, one hidden layer, and one output layer. Each unit of the hidden layer and output layer is a single perceptron, represented in the box

5.4.2 Convolutional Neural Networks

In traditional classification pipelines, images or ROIs must first be converted into numerical features, such as intensity histograms, texture descriptors, or shape-based metrics, before being fed into a model. These handcrafted features have proven useful in many clinical tasks [55, 56], with additional transformations such as wavelet decompositions offering further improvements in specific use cases [92].

CNNs represent a paradigm shift by automating the feature extraction process. These models are designed to take entire images or ROIs as input and internally compute increasingly abstract feature representations, which are then passed through fully connected layers, often referred to as dense layers, for final classification [57]. This end-to-end structure makes CNNs highly effective for radiological tasks, including lesion detection, segmentation, and diagnostic classification [93, 94, 95]. CNNs are composed of a series of layers organized in a specific sequence. The initial input layer processes the image as a three-dimensional matrix of pixel values, with dimensions corresponding to width, height, and colour channels. This is followed by convolutional layers, where small filters (kernels) slide over the input to extract local patterns. These filters are learnable parameters that are optimized during training to detect specific features such as edges, textures, or other relevant structures [96]. After each convolution operation, activation functions such as ReLU are applied to introduce non-linearity. To reduce the spatial dimensionality and control overfitting, pooling layers are inserted periodically. Pooling layers (e.g., max-pooling or average-pooling) aggregate the information in small regions, preserving essential features while discarding less informative details [97]. As the data progresses through successive convolutional and pooling layers, the network captures increasingly abstract and high-level features. These so-called *deep features* are then flattened and passed through dense layers to produce the final classification output. Figure [21] illustrates the convolution process.

5.4.3 Choosing the Right Classifier in Medical Imaging

Selecting an appropriate classifier for radiological tasks involves balancing performance, interpretability, data availability, and computational constraints. According to the “No Free Lunch” theorem [98], no single algorithm performs best across all problems, which

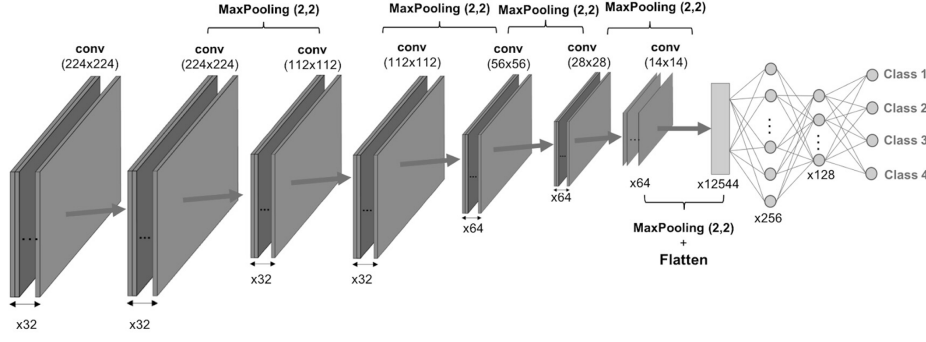


Figure 21: Example of CNN architecture. The input images are fed into the convolutional and pooling layers for feature extraction. In the end, the resulting flattened feature vector is fed into the dense layer to perform the classification task

underscores the importance of context-specific decision-making [99, 100]. Shallow models, such as LR and SVM, are well-suited to small or medium-sized datasets, especially when the features are well-crafted and low-dimensional. Their relatively simple structure allows for greater transparency and often requires fewer computational resources. Ensemble methods such as RF and gradient increase offer improved accuracy while maintaining partial interpretability and have been demonstrated to be effective in many radiological applications [77, 101]. In contrast, deep learning methods demand large annotated datasets to fully exploit their capabilities. When data is limited, transfer learning from pre-trained models can mitigate this challenge by leveraging features learned from large external datasets. Nevertheless, deep models typically require access to GPUs and substantial memory, limiting their use in resource-constrained environments. In summary, the classifier choice should reflect the specific demands of the task: shallow models are often preferred when interpretability, limited data, or computational efficiency are priorities, while deep learning is more appropriate for large-scale, high-dimensional problems where performance is paramount. In some cases, hybrid frameworks that combine both paradigms may offer optimal performance and usability.

5.5 Evaluation Metrics

In the field of medical image analysis, the evaluation of classification models requires the use of specific statistical metrics that reflect both diagnostic accuracy and clinical relevance. Among the most widely adopted indicators are accuracy, sensitivity, specificity, and the Area Under the Receiver Operating Characteristic curve (AUROC). These metrics offer complementary insights into the performance of a model, particularly in scenarios characterized by class imbalance or asymmetric costs of misclassification. Accuracy represents the proportion of correctly classified instances among all samples and provides a general measure of performance. However, it may be misleading in datasets where one class significantly dominates the others. In such cases, a model might achieve high accuracy by predominantly predicting the majority class, while failing to detect the minority one, often the clinically significant class, such as malignant lesions. To address this limitation, sensitivity (also known as recall or true positive rate) is used to quantify the model’s ability to correctly identify positive cases, which in a medical context usually correspond to pathological conditions. High sensitivity is crucial when the goal

is early detection, as in the case of melanoma or breast cancer, where missing a true positive may lead to severe clinical consequences. Conversely, specificity measures the proportion of true negatives correctly identified, highlighting the model’s effectiveness in avoiding false positives. This is particularly important in reducing unnecessary anxiety, additional diagnostic procedures, or invasive treatments that can result from incorrect positive classifications. AUROC, or Area Under the ROC Curve, provides a global view of the classifier’s ability to discriminate between classes across all possible thresholds. It is especially valuable in binary classification problems, offering a single scalar value that summarizes the trade-off between sensitivity and specificity. A model with AUROC close to 1 is considered to have excellent discriminatory capacity, while an AUROC near 0.5 indicates performance equivalent to random guessing.

These metrics are mathematically defined as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$Sensitivity = \frac{TP}{TP + FN} \quad (2)$$

$$Specificity = \frac{TN}{TN + FP} \quad (3)$$

Here, TP , TN , FP , and FN denote the number of true positives, true negatives, false positives, and false negatives, respectively.

The selection of appropriate evaluation metrics must align with the specific clinical objective. In screening applications, maximizing sensitivity is often prioritized to ensure that no pathological case is overlooked. In contrast, in diagnostic confirmation or resource-limited settings, specificity may be emphasized to avoid excessive false positives. Therefore, a careful interpretation of these indicators in relation to the clinical context is essential when evaluating the performance of a medical AI system.

6 Explainable Artificial Intelligence: Concepts and Techniques

The increasing integration of AI into high-stakes domains such as medicine, autonomous driving, and finance has introduced new demands for transparency and interpretability. While AI models, particularly deep neural networks, achieve state-of-the-art performance in tasks ranging from image recognition to diagnosis support, their internal decision-making logic remains often inaccessible to users. This lack of interpretability, commonly referred to as the "black-box" problem, has prompted the development of XAI, a research field focused on providing human-understandable explanations for algorithmic outputs [102, 103]. Explainability is particularly crucial in medical applications, where understanding the rationale behind AI predictions is essential to ensure trust, safety, and clinical accountability. In such contexts, the goal is twofold: first, to increase trust and adoption among end users, especially in domains where decisions carry significant consequences, such as healthcare; and second, to ensure accountability and facilitate model debugging, validation, and bias detection [104]. In clinical contexts, for instance, understanding the rationale behind an AI diagnosis can support physicians in decision-making processes, rather than merely providing an opaque output. As such, XAI becomes a cornerstone for developing AI systems that are not only accurate but also safe, ethical, and interpretable.

6.1 Global and Local Explanation Strategies

Explanation techniques in XAI are often categorized based on their scope. A key distinction is made between *global* and *local* methods. Global explanations aim to provide a comprehensive understanding of how the model behaves across the entire dataset, revealing general patterns or feature importance. Local explanations, on the other hand, focus on specific individual predictions, allowing users to understand why the model made a certain decision in a given instance. Among the most prominent local explanation methods are LIME (Local Interpretable Model-agnostic Explanations) [105] and SHAP (SHapley Additive exPlanations) [106]. LIME operates by generating perturbed versions of the input and learning a simpler interpretable model around that instance. SHAP builds on game theory to assign contribution scores (Shapley values) to each feature, with formal guarantees of consistency and local accuracy. A further distinction can be drawn between *model-agnostic* methods, which can be applied to any model by observing inputs and outputs, and *model-specific* methods, which leverage the internal structure of the model. While the former are flexible and broadly applicable, the latter can provide deeper insights when applied to architectures such as neural networks or tree ensembles. In addition, explainability tools may differ in their mode of presentation: some provide quantitative feature scores requiring expert interpretation, while others offer visual interfaces, such as bar plots or saliency maps, that facilitate communication of results to a wider audience. These visualizations are especially helpful in sensitive domains like healthcare, where clarity and interpretability are key to decision support.

6.2 Explainability in Deep Learning and Medical Imaging

DL models, especially CNNs and transformer-based architectures, have set new benchmarks in image classification, segmentation, and natural language understanding. However, their depth and complexity often exacerbate the black-box issue. Unlike linear

models or decision trees, deep networks do not offer easily interpretable structures or parameters. As such, specialized techniques are required to uncover the decision logic encoded within their multiple layers of abstraction. Among the most widely used explanation techniques for deep neural networks is Gradient-weighted Class Activation Mapping (Grad-CAM) [107]. Grad-CAM utilizes the gradients of target class scores flowing into the final convolutional layers to produce coarse localization maps, highlighting the regions of the input image that are most influential for a particular classification. These heatmaps provide visual explanations that can be overlaid on the original input, allowing users to inspect whether the network focused on clinically meaningful regions, such as tumours or lesions, in medical imaging scenarios. Integrated Gradients [108] is another important method that attributes the prediction of a deep network to its input features by integrating the gradients of the output with respect to the input along a path from a baseline input (typically zero) to the actual input. This method satisfies desirable axioms such as sensitivity and implementation invariance, making it theoretically grounded and reliable for feature attribution. Transformer-based models, such as the Vision Transformer (ViT) [109], introduce additional interpretability avenues via their attention mechanisms [110]. Attention maps, which quantify the importance of each input patch relative to others, can be visualized to assess which regions the model prioritized during inference. These attention weights provide native interpretability and are particularly valuable in domains like medical imaging, where spatial context is crucial.

In radiological applications, the use of XAI has been particularly impactful. DL models have shown remarkable capabilities to detect abnormalities in CT, magnetic resonance, and dermatoscopic images. However, the lack of transparency in their decision-making has been a barrier to clinical adoption. Techniques such as Grad-CAM and Integrated Gradients are increasingly used to generate explanations aligned with expert reasoning. For example, in the classification of skin cancers, heat maps can highlight the borders of the lesion or the specific texture patterns that led the model to predict melanoma rather than benign nevus. Moreover, explainability contributes not only to the acceptance of AI tools in clinical practice but also to their safety and robustness. In longitudinal studies or clinical trials, being able to trace model behaviour over time or across patient cohorts allows researchers to verify consistency and detect possible drifts or biases. This is especially critical in high-risk applications, such as brain lesion detection or tumour classification, where misclassification can have severe consequences. The integration of explainability into medical AI systems thus enables a symbiotic relationship between human expertise and machine precision. Rather than replacing radiologists, explainable AI acts as a cognitive aid, enhancing diagnostic confidence, identifying edge cases, and supporting clinical decisions through transparent and interpretable insights.

7 AI Applications in Radiological and Oncological Imaging

The graphical environment provided by DicomOS was used to test deep learning models tailored to the classification of radiological images in oncological contexts. Specifically, the integrated interface enabled the management of complete classification pipelines for both dermoscopic and mammographic images, from dataset selection to model evaluation, without requiring terminal-based scripting. This approach facilitated the standardized application of ViTs with various data augmentation strategies, thereby promoting reproducibility and ease of experimentation. In real-world clinical scenarios, available datasets are frequently limited in size and often exhibit significant class imbalance. These challenges necessitate the use of data augmentation techniques capable of simulating additional observations, thus enhancing the generalization capabilities of learning algorithms. Classical transformations include geometric operations, illumination adjustments, contrast enhancement, and noise injection. More recent approaches, such as Diffusion Models, allow the generation of synthetic samples by learning the underlying data distribution and sampling new images that realistically mimic clinical variability. ViTs have recently emerged as a powerful alternative to traditional CNNs for medical image analysis. Unlike convolutional architectures, ViTs process images by dividing them into non-overlapping patches and leveraging multi-head self-attention mechanisms to capture long-range dependencies and spatial relationships across the entire image. This global reasoning capability proves particularly effective in diagnostic tasks such as the recognition of melanoma or the classification of breast lesions.

7.1 Transformer Architectures: Foundations and Relevance in Medical AI

The Transformer architecture, first introduced by [110], marked a fundamental departure from traditional sequence modelling techniques by replacing recurrence and convolution with a novel mechanism known as self-attention. Originally developed for natural language processing tasks such as machine translation, Transformers have since been successfully adapted to a wide range of domains, including computer vision, speech processing, and, more recently, medical imaging. Their ability to process input data as a whole, rather than sequentially or locally, allows them to model global dependencies and contextual relationships with exceptional efficiency.

At the core of the Transformer lies the self-attention mechanism, which enables the model to assign dynamic weights to different elements in the input sequence when computing contextual representations. Each token in the sequence is associated with three vectors: a query, a key, and a value. The attention score is computed as a scaled dot product between queries and keys, followed by a softmax operation that normalizes the weights. These weights are then applied to the value vectors to produce a weighted sum that represents the contextual embedding of each token. This process is performed simultaneously across multiple “heads” in a multi-head attention layer, allowing the model to capture diverse and complementary aspects of the data. A standard Transformer encoder is composed of stacked layers, each consisting of a multi-head self-attention module and a feed-forward neural network, both wrapped with residual connections and layer normalization. The absence of recurrence and convolution allows Transformers to benefit from massive parallelization during training, making them more scalable and faster to

train on large datasets compared to recurrent neural networks (RNNs) and long short-term memory (LSTM) models. Transformers also integrate positional encodings to retain information about the order of elements in the input sequence. Since the model architecture itself is permutation-invariant, positional encodings, either sinusoidal or learned, are added to the input embeddings to preserve sequence order, which is essential for tasks such as language modelling or image patch processing. The flexibility and expressiveness of the Transformer architecture have led to its rapid adoption across domains. In medical AI, the ability to capture long-range dependencies and contextual relationships has proven especially valuable in complex diagnostic scenarios. Applications include clinical text summarization, drug discovery, genomic sequence modelling, and increasingly, the analysis of high-dimensional medical images. By modelling the intricate relationships within and across image regions, Transformers provide an alternative to convolutional architectures, particularly in tasks where spatial context plays a critical role in interpretation. The success of Transformers in vision tasks has led to the development of domain-specific adaptations, most notably the Vision Transformer (ViT), which repurposes the self-attention framework for image classification and analysis. This evolution has unlocked new possibilities for applying attention-based models to radiology and oncology, where interpretability, precision, and context-awareness are paramount.

7.1.1 Vision Transformers for Medical Image Classification

ViTs architecture represents a significant paradigm shift in computer vision, adapting the Transformer framework, originally introduced for sequence modelling in natural language processing, to the domain of image analysis [111]. Unlike CNNs, which exploit local spatial hierarchies through convolutional filters, ViTs operate on sequences of image patches and leverage self-attention mechanisms to model global dependencies among different regions of an image. This distinctive approach enables the model to capture long-range spatial relationships, a property particularly beneficial for medical imaging tasks where critical diagnostic information may be distributed across non-contiguous image regions. In a typical ViT architecture, the input image is first divided into a sequence of fixed-size non-overlapping patches, each of which is flattened into a one-dimensional vector. These vectors are then projected linearly into a latent space of predefined dimensionality, yielding a set of patch embeddings that serve as input tokens to the model. A positional encoding is added to each patch embedding to preserve spatial information lost during the flattening process. This enriched sequence is then passed through a stack of Transformer encoder blocks, each comprising a multi-head self-attention mechanism and a feed-forward neural network, both followed by layer normalization and residual connections. The self-attention module within each Transformer layer plays a central role in the representational power of the model. It enables the network to weigh the relative importance of different image patches when computing the representation of a given region. This mechanism allows the ViT to effectively model contextual dependencies at multiple scales, making it particularly suitable for analyzing complex patterns in high-resolution medical images. The attention weights are computed using a scaled dot-product between query, key, and value matrices derived from the input embeddings, allowing the model to attend to the most informative regions dynamically. After processing the entire sequence through multiple encoder layers, a classification token, usually a learnable vector prepended to the input sequence, is extracted and passed through an MLP head. This head typically consists of two fully connected layers with a non-linear activation function, such as GELU (Gaussian Error Linear Unit) [112], and outputs the final class prediction.

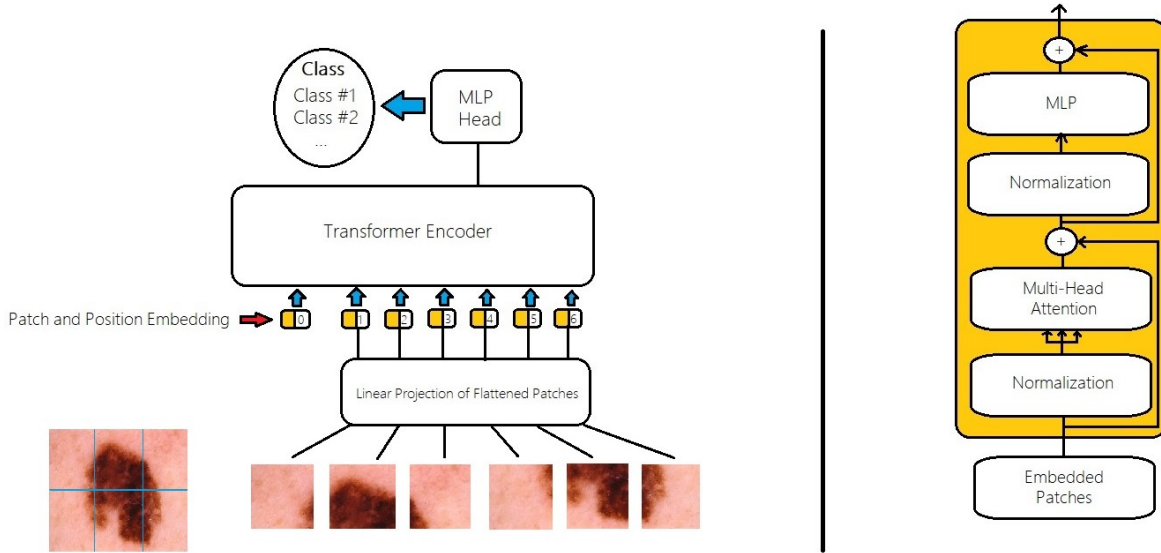


Figure 22: The computational architecture, based on the ViT, used in this study.

In Fig 22, the implemented architecture, based on the ViT, is reported. One of the major advantages of Vision Transformers lies in their architectural flexibility and scalability. They eliminate the inductive biases inherent in convolutional layers, such as locality and translation invariance, allowing the model to learn these properties directly from data. However, this also results in a higher data requirement for effective training. To mitigate this, ViTs are often pre-trained on large-scale datasets, such as ImageNet, and subsequently fine-tuned on domain-specific medical imaging datasets, which tend to be smaller and more imbalanced.

In the context of medical imaging, ViTs have demonstrated promising results across a variety of applications, including skin lesion classification, breast cancer detection, and histopathological image analysis. Their ability to model global context, combined with attention-based interpretability, offers a compelling alternative to traditional CNN-based approaches, particularly in scenarios where clinical relevance is spread across multiple spatially disjoint regions. Moreover, the self-attention maps produced by ViTs provide a form of visual explanation that can support clinical interpretability and aid in the validation of model predictions. Recent research has also explored hybrid models that integrate convolutional front-ends with Transformer-based backbones, leveraging the inductive strengths of CNNs in early feature extraction and the contextual modelling capabilities of Transformers in later stages. These hybrid architectures have been shown to achieve state-of-the-art performance in many medical image classification benchmarks, suggesting that the future of deep learning in radiology may reside in such synergistic designs. Overall, the adoption of Vision Transformers in medical image analysis reflects a broader shift toward more expressive and flexible DL architectures capable of capturing the complexity inherent in diagnostic imaging tasks. Their integration within user-friendly platforms like DicomOS ensures that even advanced AI models can be deployed and validated within clinically accessible environments, fostering a closer alignment between computational innovations and real-world medical practice.

7.2 Classical Data Augmentation in Medical Imaging

Data augmentation is a widely adopted strategy in ML that aims to artificially increase the size and diversity of a training dataset by applying a series of label-preserving transformations to the original data. This process is particularly essential in medical imaging, where datasets are often small, class-imbalanced, or limited by ethical and logistical constraints. Augmentation plays a dual role: it mitigates the risk of overfitting in DL networks and enhances the generalization ability of models by exposing them to a broader spectrum of plausible input variations [113]. In the context of medical image analysis, classical augmentation techniques typically include geometric, photometric, and noise-based transformations. Geometric augmentations involve operations such as rotation, flipping, scaling, cropping, and affine transformations, all of which preserve the anatomical structure while altering its spatial representation. These are especially effective in dermoscopy and radiography, where the diagnostic features remain invariant under such transformations [114]. Photometric transformations adjust pixel intensities to simulate variations in lighting, contrast, and colour balance, mimicking the variability found in real clinical scenarios. For grayscale modalities such as X-ray or MRI, contrast-limited adaptive histogram equalization, gamma correction, and brightness scaling are frequently used to enrich the training distribution [115]. In colour images like dermatological photographs, channel-specific transformations are often applied to simulate device or acquisition differences. Noise injection is another important technique, particularly relevant when training models to be robust to acquisition artefacts or low signal-to-noise ratios. Gaussian noise, Poisson noise, and motion blur are commonly introduced during training to improve resilience under suboptimal acquisition conditions [116].

Recent works have also proposed more advanced strategies such as elastic deformations, cutout, mixup, and random erasing [117, 118]. Elastic deformations simulate realistic tissue distortions, while cutout and random erasing mask out portions of the image to force the model to focus on global features. Mixup, on the other hand, generates new samples by convexly combining two images and their labels, encouraging smoother decision boundaries. Despite their simplicity, classical data augmentation methods have shown remarkable effectiveness in boosting classification, segmentation, and detection performance across multiple domains of medical AI [13]. However, their transformations are handcrafted and often constrained in their ability to reflect complex anatomical or pathological variability. To address this limitation, the field has recently shifted toward generative data augmentation, where synthetic images are created by learning the underlying data distribution, an approach exemplified by the emergence of Diffusion Models.

7.3 Diffuser-Based Data Augmentation for Synthetic Medical Imaging

In recent years, Diffusion Models have emerged as a powerful and flexible class of generative architectures, capable of producing high-fidelity synthetic images that rival or surpass those generated by traditional approaches such as Variational Autoencoders (VAEs) [119] and Generative Adversarial Networks (GANs) [120]. Particularly in medical imaging, where the availability of large annotated datasets is limited and data imbalance remains a critical issue, Diffusion Models offer a promising solution for augmenting training corpora with realistic, label-consistent synthetic examples [121]. The specific family of models employed in this work is known as Denoising Diffusion Probabilistic Models (DDPMs) [122]. These models are grounded in a generative framework that learns to reverse a Markovian

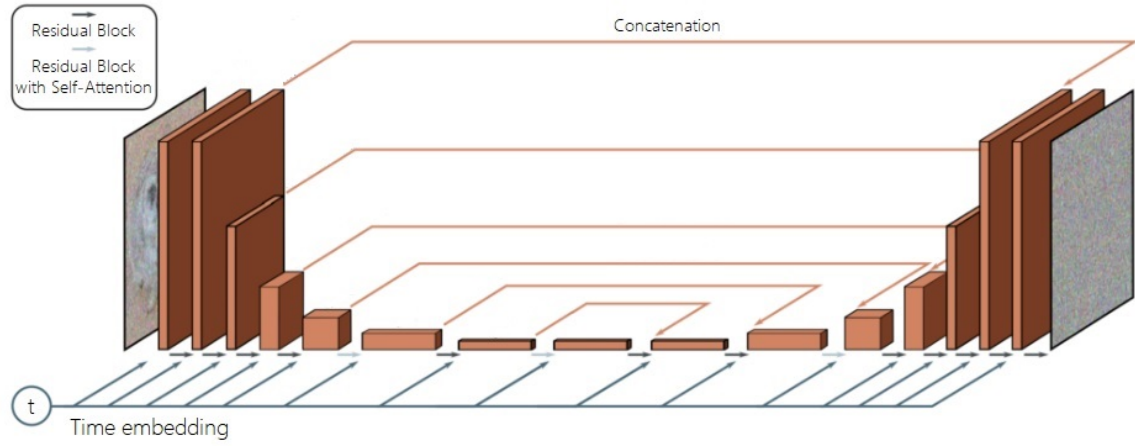


Figure 23: The Diffuser architecture used for the generation of synthetic images.

process of progressive noise corruption. During training, each input image is repeatedly perturbed by Gaussian noise over a series of time steps, gradually transforming it into a distribution that approximates pure white noise. The DDPM is then optimized to invert this process-step-by-step-restoring the corrupted image to its original state. By learning this denoising trajectory, the model implicitly captures the underlying data distribution and can be used to sample entirely new synthetic images from noise.

To implement this generative process, the model architecture is typically based on a U-Net [123], enhanced with residual connections and self-attention layers. These augmentations enable the network to effectively model both local texture and global structure, which are essential in medical images containing subtle diagnostic patterns. The denoising network is conditioned on the current time step via learned embeddings that are injected into multiple layers of the U-Net, allowing the model to track the progression of the reverse diffusion process. The flexibility of Diffusion Models extends beyond unconditional generation. In advanced applications, conditioning information such as segmentation masks, class labels, or textual prompts can be integrated to guide the generation process. In the context of this work, a purely image-conditioned DDPM was trained using real melanoma images, without auxiliary information such as textual descriptors. The goal was to enrich the training dataset with high-quality artificial samples that reflected the statistical properties of both benign and malignant lesions, enhancing the robustness and diversity of the classification model. Compared to GANs, which often suffer from training instability and mode collapse, Diffusion Models exhibit superior convergence properties and generate more diverse outputs. They also outperform VAEs in terms of image realism, while avoiding the blurriness commonly associated with reconstruction-based methods. Recent studies have demonstrated the potential of Diffuser-based architectures in medical domains such as dermatology, mammography, and retinal imaging [124].

Figure 23 illustrates the architecture of the Diffusion Model used for synthetic image generation in this study. By leveraging this approach, it was possible to expand the training set with realistic, high-resolution dermoscopic images, thereby reducing class imbalance and improving model generalization without compromising clinical validity.

8 Vision Transformer Applications in Oncological Imaging

Breast cancer and melanoma are among the most prevalent and impactful cancers worldwide. According to the World Health Organization, skin cancer represents the most commonly diagnosed cancer, while breast cancer is the leading cause of cancer-related mortality among women. Early detection is crucial in both cases, as it significantly improves patient outcomes and survival rates. In this context, radiological and dermatoscopic imaging plays a central role in screening and diagnosis. However, the accuracy of image-based diagnosis is still affected by multiple factors, including human perception, variability in lesion appearance, breast tissue density, and class imbalance in available datasets. Therefore, the integration of AI-based tools into oncological imaging workflows is becoming increasingly essential. Recent advances in deep learning have introduced ViTs as promising alternatives to CNNs. By leveraging self-attention mechanisms, ViTs are capable of modelling global spatial relationships within the image, thus overcoming some of the intrinsic limitations of CNNs in capturing long-range dependencies. This architectural flexibility has led to successful applications of ViTs in both mammographic and dermatoscopic image classification, particularly in scenarios where complex contextual cues are needed for accurate diagnosis.

In parallel, the performance of ViT-based classifiers is highly dependent on the availability and diversity of the training data. In clinical contexts, datasets are often limited and highly imbalanced, with a small proportion of malignant cases. To address this issue, data augmentation strategies have been widely adopted. Alongside traditional augmentation methods (e.g., geometric transformations, colour shifts, noise injection), this work explores the use of DDPMs to generate synthetic samples representative of benign and malignant lesions. These synthetic images are used to enrich the training data and improve the generalization capabilities of the model. The experiments presented in this section were conducted entirely within the DicomOS environment, which enabled a complete AI workflow, from data selection to model training and evaluation, through a user-friendly graphical interface. This unified framework promotes reproducibility and facilitates the integration of advanced AI tools into real-world radiological practice. The following subsections describe the two experimental pipelines: one dedicated to melanoma detection using dermatoscopic images, and the other focused on mammogram classification. Both pipelines adopt a ViT-based architecture, with a comparative evaluation of classical and Diffuser-based data augmentation techniques.

8.1 Dermatoscopic Image Classification for Melanoma Detection

8.1.1 Dataset Description: ISIC 2017 Challenge

The first application explored in this study focuses on the classification of dermoscopic images of cutaneous lesions, to distinguish between malignant melanoma, nevocellular nevus, and seborrheic keratosis. The dataset used for this task was derived from the ISIC Archive¹, and was initially released as part of the ISIC 2017 Challenge – Skin Lesion Analysis Towards Melanoma Detection [125]. The ISIC 2017 Challenge proposed three distinct tasks: lesion segmentation, dermoscopic feature detection, and lesion classification. The

¹<https://challenge.isic-archive.com/data/#2017>

present work specifically addresses the third task, which involves assigning dermoscopic images to one of three diagnostic categories: malignant melanoma (MM), nevocellular nevus (NCN), and seborrheic keratosis (SK). From a clinical perspective, melanoma represents a malignant tumour originating from melanocytes (melanocytic origin), whereas seborrheic keratosis is a benign neoplasm of keratinocyte origin (non-melanocytic), and nevus is a benign lesion also derived from melanocytes. The dataset includes 2000 training images in JPEG format, annotated by expert dermatologists using histopathological examination or clinical assessment as reference standards. This annotation strategy ensures high label reliability, which is critical in the context of medical image classification. The training set comprises 374 melanoma cases, 254 seborrheic keratosis cases, and 1372 nevus images. A separate validation set consists of 150 samples, while the test set includes 600 images with heterogeneous characteristics in terms of size, illumination, and acquisition conditions. The complete dataset structure is summarized in Table 3. Also, Figure 24 represents examples from the three classes: MM, NCN, SK. Inhomogeneous dark spots with irregular shapes usually characterize MM; and similar features are shared with the NCN and SK classes, making the classification task particularly challenging.

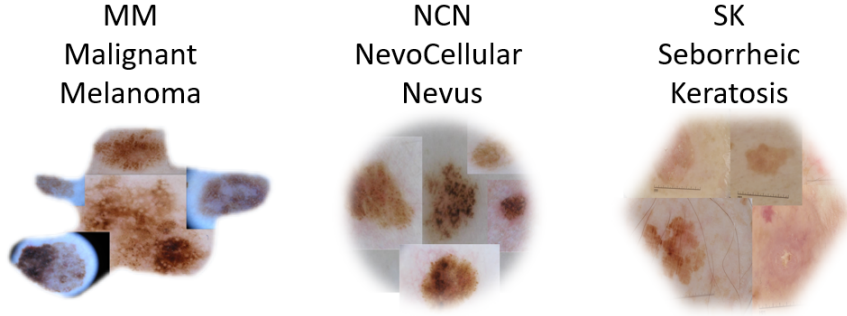


Figure 24: Some examples belong to the used dataset, with the three classes highlighted.

Table 3: ISIC dataset summary.

Class	Training Set	Validation Set	Test Set
Melanoma (MM)	374	30	118
Nevus (NCN)	1372	78	392
Seborrheic Keratosis (SK)	254	42	90
Total	2000	150	600

8.1.2 Classical Data Augmentation

To enhance the diversity of the training set and address class imbalance, a set of conventional augmentation techniques was applied to the original melanoma and seborrheic keratosis images. The goal was to simulate the variability typically encountered in clinical scenarios, thereby improving the robustness and generalization capabilities of the model. These transformations included brightness adjustment, contrast enhancement using CLAHE [126], gamma correction, hue and saturation shifts, as well as the injection of Gaussian noise (mean=0, standard deviation=25). Rotations and flips were also used to account for geometric variance. By applying these transformations to the 374 melanoma images, a total of 2992 augmented samples were produced. A random selection step was

then performed to retain 998 images, ensuring parity with the number of nevus samples. An equivalent procedure was adopted for the SK class to maintain a balanced dataset. Figure 25 illustrates some examples of the classical transformations.

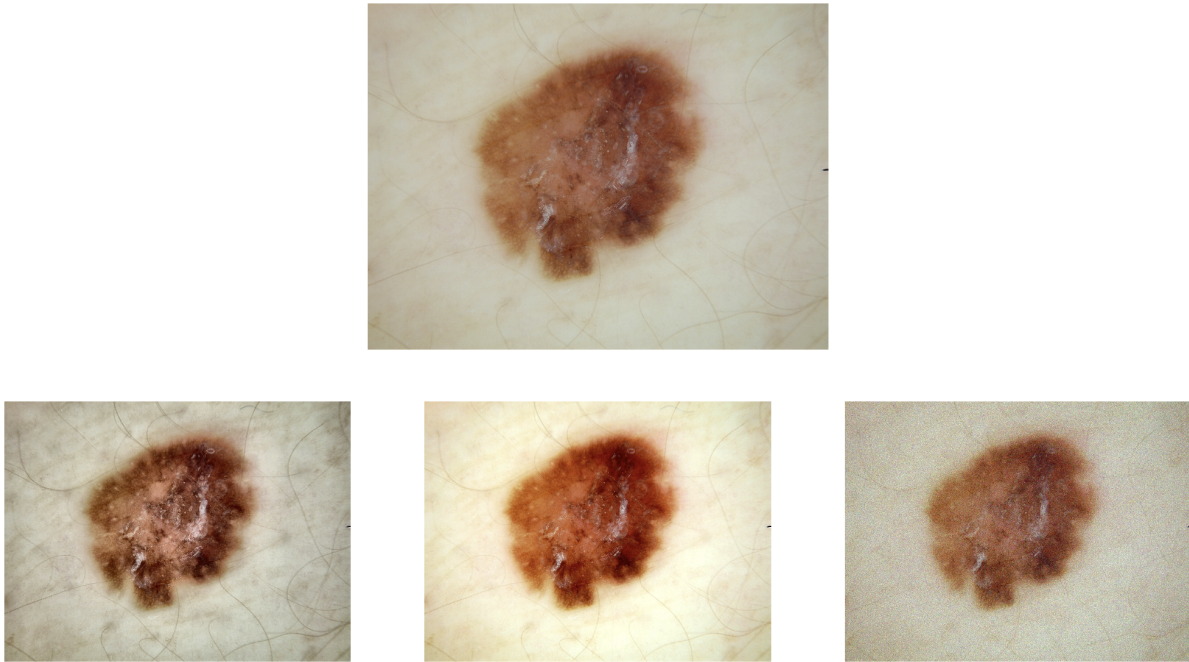


Figure 25: Pre-processing examples applied to a dermatoscopic image. Top: original melanoma image. Bottom (from left to right): image after CLAHE, contrast enhancement, and Gaussian noise addition.

8.1.3 Methodology

A ViT was fine-tuned using 32×32 patches, employing the same hyperparameters described in [127]. Initially, the model was trained using the augmented training set based solely on conventional DA techniques. RandAugment [128] was adopted as the primary pipeline to randomly select and apply a sequence of transformations. To further boost sensitivity in detecting melanoma, especially given its clinical importance, a DDPM [122] was trained on melanoma and SK images from the original training set. The goal was to generate additional synthetic samples to further balance the dataset. In total, 2116 synthetic images were created (998 melanoma and 1118 SK), thereby equalizing the training sample size across all three classes (Table 4).

Table 4: ISIC dataset after augmentation.

Class	Training Set	Validation Set	Test Set
Melanoma (MM)	1372 (374+998)	30	118
Nevus (NCN)	1372	78	392
Seborrheic Keratosis (SK)	1372 (254+1118)	42	90

Figure 26 shows a comparison between a real melanoma image and one generated through the diffusion model. The synthetic image successfully preserves the visual traits of the original while introducing additional variability that enhances model generalization.

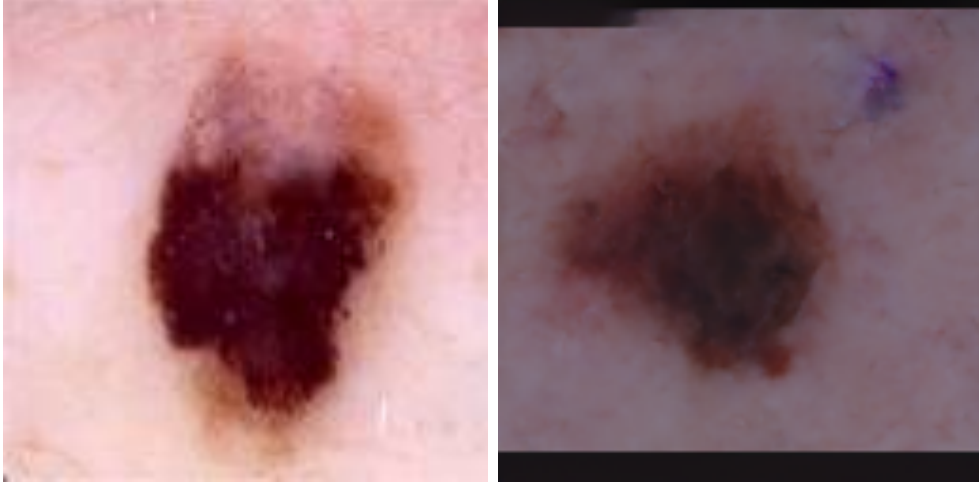


Figure 26: Real melanoma image (left) vs. synthetic melanoma image (right) generated using DDPM.

8.1.4 Results and Discussion on Melanoma Classification

The experimental evaluation highlighted the substantial impact of data augmentation (DA) strategies on the performance of ViT-based models in classifying dermoscopic images. The analysis focused on three scenarios: a baseline model trained without any augmentation, a model trained with classical augmentation techniques, and a model enhanced with synthetic samples generated through a DDPM.

Figure 27 illustrates the confusion matrices obtained from the ViT-based classifier under each experimental setting. A notable improvement in sensitivity for the *Melanoma* class was observed as the number of false negatives decreased progressively across the three setups, from 17 in the baseline to 13 with classical DA, and finally to 10 when using Diffuser-generated samples. These results provide strong evidence that the inclusion of high-quality synthetic images contributes to the more robust detection of clinically relevant but underrepresented lesion types.

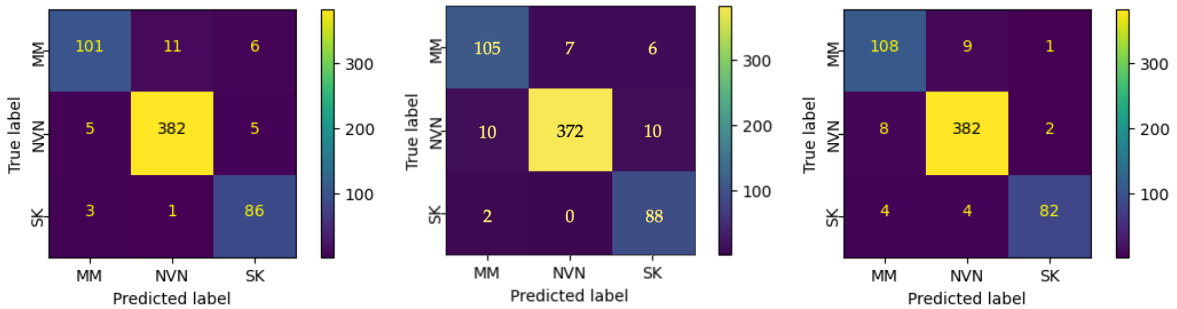


Figure 27: Confusion matrices obtained by the ViT-based classifier: (left) without DA, (centre) with classical DA, (right) with Diffuser-based DA.

Table 10 summarizes the classification metrics for each class using a one-vs-rest evaluation scheme. For the *Melanoma* category, the model trained with Diffuser-based DA exhibited the highest sensitivity (0.915) and AUROC (0.945), while preserving a high specificity (0.975). Compared to the baseline configuration, the improvement in sensitivity corresponds to a 6% gain, which is particularly important for the early detection of

malignant lesions. This confirms that the enriched feature diversity introduced by synthetic data contributes to improving the model’s ability to correctly recognize complex skin patterns that are often underrepresented in clinical datasets.

Table 5: Performance comparison between ViT-based classifier without DA, with classical DA, and with Diffuser-based DA. The evaluation followed a one-*vs.*-rest scheme.

Class	Evaluation metric	ViT-based	ViT-based + Classic DA	Vit-based + Diffusers DA
Melanoma (MM)	Accuracy	0.958	0.958	0.963
	Sensitivity	0.856	0.889	0.915
	Specificity	0.983	0.975	0.975
	AUROC	0.919	0.932	0.945
Nevus (NCN)	Accuracy	0.963	0.955	0.962
	Sensitivity	0.974	0.948	0.974
	Specificity	0.942	0.966	0.938
	AUROC	0.958	0.957	0.956
Seborrheic Keratosis (SK)	Accuracy	0.975	0.970	0.982
	Sensitivity	0.955	0.977	0.911
	Specificity	0.978	0.968	0.994
	AUROC	0.967	0.973	0.953

For the *Nevus* class, which is the most represented category in the training set, performance metrics remained stable across all configurations. The model exhibited consistently high accuracy and sensitivity, suggesting that this class was already sufficiently modelled by the original data distribution. Minor fluctuations in specificity and AUROC were observed, likely due to the increased complexity of balancing the remaining two classes.

In the case of *Seborrheic Keratosis*, the addition of classical augmentation led to slight gains in sensitivity and AUROC, while the Diffuser-augmented setup improved accuracy and specificity to 0.982 and 0.994, respectively. However, this was accompanied by a reduction in sensitivity (0.911), indicating a possible shift in the classifier’s decision boundary. This could be interpreted as a tendency to reduce false positives at the expense of increased false negatives, a trade-off that may be acceptable in low-risk lesions but undesirable in malignant cases. From a broader perspective, the adoption of Diffuser-based DA proved most advantageous for melanoma classification, where the cost of misclassification is highest. The method achieved a reduction of 41% in false negatives compared to the baseline and 23% when compared to classical augmentation. These improvements are not only statistically meaningful but also clinically significant, as they directly support early-stage detection and treatment planning. Moreover, the integration of Diffuser models aligns well with the current shift toward generative AI in medical imaging. Unlike classical augmentation, which applies predefined transformations, Diffusers learn the underlying data distribution and generate realistic, diverse instances that improve generalization. This is particularly useful in domains such as dermatology, where image variability is high and manual annotations are costly and subjective.

In conclusion, the combination of a ViT-based architecture and Diffuser-generated synthetic data offers a powerful and scalable solution for melanoma detection from dermoscopic images. The approach effectively mitigates class imbalance, enhances the model’s robustness, and ultimately supports more accurate and sensitive classification, all of which are crucial in the design of reliable AI-assisted diagnostic tools.

8.2 Transformer-Based Approach to Melanoma Detection: Additional Findings

8.2.1 Methodology

The methodology adopted in this study is centred on the use of ViTs as the core architecture for dermoscopic image classification. Unlike traditional convolutional networks, ViTs are capable of capturing long-range spatial dependencies, which is particularly valuable in analyzing complex medical images. Before feeding the images into the model, a sequence of pre-processing steps was applied to standardize and augment the data. Subsequently, the model was trained in an end-to-end fashion, from raw input to class prediction. The entire workflow is summarized in Figure 28 and detailed in the following sections.

8.2.2 Pre-processing and Model Pipeline

The dataset was used as provided in the original ISIC 2017 Challenge, maintaining the same splits for training, validation, and test sets. Each image initially entered the pipeline in its raw format and underwent resizing to a resolution of 384×384 pixels, which was chosen as compatible with the input requirements of the ViT architecture. During training, a random resized cropping operation was applied to the images to introduce variability and reduce overfitting. In addition to cropping, the RandAugment framework was employed to further enhance data diversity, applying randomized transformations such as colour jitter, contrast changes, and geometric distortions. These augmentation steps were exclusively applied to the training data, leaving the validation and test sets unaltered. Once pre-processed, each image was divided into non-overlapping patches of 32×32 pixels. Each patch was then flattened and transformed into a token embedding, which was enriched with positional encoding to preserve spatial relationships. These encoded tokens were concatenated to form the input sequence for the Transformer. The embedded sequence was normalized and passed through the Transformer Encoder, where multi-head self-attention mechanisms operated in parallel to extract complex inter-patch dependencies. The output of the attention layers was combined with the original input through residual connections and subjected to further normalization. An MLP, composed of fully connected layers with GELU activation, processed the resulting representation. The final layer, a fully connected classification head, produced the model's output, assigning each image to one of the target classes.

8.2.3 Multi-Head Attention Mechanism

At the core of the ViT lies the multi-head attention mechanism, which enables the model to weigh the importance of different regions of an image when forming its internal representation. This mechanism is inspired by the original Transformer design developed for Natural Language Processing tasks, where the contextual relationship between words is fundamental. In the context of vision, the same principle allows the model to attend to different parts of the image simultaneously and interpret their relationships. Each input token is projected into three vectors: query, key, and value. The dot product between the query and key vectors determines the level of attention that a particular region should receive. This score is scaled by the dimensionality of the key vectors and passed through a softmax function to generate attention weights. These weights are then used to compute a weighted sum of the value vectors, producing the final output of the attention layer. The

process is replicated across multiple heads, with each head learning to focus on different aspects of the input. This ensemble of attention pathways enables the model to develop a rich and diversified understanding of the image content.

8.2.4 Experimental Setup

To determine the most effective configuration, the experimental process was structured into two primary phases. The first phase aimed to identify the optimal learning rate by evaluating the model performance across a range of values. Starting from a relatively high learning rate of 1.5×10^{-5} , the value was progressively decreased until 5×10^{-7} , monitoring accuracy on the validation set at each iteration. The second phase consisted of an ablation study, which examined the relationship between network depth and model performance. This analysis was carried out to understand how increasing the number of Transformer layers influenced classification accuracy and to identify the appropriate balance between expressiveness and computational efficiency. The combination of these two phases allowed for a rigorous assessment of hyperparameters and contributed to optimizing the final model.

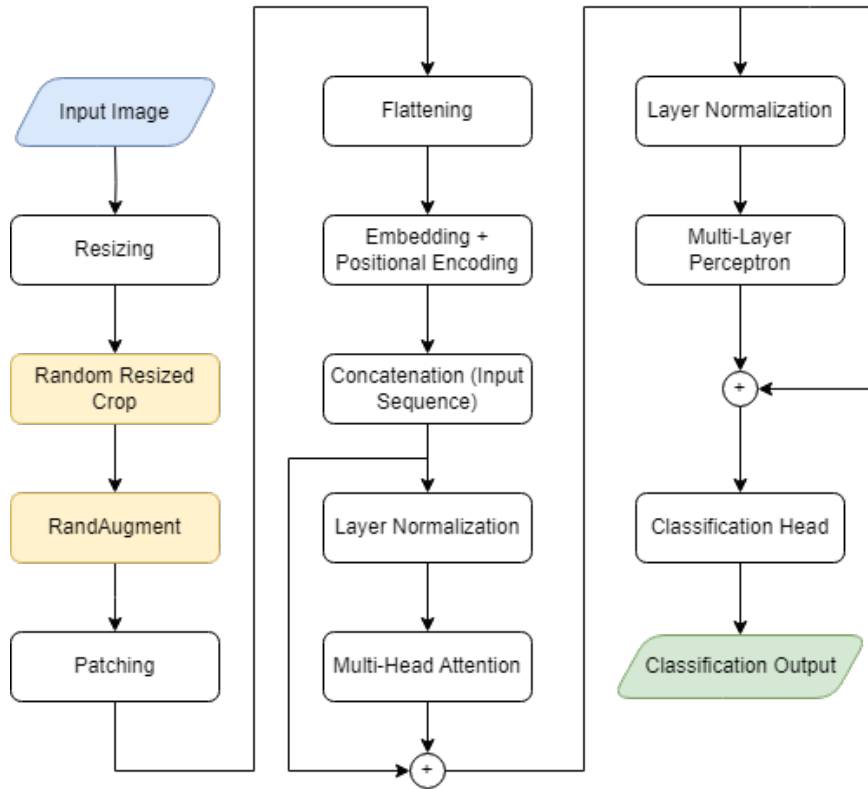


Figure 28: The implemented end-to-end workflow.

8.2.5 Experimental Results

All experiments were carried out using the Google Colab platform, leveraging its GPU acceleration and seamless integration with the pre-trained models of Hugging Face. The model architecture chosen for this study was ViT-Large, which contains approximately 307 million parameters. This model operates on input images of resolution 384×384 and processes them using a patch size of 32×32 pixels.

The ViT-Large model was initialised using weights pre-trained on the ImageNet dataset [129] and subsequently fine-tuned on the ISIC 2017 skin lesion dataset for 40 epochs. The training process adopted the hyperparameter configuration detailed in Table 6. To thoroughly evaluate the architecture, a series of experiments were designed, each varying in learning rate and model depth. These experimental setups are reported in Table 7. A linear learning rate scheduler with warm-up was employed during training. This strategy involves a gradual increase in the learning rate from a low initial value up to a specified maximum, followed by a linear decay as training progresses. The goal is to allow the optimizer to initially explore the loss landscape without overshooting minima while stabilizing convergence in the final stages. Different learning rate targets were tested: 0.5×10^{-6} , 0.5×10^{-5} , 1.0×10^{-5} , and 1.5×10^{-5} . All experiments were conducted while fixing the number of layers at 24.

The observed training dynamics indicate that excessively high learning rates led to unstable convergence, as the model struggled to minimize the loss function effectively. Conversely, very small learning rates achieved stable convergence but required a larger number of epochs and posed a risk of getting trapped in local minima. Among the configurations yielding comparable performance (i.e., 94.8% accuracy), the intermediate learning rate of 1.0×10^{-5} was selected as the optimal choice, offering a good compromise between convergence speed and generalization.

Table 6: Model hyperparameters

Hyperparameter	Value
Number of layers	24
Number of attention heads	16
Hidden size	1024
Optimizer	Adam ($\beta_1 = 0.99$, $\beta_2 = 0.999$, $\epsilon = 10^{-8}$)
Learning rate	1.0×10^{-5}
Scheduler	Linear with warm-up (10% of total epochs)
Batch size	16
Gradient accumulation steps	4

The list of tested configurations, including variations in learning rate and number of Transformer layers, is shown in Table 7. While the learning rate proved influential, the number of layers had a significantly more pronounced impact on model accuracy.

To further investigate the effect of model depth, additional experiments were conducted by varying the number of Transformer layers from 2 to 28. The relationship between layer count and classification accuracy is plotted in Figure 29. As expected, performance generally improved with deeper models, but the gain diminished beyond 24 layers. A performance plateau was observed in the range of 24–26 layers, while configurations with more than 26 layers led to a slight degradation, possibly due to overfitting or gradient instability. In light of these findings, and adhering to the principle of model parsimony, a depth of 24 layers was selected as the best trade-off between complexity and performance.

The training dynamics of the selected configuration (#2) are depicted in Figure 30, which presents the training and validation loss curves. A sharp decrease in both curves occurs around epoch 5, aligning with the end of the learning rate warm-up phase. The periodic spikes observed thereafter are likely attributable to the relatively small batch size

Table 7: Tested configurations and corresponding accuracy.

Config	Batch Size	Learning Rate	Layers	Accuracy
#1	16	1.5×10^{-5}	24	0.948
#2 (Selected)	16	1.0×10^{-5}	24	0.948
#3	16	0.5×10^{-5}	24	0.948
#4	16	0.5×10^{-6}	24	0.945
#5	16	1.0×10^{-5}	20	0.925
#6	16	1.0×10^{-5}	16	0.902
#7	16	1.0×10^{-5}	12	0.850
#8	16	1.0×10^{-5}	10	0.817
#9	16	1.0×10^{-5}	8	0.788
#10	16	1.0×10^{-5}	4	0.707
#11	16	1.0×10^{-5}	2	0.655

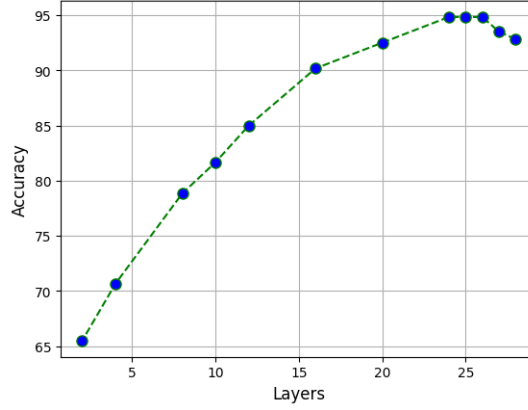


Figure 29: Accuracy trend as a function of the number of Transformer layers.

used during training, which can amplify gradient fluctuations.

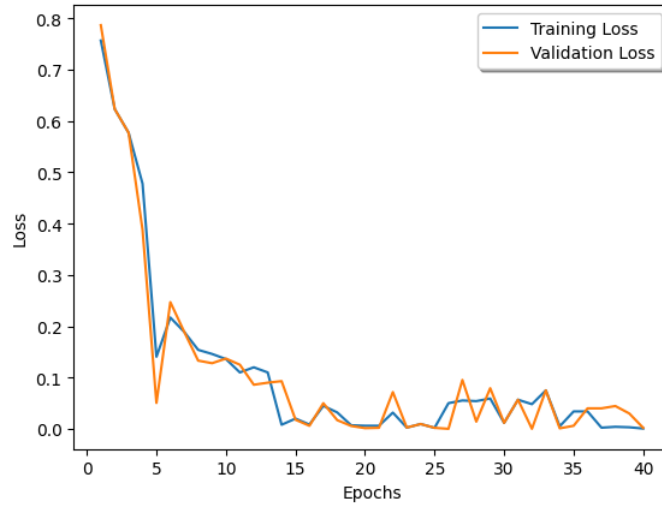


Figure 30: Loss curves for training and validation (Configuration #2 – Best model).

Figure [31](#) shows the confusion matrix corresponding to the best-performing configu-

ration. The matrix reveals a strong diagonal pattern, indicating that most test samples were correctly classified. Misclassifications were relatively rare, suggesting that the model was able to learn discriminative features effectively across the three lesion categories.

True label	Melanoma	101	11	6
	Nevus	5	382	5
	Seborrheic_keratosis	3	1	86
		Melanoma	Nevus	Seborrheic_keratosis
		Predicted label		

Figure 31: Confusion matrix for Configuration #2 (ViT-Large, best setting).

The experimental results confirm the effectiveness of Vision Transformer architectures in classifying dermoscopic images when supported by a careful configuration of hyperparameters. The analysis of layer depth, learning rate, and convergence behaviour has demonstrated that model performance is highly sensitive to architectural tuning. With appropriate training settings, ViTs exhibit robust performance and generalization, even on moderately sized datasets such as ISIC 2017, reinforcing their suitability for computer-aided dermatological diagnosis.

8.2.6 Discussion

In recent years, a growing number of studies have focused on enhancing the diagnostic process for skin lesions using advanced DL techniques, with particular attention to melanoma detection. In this context, numerous approaches have been proposed that exploit attention mechanisms and Transformer-based architectures. For instance, U-Net variants and attention-based segmentation networks such as TransUNet have demonstrated superior performance in isolating lesion boundaries, achieving high accuracy and segmentation quality [130]. Similarly, Fully Transformer Networks (FTNs) have been employed to extract multi-scale spatial information, with promising results on the ISIC datasets [131]. Other contributions have leveraged dual-stage attention mechanisms or hierarchical vision structures to enhance skin lesion classification and segmentation, showing strong potential for deployment in clinical settings [132, 133]. Furthermore, recent efforts have successfully employed ViT architectures for melanoma classification by incorporating medical-specific training strategies and data augmentation methods, confirming the relevance of self-attention in the medical imaging domain [134].

Despite these advancements, the full potential of Vision Transformers for melanoma classification remains underexplored. The current study aimed to address this gap by introducing a ViT-based model tailored for dermoscopic image classification, designed to differentiate among MM, NCN, and SK. The experiments were carried out using the publicly available ISIC 2017 dataset, allowing a direct comparison with existing literature using the same benchmark. The performance of the proposed model, particularly in terms of classification accuracy and sensitivity, significantly surpasses that of previously

published approaches (see Table 8). These gains can be attributed to the model’s inherent ability to model long-range dependencies and spatial relations in the image through the self-attention mechanism, which plays a pivotal role in capturing global contextual patterns.

Table 8: Model Test Performance on the ISIC 2017 dataset.

Model Name	Accuracy	Sensitivity	Specificity	AUROC
IRv2+Soft Attention [135]	0.904	0.916	0.833	0.959
ARL-CNN50 [136]	0.868	0.878	0.867	0.958
SEnet50 [137]	0.863	0.856	0.865	0.952
RAN50 [138]	0.862	0.878	0.859	0.942
ResNet50 [139]	0.842	0.867	0.837	0.948
Proposed ViT-based approach	0.948	0.928	0.967	0.948

It is worth emphasizing that, despite the number of lesion classes being fixed by the ISIC challenge (i.e., MM, NCN, SK), the model was able to maintain high classification performance across all categories. The accuracy of 94.8%, together with a sensitivity of 92.8% and a specificity of 96.7%, demonstrates the effectiveness of the model in both detecting positive cases and reducing false alarms. Although the AUROC value remains slightly lower than some alternatives, the overall trade-off favours the proposed architecture, especially in terms of real-world clinical usability.

The role of hyperparameter tuning emerged as particularly relevant. A detailed analysis of model depth revealed that increasing the number of layers improves classification accuracy up to a saturation point, beyond which further depth may introduce unnecessary complexity without performance gains. Interestingly, even with reduced architectural depth (e.g., 20 layers), the model retained performance levels above the state-of-the-art, which is encouraging for low-resource deployment scenarios. This is especially relevant in mobile and embedded health monitoring platforms, where computational efficiency is critical. Moreover, the study explored the impact of learning rate scheduling and optimizer dynamics. The use of a linear warm-up schedule proved beneficial in preventing early instability during the initial training stages. Configurations employing a balanced learning rate showed better convergence behaviour and generalization, while extreme values led to either underfitting or instability. The results of the ablation study demonstrate the sensitivity of Transformer models to these design choices. From an architectural standpoint, the ViT’s ability to capture cross-patch dependencies through multi-head attention mechanisms enhances its capacity to identify both local features and global structures. This is particularly useful in dermatoscopic images, where lesion asymmetry, border irregularity, and colour variation are important visual cues for diagnosis. Furthermore, attention-based visualization (e.g., attention maps) provides a degree of interpretability, enabling clinical users to verify that the network is focusing on medically relevant regions of the image. Despite these promising results, training ViTs on relatively small datasets such as ISIC 2017 still presents challenges. The observed fluctuations in training and validation loss curves, particularly in early epochs, suggest a sensitivity to small batch sizes and data variability. This may be linked to internal covariate shifts and unstable gradients, which could be mitigated through more aggressive regularization strategies or larger-scale pretraining. One possible avenue for future improvement is the integration of data augmentation via generative models, such as GANs or Diffusion Models, which

can help alleviate overfitting and increase training diversity. In conclusion, the ViT-based framework presented in this study strikes a balance between architectural complexity and clinical relevance. It delivers high sensitivity and specificity, key requirements in melanoma screening while maintaining an explainable and scalable design. Further research will explore the incorporation of generative data augmentation, multi-modal data fusion (e.g., patient history, metadata), and transformer variants with reduced complexity for deployment on edge devices in teledermatology scenarios.

8.3 ViT-based Classification of Mammogram Images

8.3.1 Dataset

The dataset employed in this study is the Curated Breast Imaging Subset of the Digital Database for Screening Mammography (CBIS-DDSM) [140], which consists of digitized mammographic images in .png format. This collection focuses on annotated cases with confirmed benign or malignant diagnoses. For the present work, only samples from these two categories were considered.

The dataset was partitioned into training, validation, and testing subsets. The training set initially comprised 930 mammographic images, evenly distributed between benign and malignant cases. For model evaluation on unseen data, the test set included 348 images, with 144 malignant and 204 benign cases. To support model validation during training, a separate validation subset containing 233 samples was extracted from the original training set. This set consisted of 116 malignant and 117 benign images, providing a balanced foundation for tuning and monitoring the model’s learning behaviour.

8.3.2 Image Pre-processing

A dedicated pre-processing pipeline was implemented to standardize and enhance the mammographic images before training. The procedure consisted of manual cropping, threshold-based segmentation, and automatic resizing. Initially, manual cropping was applied to remove visible artefacts such as radiopaque markers, ensuring that the retained regions exclusively captured relevant breast anatomy. This step was crucial to prevent irrelevant structures from interfering with the feature extraction process. Subsequently, Otsu’s thresholding method was adopted to isolate the breast area from the background. This automated segmentation technique emphasized the tissue of interest and mitigated the influence of background noise caused by non-uniform black regions. To finalize the preprocessing phase, all images were automatically cropped and resized to a fixed resolution of 640×640 pixels. Standardizing image dimensions facilitated consistent analysis across the dataset and ensured compatibility with the model input size, while also minimizing variability during training. An overview of the entire pre-processing workflow is depicted in Figure 32.

8.3.3 Geometric Data Augmentation

To improve model generalization and increase the diversity of training samples, classical geometric data augmentation techniques were applied. These transformations aimed to simulate various acquisition conditions and anatomical variations, making the model more resilient to real-world variability. Horizontal and vertical flips were used to introduce variations in breast orientation, while random diagonal rotations in the range of -30° to $+30^\circ$

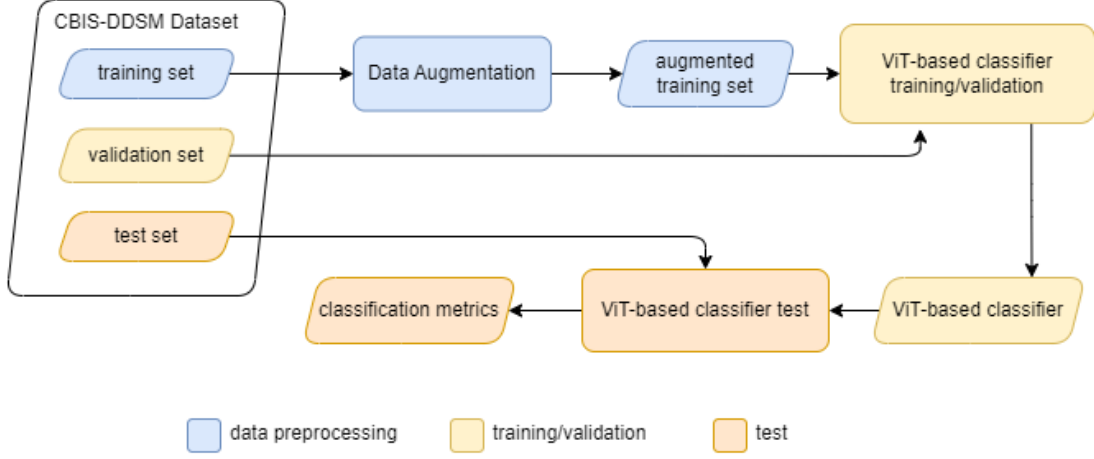


Figure 32: Overview of the image pre-processing pipeline applied to the CBIS-DDSM dataset.

were applied to mimic typical angular shifts occurring during mammographic screening. These transformations allowed the model to learn rotational invariance in lesion localization. Additionally, exposure adjustments were introduced to replicate differences in brightness and contrast that may occur across imaging devices or acquisition protocols. The resulting variations helped the model adapt to different lighting conditions. Gaussian noise with a mean 0 and standard deviation 0.1 was added to mimic sensor-related noise, encouraging the model to focus on robust features rather than image artefacts. Contrast enhancement and CLAHE were also applied to improve local contrast and highlight subtle patterns in the tissue. Further augmentations included saturation manipulation to explore colour distribution variability and Gaussian blurring to reduce high-frequency noise, enhancing the model’s focus on larger structures. After applying these transformations, the dataset was expanded by generating an additional 8373 samples, 4188 benign and 4185 malignant, resulting in a more balanced and representative training set. This augmentation strategy proved essential to boost the model’s robustness and reduce overfitting.

8.3.4 Diffusion-Based Augmentation

In addition to classical augmentations, a generative approach was employed using DDPMs. Two separate DDPMs were trained: one on benign samples and one on malignant samples, each for 200 epochs with a learning rate of 1×10^{-4} . The goal was to generate synthetic images that would enrich the dataset while preserving the underlying data distribution of each class.

At the end of the training phase, each DDPM produced 2000 new images, yielding a total of 4000 synthetic samples, 2000 for each class. These images were subsequently incorporated into the training set, further improving the dataset’s size and balance. The use of Diffusion Models offered a powerful means of generating highly realistic images, which effectively enhanced the training diversity and allowed for better representation of rare or under-sampled lesion types.

8.3.5 Experimental Setup

To systematically assess the impact of each data augmentation strategy, four experimental configurations were evaluated:

1. A baseline Vision Transformer (ViT) model, pre-trained on the ImageNet dataset [141], was fine-tuned for 100 epochs using the original CBIS-DDSM training set.
2. The same ViT model was fine-tuned for 50 epochs using a training set enhanced by geometric data augmentation techniques.
3. Another experiment was conducted where the model was fine-tuned for 50 epochs using a training set augmented with Diffuser-generated synthetic images.
4. A final configuration combined both augmentation methods, with the ViT model trained for 50 epochs on a dataset that included original, geometrically augmented, and Diffuser-generated images.

In all cases, the ViT-Base model architecture was adopted, with its structure and hyperparameters summarized in Table 9.

Table 9: Vision Transformer (ViT-Base) hyperparameters.

Hyperparameter	Value
Number of layers	12
Number of attention heads	12
Hidden dimension	768
Input image resolution	224×224
Patch size	16×16
Optimizer	Adam ($\beta_1 = 0.99$, $\beta_2 = 0.999$, $\epsilon = 10^{-8}$)
Learning rate	1×10^{-5}
Scheduler	Linear decay with 10% warm-up
Batch size	16
Gradient accumulation steps	4

8.3.6 Results and Discussion on Mammography Classification

The experimental evaluation aimed to assess the effectiveness of different data augmentation (DA) strategies in improving the classification performance on the CBIS-DDSM dataset. The impact of each augmentation configuration was analyzed in terms of classification metrics, training dynamics, and model interpretability.

Figure 33 displays the confusion matrices for the test set under four experimental conditions: no augmentation, geometric DA only, Diffuser-based DA only, and the combination of both. These visualizations allow for a qualitative inspection of the model’s behaviour across the benign and malignant categories.

Geometric augmentation improves the model’s sensitivity to benign cases, although with a modest reduction in the ability to detect malignancies. This may indicate a slight shift in the decision boundary, which tends to favour the dominant class in the augmented

data. This trade-off suggests that while geometrically augmented images increase exposure to structural variability, they might not sufficiently address class imbalance or the intrinsic complexity of malignant patterns. In contrast, the use of synthetic images generated through a DDPM significantly enhanced the classifier’s ability to detect malignant lesions. The generated samples are not simply transformations of existing data but represent novel patterns synthesized from learned distributions, thus introducing truly new information into the training process.

The combined use of both augmentation strategies further reinforced this trend, leading to a more balanced performance across classes. This hybrid approach seems to unite the strengths of both methods: the variability introduced by geometric transformations and the semantic richness of the synthetic samples.

This tendency is quantitatively confirmed by the results reported in Table 10, where accuracy, sensitivity, and specificity show a consistent improvement when augmentation techniques are applied. Notably, the highest sensitivity (88.89%) is achieved in the combined augmentation setting, indicating the model’s improved capability to correctly identify malignant cases, which are of critical importance in clinical diagnosis. The specificity also benefits from this approach, rising from 64.58% (no augmentation) to 68.63% in the combined setup, indicating better control over false positives.

Moreover, the consistent increase in overall accuracy from 65.52% to 77.01% across experiments suggests that ViTs can fully exploit the increased variety and density of training examples. This confirms the importance of data-centric strategies in medical AI, where model improvements are often constrained by dataset limitations.

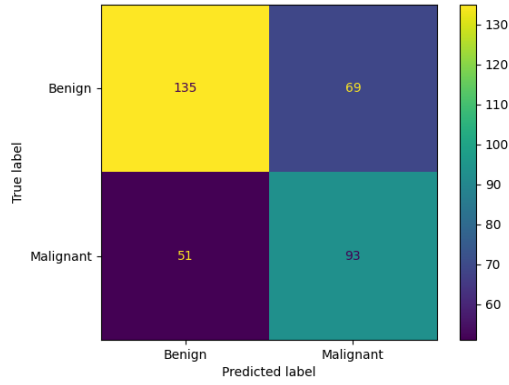
Table 10: Classification results on the test set with different data augmentation techniques.

Approach	Accuracy	Sensitivity	Specificity
no augmentation	65.52%	66.18%	64.58%
geometric DA	69.25%	75.98%	59.72%
diffuser-based DA	74.42%	88.19%	64.70%
geometric + diffuser-based DA	77.01%	88.89%	68.63%

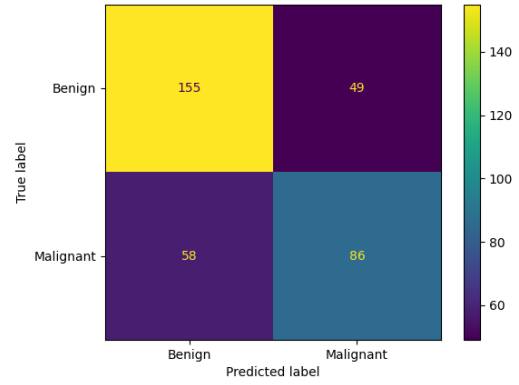
The evolution of the training and validation losses, plotted in Figure 34, further supports these findings. The curves reveal that the use of Diffuser-based images contributes to stabilizing the training process, as indicated by the narrowing gap between training and validation losses. This implies a reduction in overfitting and an increased ability of the model to generalize to unseen data. Moreover, the combination of classical and synthetic augmentation led to smoother loss trajectories, suggesting that such data diversity supports a more regular and robust optimization path.

Notably, the stabilization effect observed with Diffuser augmentation indicates that the generated samples do not introduce noise or ambiguity, but rather enhance the signal space of the learning problem. This is particularly relevant in high-stakes applications such as cancer detection, where generalization on rare and difficult samples is essential.

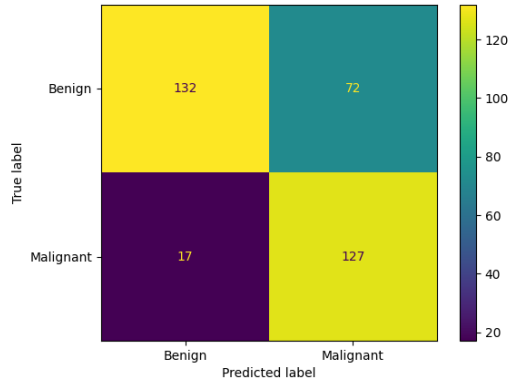
When comparing the proposed method with existing literature on the same dataset (see Table 11), the ViT-based model with combined augmentation outperforms various CNN-based approaches such as max pooling, average pooling, and multiple instance learning. This is a notable result, especially considering that the ViT architecture used is relatively shallow compared to large-scale convolutional backbones, highlighting the ef-



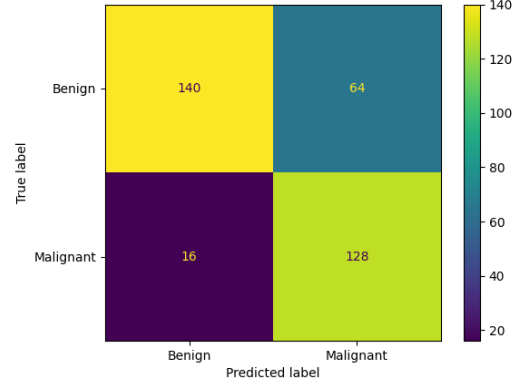
(a) original dataset (no augmentation)



(b) geometric DA

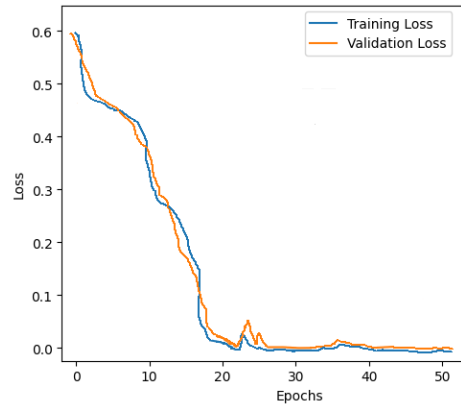


(c) diffuser-based DA

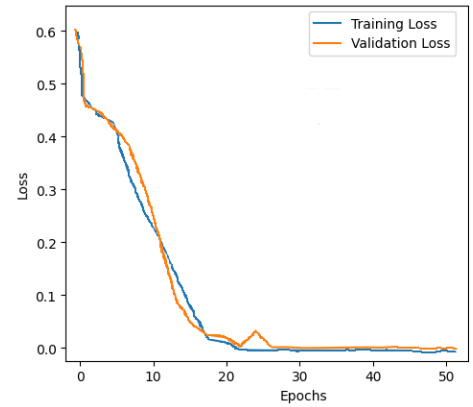


(d) geometric + diffuser-based DA

Figure 33: Confusion matrices on the test set for different augmentation strategies.



(a) training and validation losses (diffuser-based DA)



(b) training and validation losses (geometric + diffuser-based DA)

Figure 34: Loss curves during training for models using Diffuser-based augmentation.

fectiveness of Transformers in limited data regimes when supported by adequate data augmentation.

The interpretability of the model was also evaluated through attention map visualiza-

Table 11: Accuracy comparison with literature approaches on CBIS-DDSM dataset.

Approach	Accuracy
deep multiple instance learning [142]	74.2%
max pooling [143]	67.5%
average pooling [143]	70.3%
region-based group-max pooling [143]	76.2%
global group-max pooling [143]	76.7%
proposed approach	77.01%

tion. Figure 35 presents an example of a ViT attention map over a malignant lesion. These maps reflect the inner attention mechanisms of the network and highlight regions deemed most relevant during inference. Such interpretability tools are crucial for real-world clinical adoption, as they enable physicians to validate AI decisions and trust automated support tools. In particular, the ability to localize abnormal regions aligns with the standard visual analysis workflow of radiologists and provides a bridge between black-box AI outputs and explainable decisions.

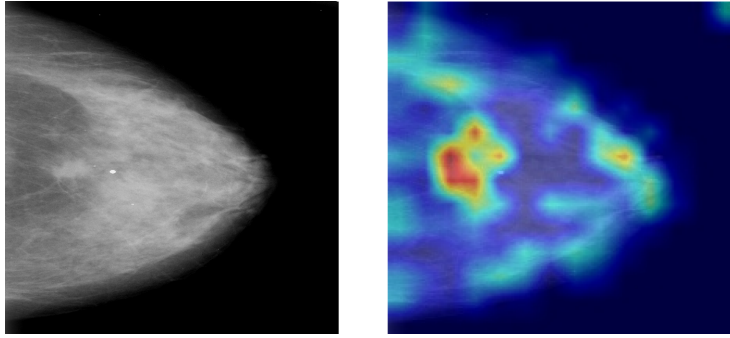


Figure 35: Example of attention map highlighting a malignant lesion.

In summary, this study demonstrates that the integration of geometric and diffusion-based augmentation strategies, when paired with a Vision Transformer architecture leads to significant improvements in classification performance and model reliability. The use of synthetic images generated by Diffusion Models emerges as a particularly effective strategy to overcome class imbalance and data scarcity while maintaining clinical relevance and interpretability. These results support the use of hybrid augmentation strategies in future diagnostic pipelines and encourage further exploration of generative models in radiology-focused AI research.

9 Graph-Based and Explainable Approaches to Traumatic Brain Injury Classification

9.1 Introduction

Traumatic brain injury (TBI) constitutes a major global health concern, affecting over 60 million people annually and representing one of the primary causes of death and long-term disability in individuals under 45 years of age [144, 145]. The multifaceted nature of TBI, encompassing a wide spectrum of causes, symptoms, and recovery outcomes, complicates both diagnosis and treatment. In particular, the clinical assessment of patients with prolonged disorders of consciousness, such as those in minimally conscious (MCS) or vegetative (VS) states, presents significant challenges due to the heterogeneity of brain damage and the limitations of standard neuroimaging techniques [146, 147].

Diffusion-weighted imaging (DWI) has emerged as a pivotal tool for investigating the microstructural alterations in white matter following traumatic injury. By reconstructing the structural connectome of the brain through tractography, it is possible to derive graph representations where nodes correspond to anatomical regions and edges reflect the strength of structural connections. These connectomes allow the extraction of complex network measures that may capture disconnection patterns or compensatory mechanisms involved in the neuropathology of TBI [148]. To exploit these graph-based features, recent research has increasingly integrated ML methodologies to classify TBI patients and predict clinical outcomes. Classical ML algorithms, such as Random Forest, Support Vector Machine, Decision Trees, and Logistic Regression, have proven useful in identifying group differences and in stratifying patients based on the severity of injury or recovery potential [149]. A key limitation, however, lies in the interpretability of these models, particularly in the medical domain, where explainability is essential for clinical validation and decision-making. In response to this issue, explainable artificial intelligence (XAI) techniques have been introduced to provide insight into the decision process of predictive models. Among these, SHAP (SHapley Additive exPlanations) has gained recognition for its ability to assign interpretable importance scores to input features, enabling clinicians to understand which aspects of brain connectivity are most influential in driving a model's output [106]. Such interpretability is not only beneficial for model transparency but also critical for fostering trust in AI-driven diagnostic tools. Alongside classical ML pipelines, graph neural networks, and in particular, Graph Convolutional Networks (GCNs), have gained traction for their ability to process non-Euclidean data, such as brain graphs. By learning the topology of the brain's structural connectome, GCNs can identify complex spatial patterns associated with disease states and functional impairments [150, 151]. Their use has been demonstrated in various neurological contexts, including dementia, Parkinson's disease, and autism spectrum disorder [152, 153, 154]. However, as with other deep learning models, GCNs also suffer from limited interpretability, which restricts their clinical applicability.

To overcome this limitation, recent studies have explored the adaptation of visualization techniques such as Grad-CAM (Gradient-weighted Class Activation Mapping) to the graph domain, enabling the identification of the most relevant nodes, i.e., brain regions, contributing to a classification decision [155, 156]. This integration of GCNs with XAI tools not only enhances model transparency but also provides neurobiologically meaningful insights into which brain structures may be implicated in consciousness recovery or persistent impairment following TBI [157, 158]. The present chapter introduces two

complementary approaches to TBI classification and interpretation. The first relies on a traditional ML framework applied to graph-derived features extracted from DWI data, using SHAP to analyze feature importance and enhance interpretability. The second builds on a deep learning framework based on GCNs, enriched with Grad-CAM to visualize the node-level contributions in the brain graph space. Both methods aim to improve the diagnostic and prognostic assessment of TBI, especially in the identification of patients at risk of prolonged consciousness disorders. Despite the relatively small size and variability of the datasets, a common issue in TBI research due to the rarity and heterogeneity of severe cases, the findings presented in this chapter contribute to the growing body of work seeking to develop interpretable, data-driven tools for precision medicine in neurotrauma care. Through the integration of connectivity-based modelling and explainable AI, these approaches strive to bridge the gap between high-performance algorithms and the clinical demand for transparency, trust, and actionable insight.

9.2 Dataset

The dataset employed in this study includes DWI scans acquired from a total of 49 individuals, categorized into three major groups: patients in the acute phase following TBI, individuals with chronic disorders of consciousness (DoC), and a control group of healthy subjects. All DWI data were obtained using a Siemens Skyra 3T scanner equipped with a 32-channel head coil. The imaging protocol followed high angular resolution diffusion imaging (HARDI) standards, featuring 60 diffusion directions with a b-value of 2000 s/mm² and 10 additional non-diffusion-weighted (b0) volumes. The resulting images had an isotropic resolution of 2 mm, ensuring an accurate depiction of brain tissue microstructure and enabling the reconstruction of reliable structural connectomes [159, 160]. The acute group consisted of 18 patients who had sustained severe TBI. Selection criteria were based on clinical severity, including a Glasgow Coma Scale (GCS) score of 6 or less and the absence of eye-opening for a minimum of 24 hours. This ensured the inclusion of individuals with significant disturbances in consciousness. Although two patients in this group did not survive, their imaging data were retained for analysis to provide insight into the early structural alterations associated with critical TBI. Of the surviving patients, nine underwent a second DWI scan approximately five months after the initial acquisition, allowing for longitudinal assessment of changes in brain connectivity associated with recovery of consciousness. The chronic group was composed of individuals examined at least five months after the traumatic event. It included six patients diagnosed with long-standing disorders of consciousness, such as VS and MCS, as well as the nine previously mentioned patients from the acute group who had recovered consciousness and participated in the follow-up imaging session. Clinical evaluation of the DoC patients was performed using the Coma Recovery Scale-Revised (CRS-R), which helped to characterize the severity and nature of consciousness impairment. Patients in VS exhibited wakefulness without signs of awareness, while those in MCS demonstrated intermittent intentional behaviours, such as visual tracking or response to stimuli, though without reliable communication. Within the MCS subgroup, patients were further distinguished into MCS+ or MCS- based on the complexity of their behavioural responses [161, 162, 163, 164]. To provide a baseline for comparison, a control group of 16 healthy volunteers was also included in the dataset. These individuals were matched to the patient groups in terms of age and sex and exhibited no known neurological or psychiatric conditions. Their inclusion enabled the identification of deviations in brain network architecture attributable to traumatic injury,

by serving as a normative benchmark in both cross-sectional and longitudinal analyses.

9.3 Image Preprocessing and Connectome Construction

while offering valuable insights into the brain’s microstructural organization, DWI is inherently susceptible to several types of artefacts that can compromise the fidelity of the acquired data. In the present work, a rigorous preprocessing pipeline was implemented to mitigate such imperfections and prepare the data for subsequent analysis. The operations were carried out using established neuroimaging software packages, including FSL, MRtrix3, and DIPY [165, 166, 167].

Due to the absence of high-resolution anatomical references such as T1- or T2-weighted images, the b0 volume extracted from the DWI acquisition was used as the anatomical reference throughout the pipeline. The first stage of preprocessing involved the removal of non-brain tissues through FSL’s Brain Extraction Tool (BET2), which enabled more accurate isolation of cerebral structures [168]. The data were then corrected for susceptibility-induced distortions and eddy currents, which are common in diffusion imaging due to the rapid switching of magnetic gradients [169, 170]. Gibbs ringing artefacts, stemming from the image reconstruction process, were also addressed to improve spatial clarity [171]. To enhance signal quality, denoising based on the multivariate principal component analysis (MP-PCA) method was applied, effectively suppressing thermal noise while preserving structural detail [172]. At the final stages of this preprocessing phase, bias field inhomogeneities were corrected using the N4 algorithm [173], and all images were resampled to an isotropic resolution of 1.5 mm³ [174]. Spatial normalization was conducted by registering each subject’s preprocessed DWI to the MNI152 template using the nonlinear registration framework available in MRtrix3. This ensured precise anatomical alignment across the cohort. Anatomical parcellation was based on the AAL3v1 atlas [175], which segments the brain into 170 regions. These served as the nodes in the structural connectome. For a subset of subjects with marked anatomical deviations, additional affine transformations were applied to better align atlas regions to individual brain anatomies, ensuring reliable node placement and improved consistency in network construction. Notably, specific anatomical structures, such as certain subdivisions of the thalamus and the anterior cingulate cortex, were not included in the final connectivity model due to limitations in their representation within the AAL3v1 atlas. Their exclusion occasionally resulted in isolated nodes, which were analyzed independently to assess the potential functional significance of disconnection patterns. Following preprocessing and registration, the structural connectivity matrices were generated via probabilistic tractography. In the absence of T1-weighted images, tissue segmentation was carried out directly on the b0 volume using the 5TTgen-FSL workflow from MRtrix3, providing a multi-tissue map to distinguish white matter, grey matter, and cerebrospinal fluid boundaries. Manual inspection and correction of segmentation masks were performed to ensure accurate anatomical representation. Fibre orientation distributions (FODs) were estimated using the single-shell three-tissue constrained spherical deconvolution (SS3T-CSD) method [176, 177], which is well suited for single-shell DWI data. These FODs were then intensity-normalized to account for inter-voxel variability. Anatomically Constrained Tractography (ACT) [178] was applied to restrict streamline generation to biologically plausible paths, leveraging tissue maps to inform the propagation of trajectories. Seeding was performed along the white-grey matter interface, with the tractography algorithm configured to enforce biologically relevant constraints. For instance, a maximum streamline length of 250 mm and a minimum FOD

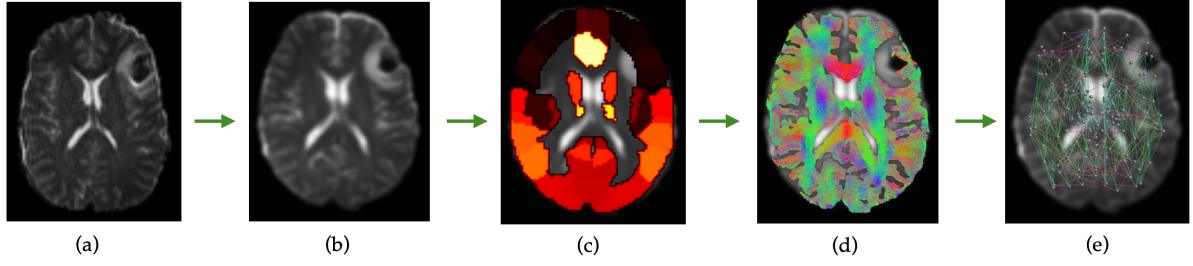


Figure 36: The comprehensive schematic delineates the procedural continuum of the enacted processing cascade: (a) unaltered original DWI, (b) preprocessed DWI, (c) DWI superimposed with the atlas, (d) tractography, and (e) connectome.

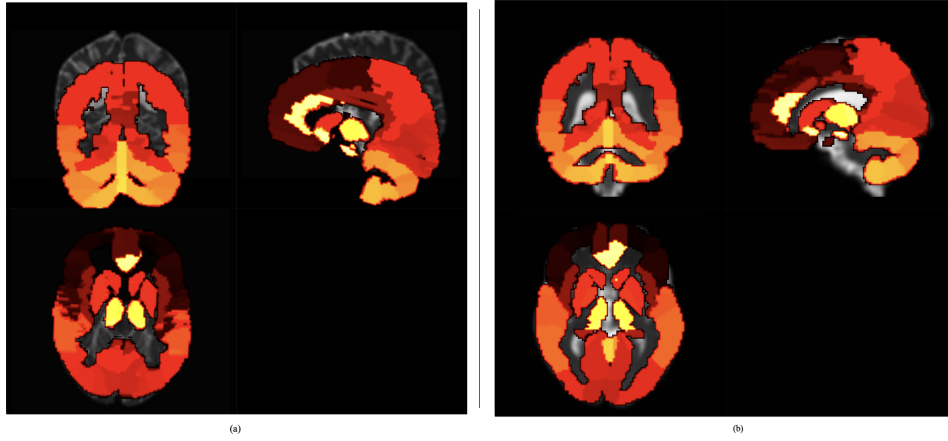


Figure 37: Panel (a): brain image before atlas alignment. Panel (b): post-registration alignment with the AAL3v1 atlas. Colours indicate distinct anatomical regions used for graph construction.

amplitude threshold of 0.06 were imposed to discard spurious trajectories. Additionally, backtracking was enabled to refine streamlines that prematurely terminated, improving anatomical plausibility. In total, 10 million streamlines were generated for each subject, allowing for a dense and statistically robust reconstruction of white matter pathways. To refine the tractograms, the SIFT2 (Spherical-deconvolution Informed Filtering of Tractograms) method was employed [179]. This technique reweights each streamline based on its alignment with the underlying FOD signal, mitigating common biases and ensuring that the resulting connectivity matrices more accurately reflect the actual anatomical structure of the brain. The outcome of this process was a weighted adjacency matrix of size 170×170 , where each element corresponds to the number of streamlines connecting a pair of brain regions. These matrices provided a quantitative representation of the brain's structural network for each subject. Zero-valued entries in these matrices indicated a lack of detected connections between specific regions. Notably, in patients with long-standing disorders of consciousness, a higher frequency of isolated nodes was observed, potentially reflecting profound disruptions in structural communication pathways.

Figure 36 visualizes the core stages of this pipeline, from raw DWI data to the final connectome. Instead Figure 37 illustrates the brain atlas before and after the registration process, highlighting the alignment achieved and ensuring that the structural connectivity accurately reflects the underlying neuroanatomical structures

9.4 The Brain as a Graph: Formal Definition and Theoretical Rationale

The human brain can be conceptually and computationally represented as a network, where complex patterns of connectivity between different regions give rise to cognitive functions, consciousness, and behavior. This conceptualization aligns with the emerging field of network neuroscience, which interprets the brain as a system of interconnected nodes linked by physical or functional pathways [180, 181]. From a structural point of view, diffusion-weighted imaging (DWI) allows the inference of white matter tracts that physically connect grey matter regions. This enables the derivation of subject-specific brain networks, known as *connectomes*, where each connection represents the presence and strength of neural communication routes. Translating this biological structure into a mathematical model allows researchers to apply formal tools from graph theory to study and quantify brain topology. Mathematically, a brain network can be described as a graph $G = (V, E)$, where V is the set of nodes (or vertices), each corresponding to a distinct anatomical brain region, and $E \subseteq V \times V$ is the set of edges that represent structural connections between pairs of regions. If two regions are connected via white matter streamlines, estimated through tractography algorithms, then an edge exists between them.

In this study, we define an undirected weighted graph, where the connections between regions are not directional, and each edge has an associated weight. This weight quantifies the strength of the connection, typically measured by the number of streamlines linking the two regions. The graph is thus represented by an adjacency matrix $\mathbf{A} \in \mathbb{R}^{N \times N}$, where N is the number of regions defined in the atlas (in our case, 170 regions from the AAL3v1 atlas). The entry A_{ij} is equal to the number of streamlines connecting node i and node j . For undirected graphs, the matrix is symmetric: $A_{ij} = A_{ji}$, and diagonal elements A_{ii} are conventionally set to zero to exclude self-loops.

The degree of a node v_i , denoted d_i , is the total weight of its incident edges:

$$d_i = \sum_{j=1}^N A_{ij} \quad (4)$$

This quantity reflects the total connectivity of a region and can be used to assess its relative importance or centrality in the network.

Other relevant matrices include the degree matrix $\mathbf{D}$, a diagonal matrix where $D_{ii} = d_i$, and the graph Laplacian:

$$\mathbf{L} = \mathbf{D} - \mathbf{A} \quad (5)$$

The Laplacian plays a fundamental role in spectral graph theory and is instrumental in clustering, diffusion modeling, and graph-based ML methods.

Modelling the brain as a graph provides a unifying framework to assess both local and global properties of brain organization. On a local scale, features such as degree, clustering coefficient, or betweenness centrality allow the identification of hubs or vulnerable nodes. On a global scale, measures like global efficiency, characteristic path length, and modularity enable comparisons across subjects or clinical groups [182]. Furthermore, this abstraction is not merely theoretical, it allows the integration of advanced computational models, including graph neural networks (GNNs), which operate directly on graph-structured data to extract latent representations of brain topology [183, 184]. These models are particularly effective in classifying neurological and psychiatric conditions based on structural connectomes.

9.5 Assessment of brain connectivity in traumatic brain injury through Statistical Analysis and Explainable AI Techniques

9.5.1 Exploration of Graph Metrics in Brain Connectomes

To characterize the structural topology of the brain, we analyzed the connectomes derived from DWI as undirected, weighted graphs. In this representation, brain regions are modelled as nodes and white matter pathways as weighted edges reflecting connection strength. This abstraction enables the quantification of both local and global network properties through graph-theoretical measures [185]. These metrics were categorized into four functional domains: segregation, integration, centrality, and resilience.

9.5.1.1 Measures of Segregation Segregation metrics reflect the brain’s ability to support specialized and localized information processing within tightly connected clusters or modules. These modules are thought to underlie functional specificity, and their presence is often associated with efficient intra-regional communication.

The primary indicator is the **Clustering Coefficient (CC)**, which quantifies how connected a node’s neighbours are to each other. For a node i , the **weighted clustering coefficient** is defined as:

$$C_i^w = \frac{1}{k_i(k_i - 1)} \sum_{j,k} (\hat{w}_{ij}\hat{w}_{jk}\hat{w}_{ki})^{1/3} \quad (6)$$

Here, k_i is the degree of node i , and $\hat{w}_{ij}$ is the normalized weight of the edge between nodes i and j .

From this node-level metric, several aggregate measures were derived:

- **AC (Average Clustering)**: the mean of the clustering coefficients across all nodes, computed using binary connections.
- **WAC (Weighted Average Clustering)**: the mean calculated using weighted edges, incorporating connection strengths.
- **GC (Global Clustering)**: a global index capturing the overall level of transitivity in the network.

Higher clustering is typically associated with a modular architecture that promotes efficient, localized computations, an important property for cortical specialization and sensory integration.

9.5.1.2 Measures of Integration While segregation focuses on local processing, integration metrics assess the brain’s capacity to coordinate information across distributed regions. These metrics reflect the efficiency of global communication and are especially relevant to complex cognitive tasks requiring widespread neural collaboration.

A foundational metric is the **Density (D)** of the network, defined as:

$$D = \frac{2|E|}{N(N - 1)} \quad (7)$$

where $|E|$ is the total number of edges and N is the number of nodes in the network.

Complementary measures include:

- **IGD (Inverse Graph Density)**: the reciprocal of density, used to quantify sparsity and the presence of weakly connected structures.

- **LCS (Largest Component Size)**: the size of the largest connected subgraph within the network, which reflects the extent of overall connectivity.

Efficient integration is a hallmark of healthy brain networks. Reduced density or fragmentation, as measured by IGD or diminished LCS, is often linked to pathological conditions such as traumatic brain injury or neurodegeneration.

9.5.1.3 Measures of Centrality Centrality metrics identify the most influential nodes, often referred to as network “hubs,” which play a pivotal role in maintaining communication efficiency and robustness.

The simplest and most commonly used metric is the **Average Degree (AD)**, which represents the average number of direct connections per node:

$$\bar{k} = \frac{1}{N} \sum_{i=1}^N k_i \quad (8)$$

Here, k_i is the degree of node i , and N is the total number of nodes. High-AD nodes may facilitate information flow across distant modules and prevent isolation. Damage to these hubs, especially in traumatic brain injury, may severely impair global connectivity and cognitive functioning.

9.5.1.4 Measures of Resilience Resilience describes the network’s robustness to node removal or edge disruption, offering insights into its fault tolerance. In clinical settings, these metrics are particularly valuable for assessing how brain injury alters the structural integrity of the connectome.

A key resilience metric is **Assortativity**, which measures the tendency of nodes to connect to other nodes with similar degrees. It is calculated as:

$$r = \frac{\sum_{i,j} (k_i - \bar{k})(k_j - \bar{k})}{\sum_{i,j} (k_i - \bar{k})^2} \quad (9)$$

where k_i and k_j are the degrees of nodes connected by an edge, and $\bar{k}$ is the average degree of the graph.

Positive assortativity implies that high-degree nodes preferentially connect to other high-degree nodes, which may protect the network against random failures. Conversely, negative assortativity could indicate vulnerability to targeted damage.

Other resilience indicators include:

- **NCC (Number of Connected Components)**: reflects the fragmentation of the network into isolated clusters.

- **NIN (Number of Isolated Nodes)**: quantifies nodes with no connections, typically indicative of severe structural disconnection.

These metrics are particularly sensitive in identifying alterations associated with disorders of consciousness and the long-term effects of TBI.

9.5.2 SHAP: SHapley Additive exPlanations

SHAP (SHapley Additive exPlanations) is a model-agnostic explainability framework based on cooperative game theory, developed to interpret predictions made by complex ML models [106]. It assigns an importance value to each input feature by quantifying its contribution to the model’s output.

SHAP builds on the concept of Shapley values, introduced by Lloyd Shapley in 1953. In a cooperative game, the Shapley value represents a fair distribution of a payoff among players based on their contributions to the collective outcome. Translated into ML, the “players” are the input features, and the “payout” is the model’s prediction. SHAP evaluates how the prediction changes when a specific feature is added or removed from all possible subsets of features.

Mathematically, given a model f , an input sample $x \in \mathbb{R}^n$, and a set of features $N = \{1, 2, \dots, n\}$, the Shapley value ϕ_i for feature i is defined as:

$$\phi_i(f, x) = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(n - |S| - 1)!}{n!} [f_{S \cup \{i\}}(x) - f_S(x)] \quad (10)$$

where:

- S is a subset of features not containing i ,
- $f_S(x)$ is the expected value of the model when only features in subset S are present,
- n is the total number of features.

The key property of SHAP is additivity: the sum of all Shapley values equals the difference between the prediction and the average model output. That is:

$$f(x) = E[f(x)] + \sum_{i=1}^n \phi_i \quad (11)$$

This ensures that the explanation is consistent and complete.

SHAP has become particularly valuable in neuroimaging and medical applications, where understanding which features (e.g., brain regions or connectivity metrics) most influence a model’s classification is essential for clinical interpretability. It is capable of generating both global explanations (aggregating across samples) and local ones (explaining individual predictions), making it versatile for various diagnostic tasks.

In the context of our work, SHAP was used to interpret ML classifiers applied to graph metrics extracted from brain networks. By assigning a relevance score to each feature, SHAP enabled the identification of the most discriminative network characteristics (e.g., assortativity, average degree), aiding clinicians in understanding the mechanistic underpinnings of traumatic brain injury and guiding personalized interventions [149].

9.5.3 Machine Learning Framework for TBI Classification

To evaluate the potential of graph-derived features in discriminating between different stages of TBI, we implemented a comprehensive ML pipeline, integrating both traditional classifiers and ensemble strategies. The classification aimed to distinguish between three subject groups, acute, chronic, and control, based on topological features extracted from individual brain connectomes.

Five widely adopted ML algorithms were employed: DT, RF, LR, AdaBoost, and SVM, chosen for their proven applicability in neuroimaging-based diagnosis and for their diverse decision boundaries and complexity levels [186, 187]. Each algorithm was trained using the full set of graph metrics described in the previous section, enabling a multi-faceted interpretation of network-level changes induced by brain trauma. To ensure robust

model evaluation and minimize both overfitting and underfitting risks, a multi-tiered validation strategy was adopted. Initially, Leave-One-Out Cross-Validation (LOOCV) was performed to yield unbiased estimates of generalization performance. Following this, a grid search was employed to fine-tune the hyperparameters for each classifier, optimizing their predictive capacity. To further strengthen model robustness and mitigate overfitting, a 5-fold Nested Cross-Validation (NCV) scheme was implemented. NCV involves an outer loop for model evaluation and an inner loop for hyperparameter tuning, ensuring that parameter selection does not bias the performance assessment. The specific configurations selected for each model are reported in Table 12, representing the result of an exhaustive grid-based exploration across relevant parameter ranges.

Table 12: Hyperparameters for each model

Model	Hyperparameters
RandomForest	n_estimators: 200, max_depth: 5
AdaBoost	n_estimators: 50, learning_rate: 0.1
SVM	C: 0.1, kernel: 'linear', gamma: 'scale'
Logistic Regression	C: 10, penalty: 'l1'
Decision Tree	max_depth: 10

Beyond single classifiers, we introduced an ensemble model that aggregates predictions from RF, AdaBoost, and SVM through majority voting. This approach aimed to leverage the individual strengths of each classifier while compensating for their weaknesses, leading to potentially more robust and generalizable predictions.

9.5.4 Topological Analysis and Group-wise Comparison

The analysis of structural connectivity metrics was conducted to capture both global and local variations across the three diagnostic categories. As shown in Figure 38, the distributions of key graph-theoretical features revealed distinct trends. Notably, WAC values were elevated in the chronic cohort, suggesting increased local connectivity that might reflect long-term network reorganization following injury. Conversely, lower WAC values in the control group support the hypothesis of a more streamlined, non-redundant architecture in healthy individuals. Similarly, the density metric exhibited a clear gradient, being highest in the control group and reduced in both patient groups, likely due to diffuse axonal injury and compromised long-range connectivity. Other features, such as Assortativity and GC, showed heterogeneous patterns across conditions, indicating altered patterns of node-to-node affinity and modular cohesion following trauma.

9.5.5 Statistical Validation

A rigorous statistical analysis was conducted to evaluate the significance of observed differences. Parametric (ANOVA) and non-parametric (Kruskal-Wallis) tests were used to assess the robustness of group-level variations. Notably, metrics such as Density, Assortativity, AD, and CC displayed statistically significant differences across groups, indicating that traumatic brain injury alters network efficiency and modular organization in both acute and chronic phases. These findings were particularly pronounced in features reflecting integration (e.g., IGD, LCS) and resilience (e.g., NCC, NIN), pointing to the fragmentation of structural connectivity as a hallmark of TBI. The progressive decline

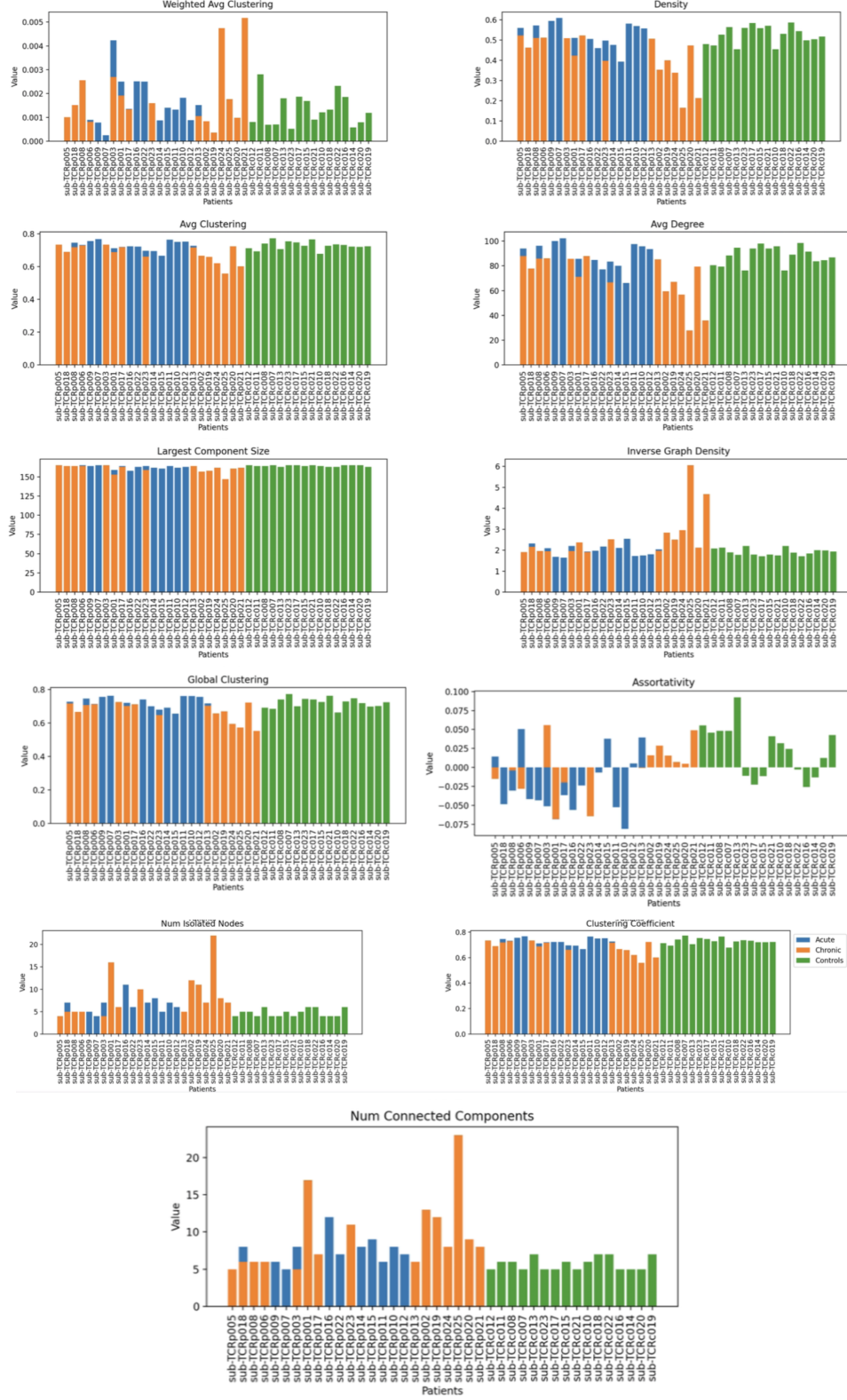


Figure 38: Comparative profiles of neural connectivity metrics. Orange represents the acute cohort, blue denotes the chronic cohort, and green corresponds to the control cohort.

from acute to chronic in certain metrics may reflect the persistence or worsening of damage over time.

9.5.6 Correlation Analysis of Network Features

To investigate interdependencies among the extracted features, we computed Pearson correlation matrices for each cohort, illustrated in Figure 39. These matrices reveal unique patterns within each group, with strong positive correlations between AD and GC in the patient groups, indicative of preserved local modularity despite global fragmentation. In contrast, controls showed more diverse and less correlated feature patterns, consistent with a more balanced network architecture.

9.5.7 Classification Performance and Feature Interpretability

The performance of each model is summarized in Table 13. Among all classifiers, the Random Forest demonstrated the best overall accuracy (63.27%), confirming its suitability for handling high-dimensional data with complex feature interactions. The ensemble model, although slightly less accurate, yielded the highest precision, recall, and F1-score, supporting its robustness and potential clinical applicability.

Table 13: Results of machine learning models for the classification of traumatic brain injuries.

Model	Accuracy	Precision	Recall	F1-Score
Decision Tree	0.5102	0.68	0.662	0.634
Random Forest	0.6327	0.831	0.832	0.849
Logistic Regression	0.5306	0.691	0.673	0.675
AdaBoost	0.5714	0.868	0.837	0.829
SVM	0.5309	0.669	0.612	0.613
Ensemble Model	0.4648	0.875	0.857	0.853

To enhance model transparency, SHAP was applied to each classifier. The SHAP summary plots in Figure 40 reveal that features such as Assortativity, AD, LCS, and IGD consistently contributed most to model predictions. Their prominence across models emphasizes their diagnostic relevance and potential as biomarkers of TBI-induced connectomic alterations.

9.6 Exploiting Graph Convolutional Networks for Insightful Classification and Explanation of Traumatic Brain Injury

9.6.1 Graph Convolutional Networks (GCNs)

Graph Convolutional Networks (GCNs) are a class of neural networks designed to operate directly on graph-structured data. Unlike traditional CNNs, which rely on regular grid structures such as images or time-series data, GCNs can learn from data where relationships among entities are naturally represented as nodes and edges, such as social networks, molecules, or brain connectomes.

In the context of neuroscience and neuroimaging, GCNs have proven to be particularly effective for modelling brain connectivity data, where each node typically represents a brain region and edges reflect structural or functional connectivity between them [183, 152, 151]. Their ability to integrate topological information and node features makes them well-suited for detecting patterns associated with neurological disorders, including TBI, dementia, and DoC.

Mathematically, a graph is defined as $G = (V, E)$, where V is the set of N nodes and $E \subseteq V \times V$ the set of edges. Each graph is associated with an adjacency matrix

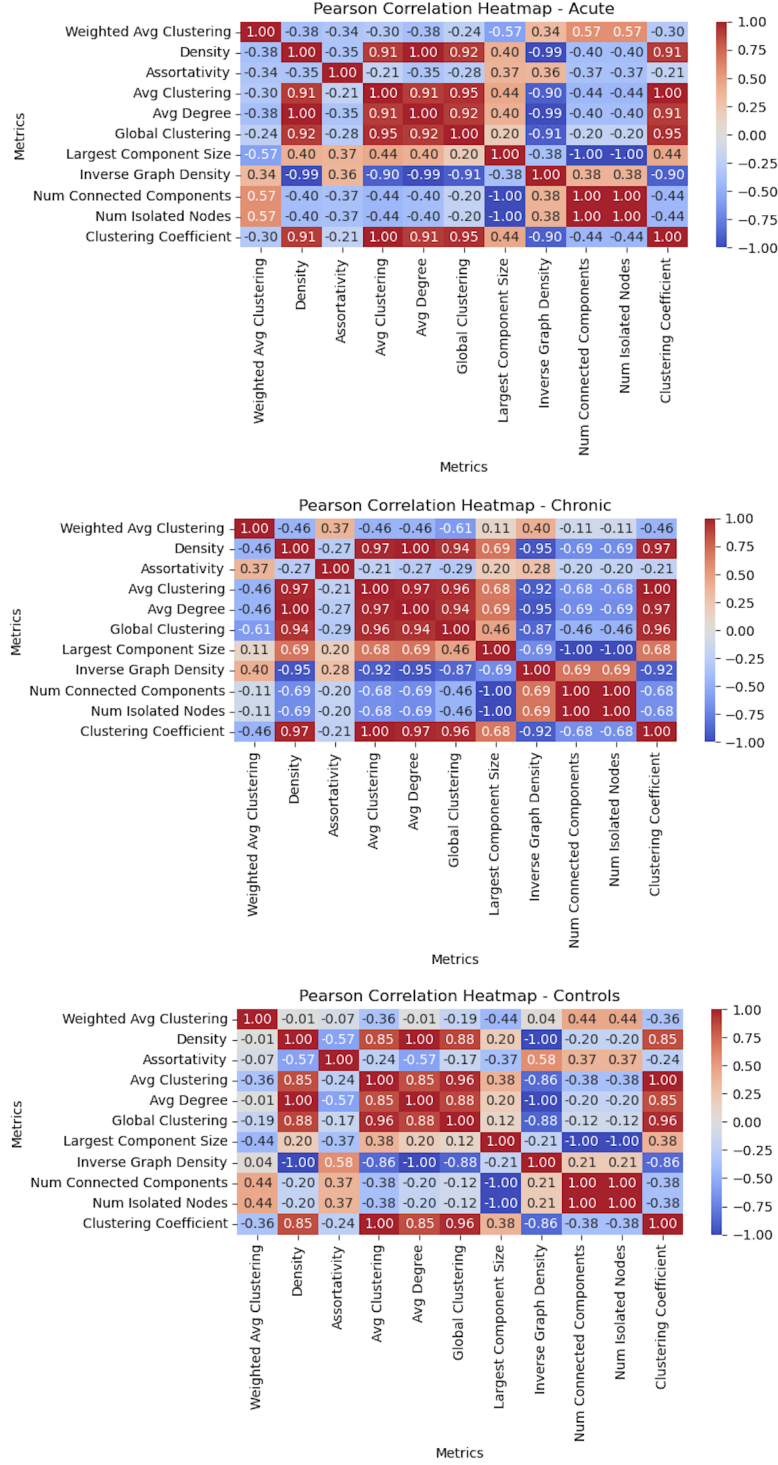


Figure 39: Pearson correlation matrices of the three classes

$\mathbf{A} \in \mathbb{R}^{N \times N}$, where $\mathbf{A}_{ij}$ indicates the connection strength between nodes i and j , and a feature matrix $\mathbf{X} \in \mathbb{R}^{N \times F}$, where each row corresponds to the F -dimensional feature vector of a node.

The forward propagation rule for a basic GCN layer, as introduced by Kipf and Welling [183], is given by:

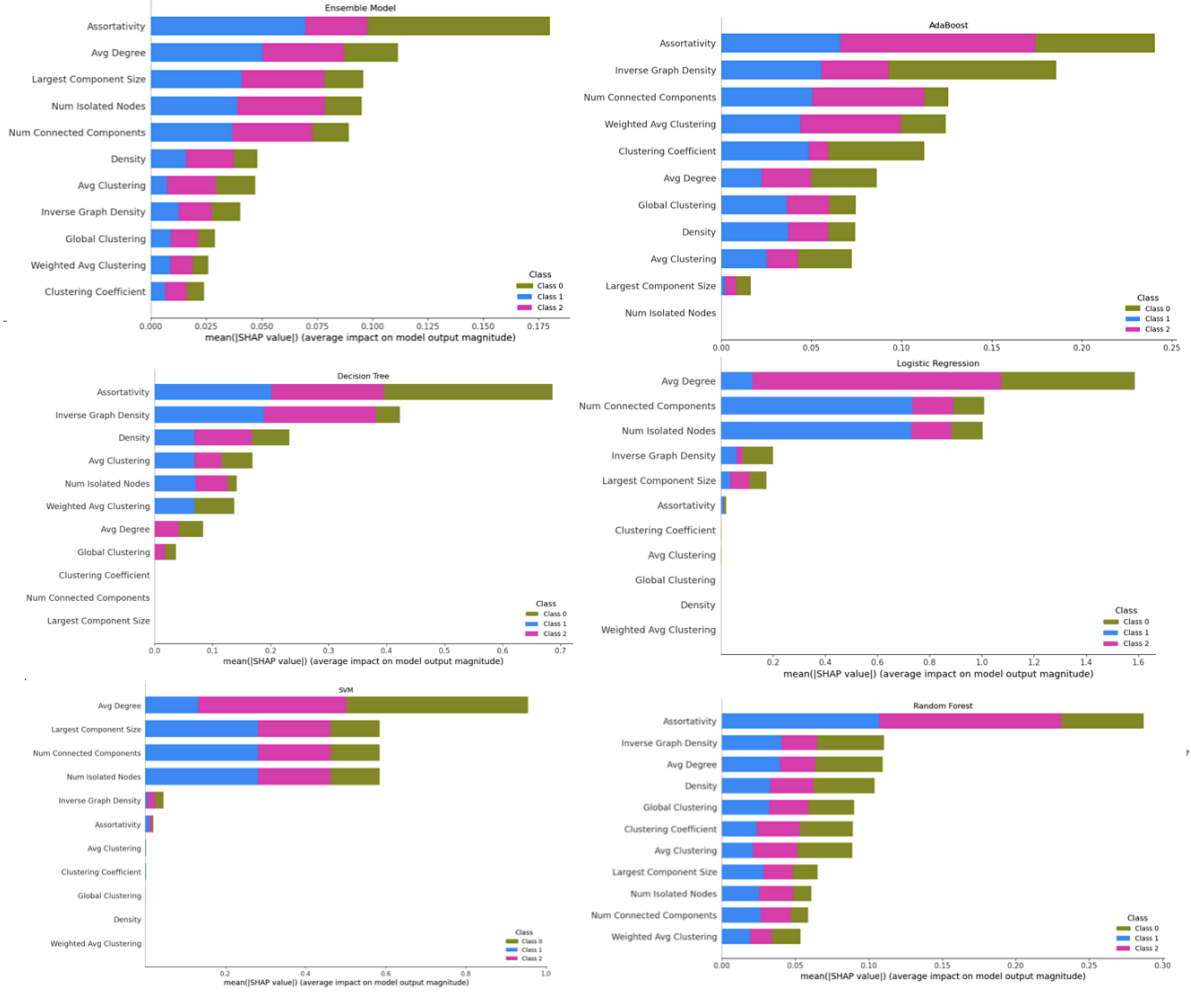


Figure 40: SHAP diagrams of models. Class 0 corresponds to the Acute group, class 1 to the Chronic cohort, class 2 to the Controls group.

$$\mathbf{H}^{(l+1)} = \sigma \left(\tilde{\mathbf{D}}^{-\frac{1}{2}} \tilde{\mathbf{A}} \tilde{\mathbf{D}}^{-\frac{1}{2}} \mathbf{H}^{(l)} \mathbf{W}^{(l)} \right) \quad (12)$$

where:

- $\tilde{\mathbf{A}} = \mathbf{A} + \mathbf{I}$ is the adjacency matrix with added self-loops;
- $\tilde{\mathbf{D}}$ is the degree matrix of $\tilde{\mathbf{A}}$;
- $\mathbf{H}^{(l)}$ is the feature matrix at layer l (with $\mathbf{H}^{(0)} = \mathbf{X}$);
- $\mathbf{W}^{(l)}$ is the learnable weight matrix at layer l ;
- $\sigma(\cdot)$ is an activation function, such as ReLU.

This propagation rule can be interpreted as performing a weighted aggregation of neighbouring node features, normalized to ensure numerical stability.

GCNs have shown promise in clinical contexts because they can naturally incorporate both regional brain features and inter-regional connectivity patterns. Recent studies have used GCNs to classify patient states (e.g., healthy vs. pathological) and predict outcomes

in TBI [188, 158], often achieving superior results compared to classical ML methods. Nevertheless, their black-box nature has raised concerns regarding explainability, particularly in critical domains such as medicine, where interpretability is vital for clinical decision-making.

9.6.2 Gradient-weighted Class Activation Mapping (Grad-CAM) for Explainable Graph Models

Gradient-weighted Class Activation Mapping (Grad-CAM) is a popular interpretability technique originally developed to explain the predictions of CNNs by generating visual saliency maps that highlight regions of the input most influential for a specific output decision. Grad-CAM belongs to the family of posthoc explainability methods that use gradient information to compute class-discriminative localization maps, thereby improving the transparency of otherwise opaque deep learning systems.

In traditional CNN applications, Grad-CAM operates on feature maps A^k from a chosen convolutional layer, computing the gradient of the output class score y^c concerning those maps. The importance weights α_k^c for each feature map channel k are calculated as:

$$\alpha_k^c = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k} \quad (13)$$

where Z is the spatial size of the feature map (i.e., number of pixels or units), and $\frac{\partial y^c}{\partial A_{ij}^k}$ denotes the partial derivative of the class score to each spatial location (i, j) in channel k . These weights capture how important each channel is for the given class. The final Grad-CAM heatmap $L_{\text{Grad-CAM}}^c$ is then obtained by a weighted linear combination of the feature maps followed by a ReLU operation:

$$L_{\text{Grad-CAM}}^c = \text{ReLU} \left(\sum_k \alpha_k^c A^k \right) \quad (14)$$

The ReLU ensures that only features with a positive influence on the output are highlighted.

9.6.2.1 Extension to Graph Neural Networks

With the increasing use of graph neural networks (GNNs) in domains such as brain connectivity, molecular analysis, and social networks, Grad-CAM has been adapted for use with non-Euclidean data structures. In GNNs, data are structured as graphs $G = (V, E)$, where V is the set of nodes and E is the set of edges. Each node $v_i \in V$ is associated with a feature vector $\mathbf{h}_i \in \mathbb{R}^d$.

In this setting, the saliency of a node is determined by computing the gradient of the target class score y^c to the node’s latent representation $\mathbf{h}_i$. The importance score of node i is given by:

$$S^c(v_i) = \text{ReLU} \left(\sum_{k=1}^d \alpha_k^c h_i^k \right) \quad (15)$$

where the weights α_k^c are calculated analogously to the CNN case, using global average pooling over all nodes:

$$\alpha_k^c = \frac{1}{|V|} \sum_{i=1}^{|V|} \frac{\partial y^c}{\partial h_i^k} \quad (16)$$

Here, h_i^k is the k -th dimension of the embedding of node i . The result is a node-level importance map that highlights the contribution of each node to the classification decision.

9.6.2.2 Saliency Interpretation and Challenges

Grad-CAM enables a form of visual reasoning, where users can inspect which parts of the input graph were most responsible for a model’s output. This is particularly relevant in high-stakes settings such as neuroscience, medicine, or finance, where interpretability is critical. By mapping node-level importance back onto domain-specific structures (e.g., brain regions), practitioners can validate model behaviour and generate new hypotheses.

However, gradient-based methods like Grad-CAM come with limitations. Their saliency maps are often sensitive to small perturbations in the input or model weights. Moreover, they may not always be faithful to the underlying decision logic of the model, especially in deep or highly regularized networks. It is therefore recommended to combine Grad-CAM with sanity checks and quantitative evaluation of saliency faithfulness, as outlined in works such as [189, 190].

9.6.3 Graph Metrics for Node-Level Analysis

In the present study, a selection of node-level graph metrics was employed to quantitatively characterize the topological roles of individual brain regions within structural connectomes derived from DWI. Each connectome was modelled as a weighted, undirected graph, where nodes correspond to anatomically defined brain regions and edges represent the strength of white matter connectivity between them. These features capture both local and global aspects of network organization, allowing for a detailed assessment of each region’s role within the broader brain network.

The features analyzed in this context include classical measures such as degree, PageRank, betweenness centrality, local efficiency, average neighbour degree, weighted clustering coefficient, eigenvector centrality, and Katz centrality. These metrics are widely used in network neuroscience to evaluate how nodes contribute to communication, influence, redundancy, and resilience within the system [185].

9.6.3.1 Degree (k_i)

This is the most direct measure of a node’s connectivity and represents the sum of edge weights linking node i to its neighbours:

$$k_i = \sum_{j=1}^N A_{ij}, \quad (17)$$

where A_{ij} denotes the weight of the edge between nodes i and j , and N is the total number of nodes in the graph [191].

9.6.3.2 PageRank (PR_i)

Originally designed for web page ranking, this metric captures node importance based on the structure of incoming connections and their respective significance:

$$PR_i = \frac{1-d}{N} + d \sum_{j \in M_i} \frac{PR_j}{k_j^{out}}, \quad (18)$$

where d is the damping factor (typically 0.85), M_i is the set of nodes with edges pointing to i , and k_j^{out} is the out-degree of node j [192].

9.6.3.3 Betweenness Centrality (BC_i)

This metric measures how frequently a node lies on the shortest paths connecting other node pairs, thereby identifying critical bridges in the network:

$$BC_i = \sum_{s \neq i \neq t} \frac{\sigma_{st}(i)}{\sigma_{st}}, \quad (19)$$

where σ_{st} is the number of shortest paths between nodes s and t , and $\sigma_{st}(i)$ denotes the number of those paths passing through node i [193].

9.6.3.4 Local Efficiency ($E_{loc}(i)$)

This index quantifies the efficiency of communication between a node's immediate neighbours when the node itself is removed:

$$E_{loc}(i) = \frac{1}{k_i(k_i - 1)} \sum_{j, h \in N_i} ((A_{jh} + A_{hj}) \cdot l_{jh}^{-1}), \quad (20)$$

with N_i representing the set of neighbors of node i and l_{jh} the shortest path between nodes j and h [194].

9.6.3.5 Average Neighbor Degree (AND_i)

This feature describes the average degree of a node's neighbours, providing insights into the local environment of the node:

$$AND_i = \frac{1}{k_i} \sum_{j \in N_i} k_j, \quad (21)$$

which is indicative of whether a node tends to be surrounded by highly connected neighbours [195].

9.6.3.6 Weighted Clustering Coefficient (C_i)

This metric evaluates the degree to which a node's neighbours are interconnected, accounting for the strength of the links:

$$C_i = \frac{1}{s_i(k_i - 1)} \sum_{j, h \in N_i} \frac{(w_{ij} + w_{ih})}{2} \cdot w_{jh}, \quad (22)$$

where $s_i = \sum_{j \in N_i} w_{ij}$ is the strength of node i and w_{jh} denotes the edge weight between neighbors j and h [196].

9.6.3.7 Eigenvector Centrality (EC_i)

This centrality measure assigns higher importance to nodes connected to other well-connected nodes. It is computed through the eigenvector of the adjacency matrix A :

$$EC_i = \frac{1}{\lambda} \sum_{j=1}^N A_{ij} EC_j, \quad (23)$$

where λ is the leading eigenvalue of the matrix [197].

9.6.3.8 Katz Centrality (KC_i)

Katz centrality extends eigenvector centrality by considering all paths (not only direct neighbors), while exponentially dampening longer walks:

$$KC_i = \alpha \sum_{j=1}^N A_{ij} KC_j + \beta, \quad (24)$$

where α is the attenuation factor (less than $1/\lambda$) and β is an additive constant representing baseline centrality [198].

To evaluate the interdependence between features, a Spearman rank correlation analysis was performed. The resulting correlation matrix, shown in Fig. 41, reveals both redundancy and complementarity among the chosen metrics. Although certain features, such as Degree and Average Neighbor Degree, exhibit high correlation, they were retained due to their distinctive interpretability and potential predictive relevance in classification tasks. Including multiple correlated yet conceptually diverse metrics can enhance model generalizability and resilience against overfitting.

Incorporating such a diverse set of topological descriptors enables the construction of a rich feature space, capturing multiscale properties of the network and facilitating subsequent ML and explainability analyses aimed at identifying clinically meaningful biomarkers in TBI patients. The selection of these specific graph metrics is not only grounded in network theory, but also guided by neuroscientific considerations related to traumatic brain injury. In particular, the degree metric offers a straightforward representation of how structurally integrated a brain region is, and reductions in degree may indicate the disruption of critical communication pathways following trauma. Similarly, PageRank and eigenvector centrality are particularly relevant in the context of TBI as they reflect a node's importance in terms of its connectivity to other influential regions, which may correspond to cortical hubs known to be involved in consciousness and cognitive control. Betweenness centrality serves as a proxy for network vulnerability: regions with high betweenness may act as critical bridges, and their compromise could lead to disconnection syndromes or altered states of consciousness. Katz centrality, by incorporating indirect paths, provides a more global assessment of a region's communicative reach and has been linked to functional resilience in disrupted networks. Metrics such as weighted clustering coefficient and local efficiency are essential for understanding localized integration and neighbourhood redundancy, properties which may reflect the capacity of the brain to reroute information in response to injury. Moreover, an average neighbour degree can help identify peripheral nodes that are surrounded by highly connected regions, revealing compensatory mechanisms in structurally weakened networks.

Altogether, these features offer a multifaceted perspective on how different brain areas participate in, or are excluded from, global communication patterns. In the context of

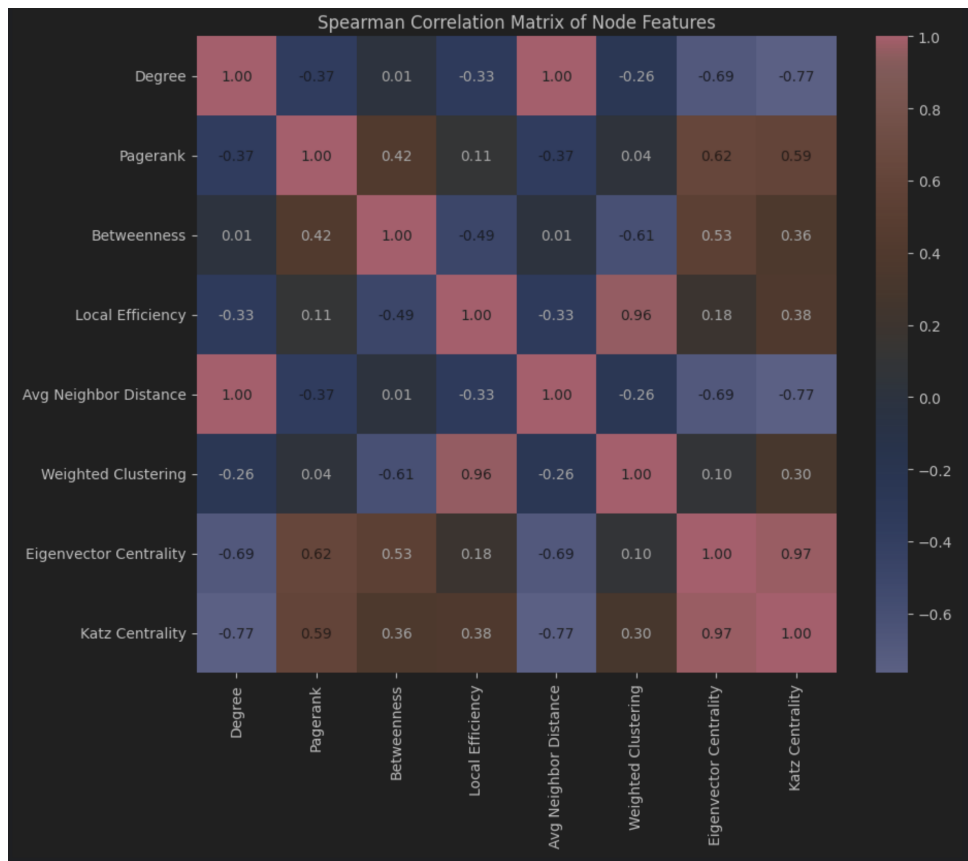


Figure 41: Spearman correlation matrix displaying relationships between various node-level metrics extracted from the brain connectomes.

TBI, such insights are critical for differentiating between patients at various stages of recovery or degrees of impairment, and for identifying potential targets for neuromodulation or rehabilitation strategies. By quantifying structural disruptions at the node level, these metrics contribute to a deeper understanding of the mechanisms underlying consciousness disorders and cognitive dysfunction in traumatic brain injury.

9.6.4 Graph Convolutional Network Architecture and Training Strategy

To analyze brain connectivity patterns and distinguish between clinical conditions in patients with TBI, we developed a GCN tailored to process structural connectomes derived from DWI. Each subject’s brain network was modelled as an undirected weighted graph, where nodes correspond to brain regions defined by the AAL3v1 atlas (170 in total), and edges represent the strength of structural connections computed via probabilistic tractography.

Each node in the graph was annotated with an 8-dimensional feature vector encoding graph-based metrics, degree, PageRank, betweenness centrality, local efficiency, average neighbour degree, weighted clustering coefficient, eigenvector centrality, and Katz centrality. These features provide a rich characterization of the local topological role of each region in the global brain network.

As summarized in Table 14, the proposed GCN architecture consists of three consecutive graph convolutional layers. The first layer processes the initial 8-dimensional input and expands it to 32 feature dimensions per node. The second layer projects these 32-dimensional representations to 64 dimensions, and the third further expands the node embeddings to 128 dimensions. Each convolutional layer is followed by a Leaky ReLU non-linearity (with negative slope 0.01), batch normalization to stabilize the training dynamics, and dropout (rate = 0.5) to mitigate overfitting. To aggregate node-level information into a subject-level embedding, we used Global Average Pooling (GAP), which computes the average of each feature across all nodes in the graph, resulting in a 128-dimensional vector summarizing the overall connectivity structure of the brain. This pooled vector is passed through three fully connected layers with progressively reduced dimensionality ($128 \rightarrow 64 \rightarrow 32 \rightarrow 16$), ending with a final classification layer with 3 output units corresponding to the acute, chronic, and control classes. The model outputs logits, which are interpreted via the softmax function during the evaluation phase. The training was performed using the Adam optimizer with an initial learning rate of 1×10^{-5} and a weight decay of 1×10^{-4} . The cross-entropy loss function was employed to guide the learning process. The model was trained over 450 epochs, providing ample opportunity for convergence.

We evaluated the model using Leave-One-Out Cross-Validation (LOOCV), in which each subject is iteratively excluded from the training set and used for testing. This approach ensures robustness and minimizes bias due to the small dataset size (49 subjects). The model’s predictive performance was quantified using accuracy, precision, recall, and F1-score. All computations were carried out using PyTorch 2.2 and Python 3.9 on a macOS system equipped with a 3.1 GHz Intel Core i5 processor and an AMD Radeon Pro 5300 GPU.

9.6.5 Interpretability via Grad-CAM: Identifying Salient Brain Regions

To interpret the predictions of our GCN and uncover which brain regions most strongly influenced the classification decision, we extended the Grad-CAM framework to graph-structured data. This method allows us to estimate the contribution of each node in the

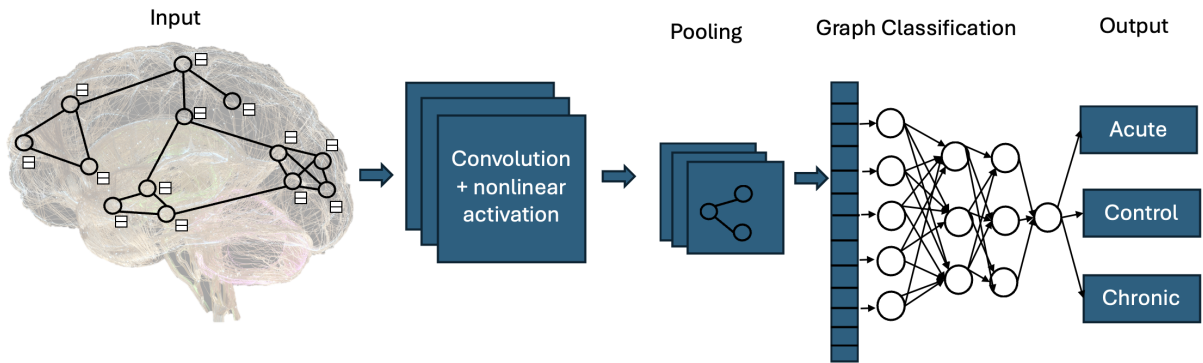


Figure 42: Overview of the implemented GCN architecture. The graph undergoes successive graph convolutions, activation and pooling operations, leading to a subject-level representation used for final classification into three groups: acute, chronic, and controls.

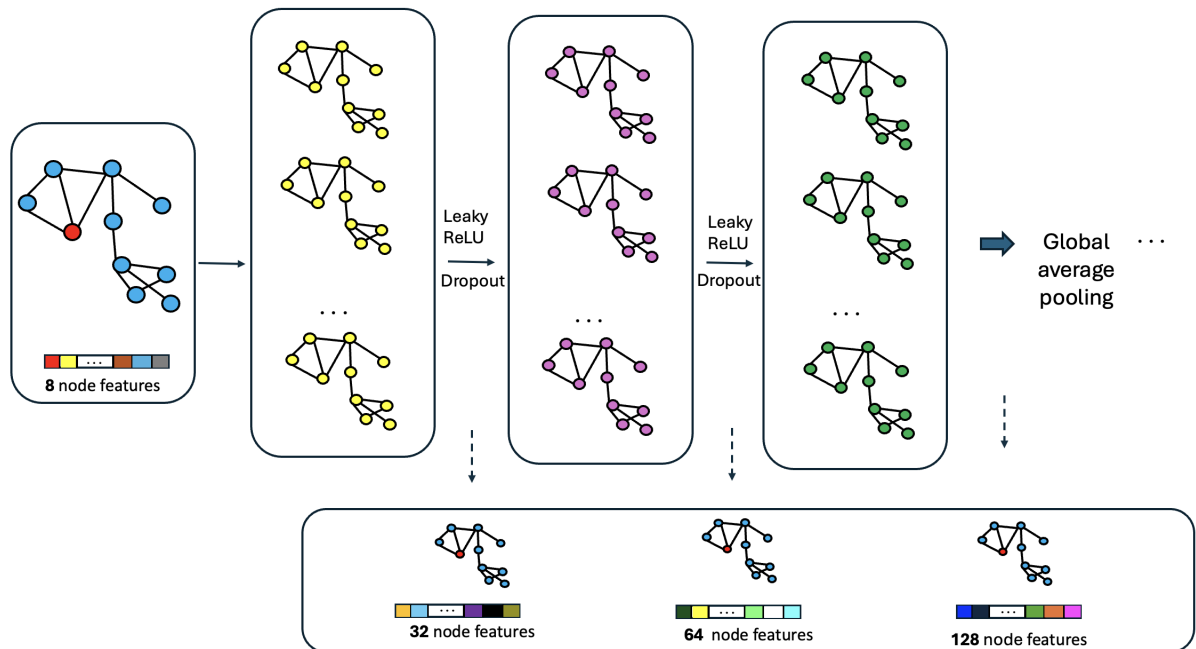


Figure 43: Detail of the feature transformation pipeline through graph convolution layers. Node features are progressively refined via GCN, nonlinearity, and dropout before global pooling.

graph to the final decision of the model, facilitating the extraction of clinically relevant neural substrates.

For each subject, we computed the gradients of the predicted class score (logit) with respect to the input node features. Let y^c denote the score corresponding to class c , and let $h_i^l \in \mathbb{R}^F$ be the feature vector of node i at layer l . The Grad-CAM score for node i is computed as:

$$S_i = \text{ReLU} \left(\sum_{f=1}^F \frac{\partial y^c}{\partial h_{i,f}^l} \right) \quad (25)$$

This expression aggregates the influence of all feature channels at a selected convolutional layer, capturing how sensitive the output is to variations in each node’s representation. Applying the ReLU ensures that only positive contributions are retained, consistent with the standard Grad-CAM rationale. After computing S_i for all nodes, we selected the top 30 nodes per subject, approximately 18% of the brain regions in the atlas, as the most relevant. This threshold was empirically validated: lower values reduced discriminative power, while larger selections introduced noise. Visual inspection confirmed that these nodes frequently aligned with known hubs or clinically implicated regions in disorders of consciousness.

This method not only highlights spatially distinct brain areas that contribute most to classification, but also offers a bridge between black-box neural networks and clinically meaningful biomarkers. Notably, similar practices are supported in the literature, where the examination of high-centrality or “rich-club” nodes has enhanced understanding of cognitive resilience and dysfunction [199, 200]. For visualization, NetworkX was used to render the brain networks, with node colours indicating importance scores and anatomical labels derived from the AAL3v1 atlas. This helped synthesize anatomical and topological insights, interpreting model decisions more intuitive for neuroscientific and clinical analysis.

9.7 Results

9.7.0.1 Performance Evaluation of the GCN Model

To evaluate the effectiveness of the proposed GCN in classifying subjects based on their brain connectivity patterns, we conducted a thorough performance assessment using standard metrics: accuracy, precision, recall, and F1-score. The model was trained and tested using a LOOCV framework, maximizing data utilization and ensuring robust validation.

On average, the GCN achieved an accuracy of 83.67%, highlighting its high reliability in correctly classifying subjects into their respective clinical groups. The model’s precision reached 81.6%, suggesting a low false positive rate, which is particularly important in medical contexts to avoid unnecessary concerns or interventions in healthy individuals. The average recall was 78%, reflecting the model’s ability to detect true positive cases, and the resulting F1-score, balancing precision and recall, was 79%.

These metrics, summarized in Table 15, confirm the robustness of the GCN in identifying neural signatures associated with TBI and distinguishing them from typical brain network patterns. Moreover, identifying the worst-performing fold within LOOCV offers a pathway to understanding which subject characteristics may limit model generalization, suggesting directions for further feature refinement or future integration of multimodal clinical data.

Table 14: Summary of GCN Architecture and Training Configuration

Parameter	Value / Description
Input features per node	8 (topological metrics: degree, PageRank, etc.)
Number of nodes per graph	170 (AAL3v1 atlas regions)
Number of classes	3 (Acute, Chronic, Control)
Graph Convolution Layer 1	$8 \rightarrow 32$ features
Graph Convolution Layer 2	$32 \rightarrow 64$ features
Graph Convolution Layer 3	$64 \rightarrow 128$ features
Activation function	Leaky ReLU ($\alpha = 0.01$)
Normalization	Batch Normalization after each layer
Regularization	Dropout (rate = 0.5)
Pooling layer	Global Average Pooling (GAP)
Fully Connected Layers	$128 \rightarrow 64 \rightarrow 32 \rightarrow 16$
Output layer	$16 \rightarrow 3$ (raw logits)
Loss function	Cross-Entropy Loss
Optimizer	Adam
Learning rate	1×10^{-5}
Weight decay	1×10^{-4}
Epochs	450
Cross-validation	Leave-One-Out (LOOCV)
Evaluation metrics	Accuracy, Precision, Recall, F1-Score
Framework	PyTorch 2.2, Python 3.9
Hardware	iMac, Intel i5, 3.1 GHz, AMD Radeon Pro 5300

Table 15: Classification Performance Metrics of the GCN Model

Metric	Average
Accuracy	83.67%
Precision	81.6%
Recall	78%
F1-Score	79%

9.7.0.2 Interpretability Analysis with Grad-CAM Beyond accuracy, interpretability plays a central role in the practical utility of AI models in neuroscience. For this reason, Grad-CAM was used to investigate the most salient brain regions contributing to the GCN’s predictions. This approach allows for node-level visualization of importance scores derived from gradient analysis. Specifically, the top 30 nodes, approximately 18% of the full 170-region brain atlas, were extracted for each subject, enabling a focused analysis of the most discriminative brain regions while minimizing noise from peripheral features.

Although node relevance was the primary target of this analysis, edge weights were also considered in visual form. Figure 44 presents the most relevant edges identified in an acute patient, coloured by their contribution to the classification. Notably, edges connecting regions such as the Raphe nuclei, VTA, and thalamic nuclei (e.g., Thal_AV_R to Thal_MDm_R) emerged as particularly relevant, aligning with known pathways involved in consciousness and arousal regulation [201, 202, 203].

Figure 45 provides a detailed comparison between significant connected regions (Panel a) and isolated yet influential nodes (Panel b). Interestingly, some nodes lacking visible connections, such as Thal_Re_L and Thal_Re_R, were nonetheless ranked as highly important by the model. Their isolated status in the anatomical atlas contrasts with their functional relevance in classification, suggesting potential disconnections or altered network embedding due to trauma.

9.7.0.3 Group-Level Patterns in Node Importance To investigate shared connectivity disruptions across subjects, we computed frequency histograms for each group, representing the number of patients in whom each brain region was identified as significant. This analysis aimed to uncover common hubs of altered connectivity at the group level. Figures 46–48 depict the distribution of significant brain regions for acute, control, and chronic patients, respectively.

In the acute group (Figure 46), the distribution appears heterogeneous, with several brain regions involved in a relatively small subset of individuals. Despite this variability, some areas consistently emerge as recurrently altered across patients. Notably, regions in the frontal cortex, including Frontal_Sup_2_R, Frontal_Inf_Orb_2_L, and Frontal_Mid_2_R, show high recurrence. Subcortical structures such as the Caudate_L, Putamen_R, and the Thalamus nuclei (e.g., Thal_LP_R, Thal_MDl_L, Thal_Re_R) also appear frequently. The diversity in affected areas may reflect the heterogeneous nature of acute traumatic brain injury and its impact on both cortical and subcortical circuits. In contrast, the control group (Figure 47) exhibits a much narrower pattern. Here, a small number of frontal and limbic regions dominate the distribution, such as Frontal_Sup_2_L, Frontal_Inf_Orb_2_R, and OFCmed_L, with consistently high values across participants. The compactness of this distribution indicates a higher degree of inter-subject consistency, as expected in a neurologically healthy population, with most subjects showing similar patterns of connectivity in these stable hub regions. The appearance of regions in the anterior cingu-

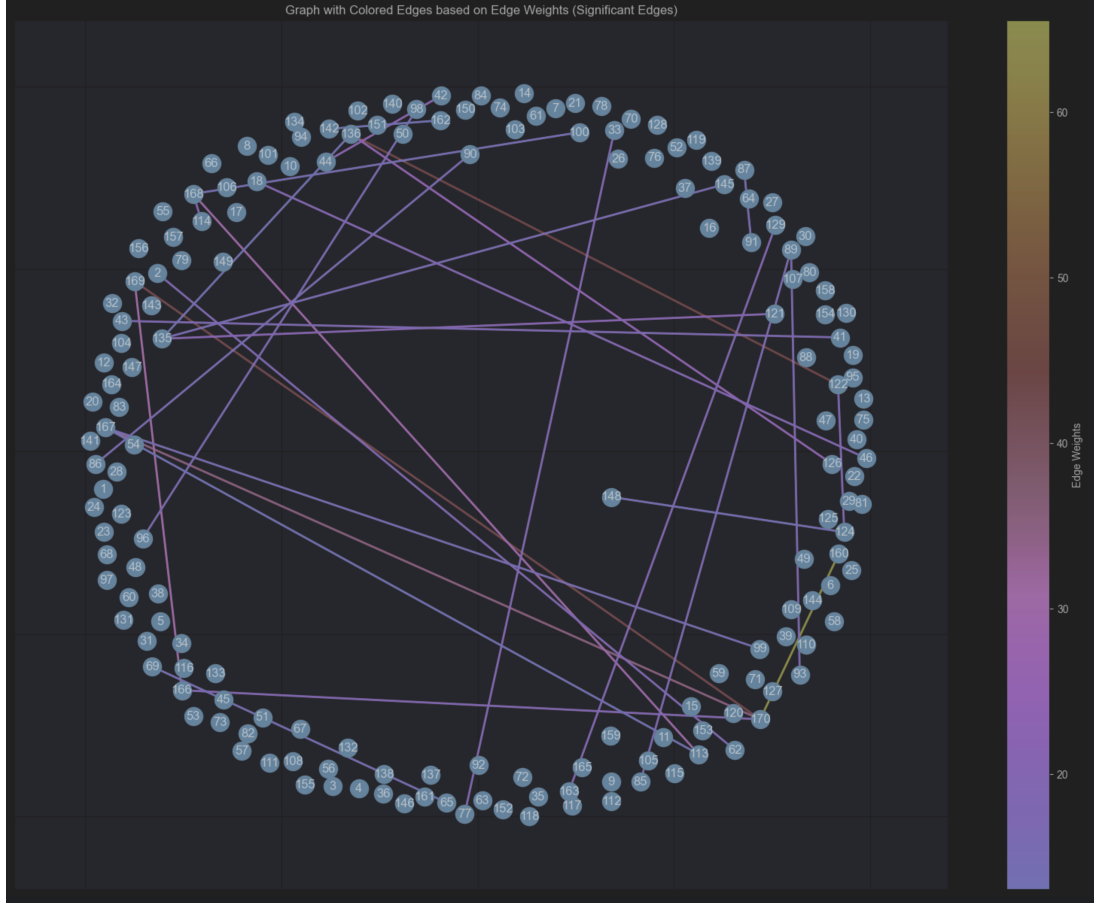


Figure 44: Graph showing all 170 nodes (representing brain regions) and the most significant edges for the acute patient shown in Fig 36. The edges are colour-coded based on their weight, with stronger connections displayed in warmer colours. This visualization emphasizes the key connections between brain regions identified through Grad-CAM analysis.

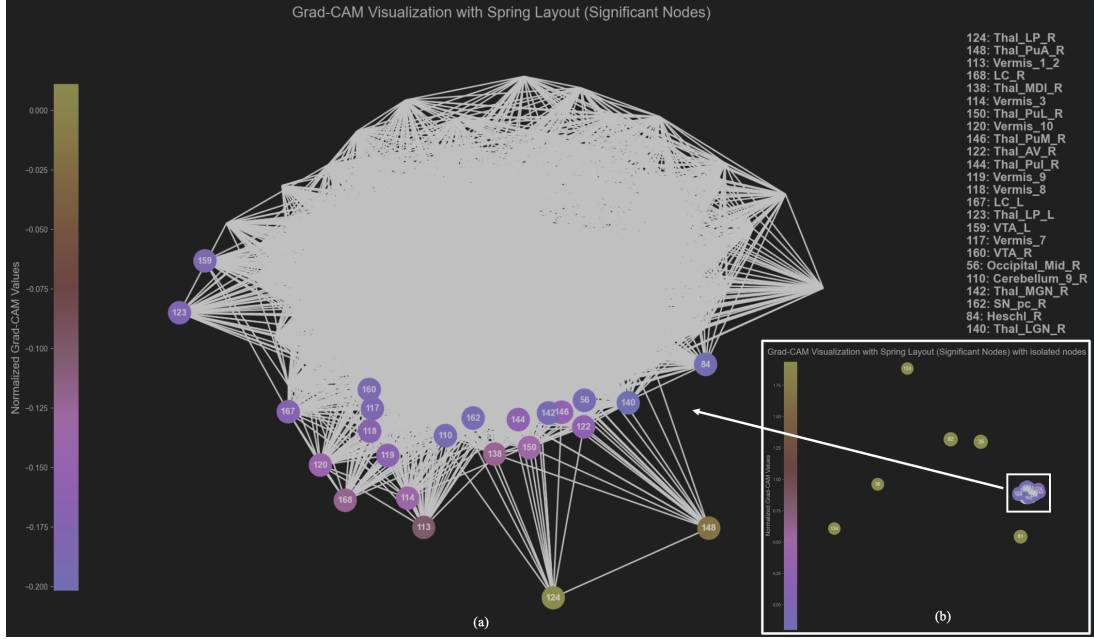


Figure 45: Panel (a) presents a focused view of the brain connectivity network, highlighting significant nodes identified using Grad-CAM, with isolated nodes excluded to emphasize the core functional interactions. Panel (b) includes the isolated nodes, marked by their high influence on network dynamics, illustrating their critical roles despite being anatomically uncharacterized.

late cortex (e.g., ACC_sup_L, ACC_pre_R) is also notable, suggesting their robustness in normal brain function. The chronic group (Figure 48) presents an intermediate configuration. While the distribution is broader than that of controls, it is more focused than the acute group's. Chronic patients frequently show involvement of regions such as the Frontal_Mid_2_L, Olfactory_L/R, Thal_LP_R, and ACC_pre_R, which are associated with long-term changes in cognitive, emotional, and attentional processes. Interestingly, subcortical structures like the Caudate and Thalamus remain prominent, suggesting persistent disconnection or compensatory rewiring in these networks. Compared to the acute state, some regions with initially isolated alterations become more consistently involved across patients, possibly reflecting the convergence of pathological adaptations over time.

These findings demonstrate that each clinical group expresses a distinctive anatomical footprint in terms of affected brain regions. The acute group is marked by variability and widespread disruption, the control group by focal and consistent connectivity, and the chronic group by a transitional profile indicating stabilization around a subset of affected hubs. This type of analysis supports the use of node-level recurrence maps as a tool to identify condition-specific biomarkers and to track the longitudinal evolution of brain connectivity after injury.

9.7.0.4 Visualization of Significant Brain Regions Across Clinical Groups To complement the numerical analysis, we generated spatial visualizations of the brain regions most frequently highlighted by Grad-CAM, providing anatomical context and facilitating interpretation. These figures (Figures 49-54) show DWI-based brain projections in axial, sagittal, and coronal planes, with colour-coded nodes representing region-specific relevance.

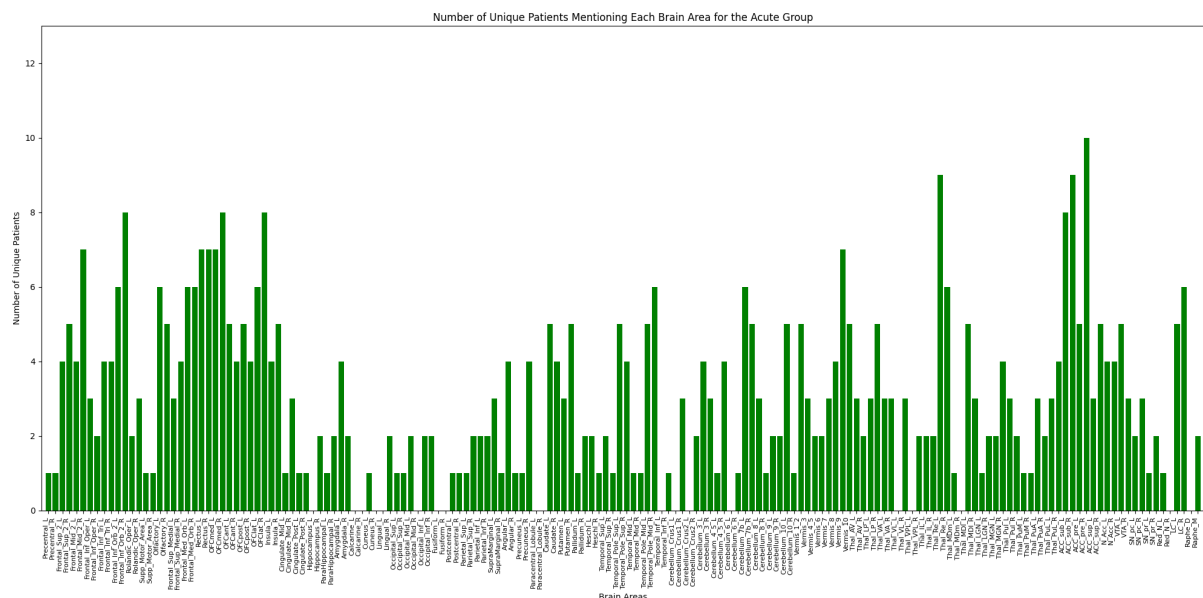


Figure 46: Histogram illustrating the distribution of significant brain regions across the acute patient group. The height of each bar represents the number of unique patients in whom each brain area was identified as significant.

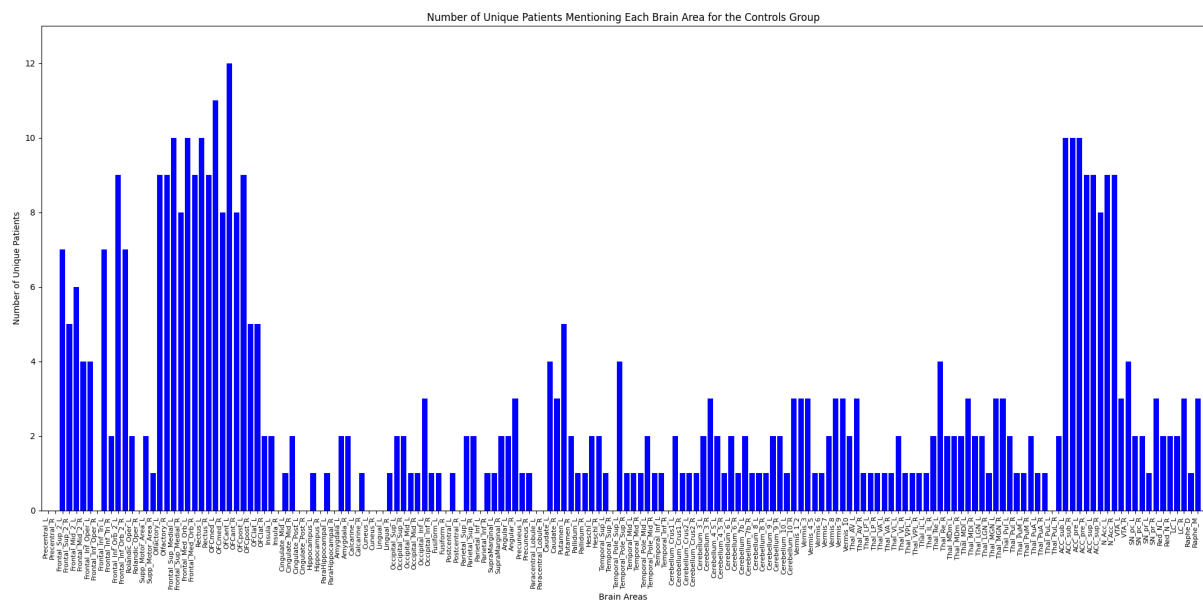


Figure 47: Histogram of brain region mentions in the controls group. The frontal cortex and cerebellum exhibit the highest concentration of mentions, suggesting these areas are most involved in maintaining normal brain function in healthy individuals.

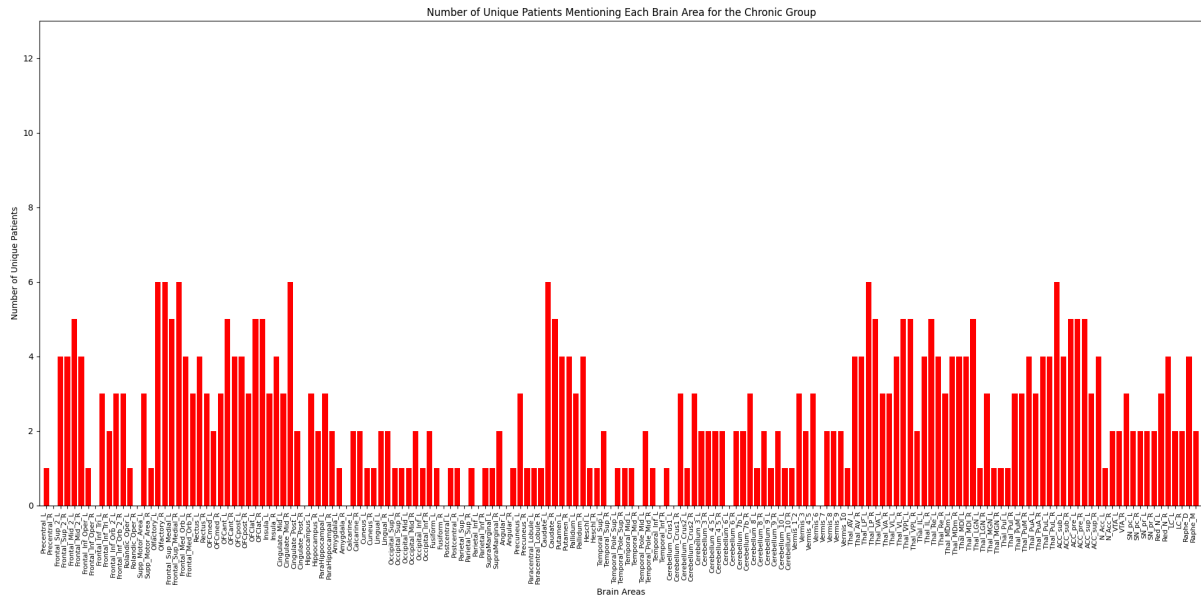


Figure 48: Histogram showing the distribution of significant brain regions in the chronic patient group. Although the overall distribution is wider than in the controls, the mentions are less concentrated than in the acute group, indicating a potential for network stabilization and adaptation over time.

In the **acute group** (Figure 49), green nodes prominently highlight frontal areas, thalamic nuclei, and the brainstem regions known to be essential in arousal, sensorimotor integration, and consciousness. Their widespread activation aligns with the diffuse network breakdown typical of the acute TBI phase.

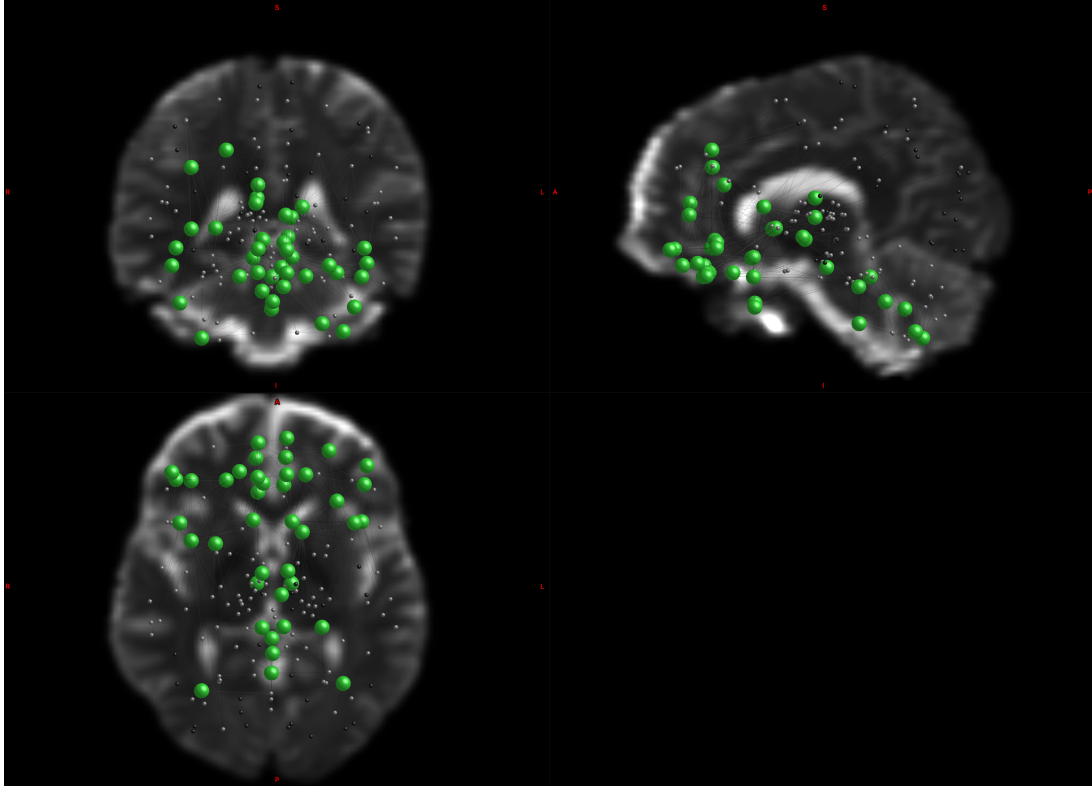


Figure 49: DWI visualization of brain networks in acute patients. The green nodes, as outlined in Table [16](#), represent the most significant areas, showing their spatial distribution across axial, sagittal, and coronal views. Key regions such as the thalamus, frontal cortex, and brainstem, which are involved in motor control, consciousness, and cognition, are disrupted in acute TBI.

In contrast, the **control group** (Figure [50](#)) shows a more concentrated and coherent pattern. Blue nodes mainly cluster in the frontal cortex and olfactory system, consistent with a stable, functionally optimized network.

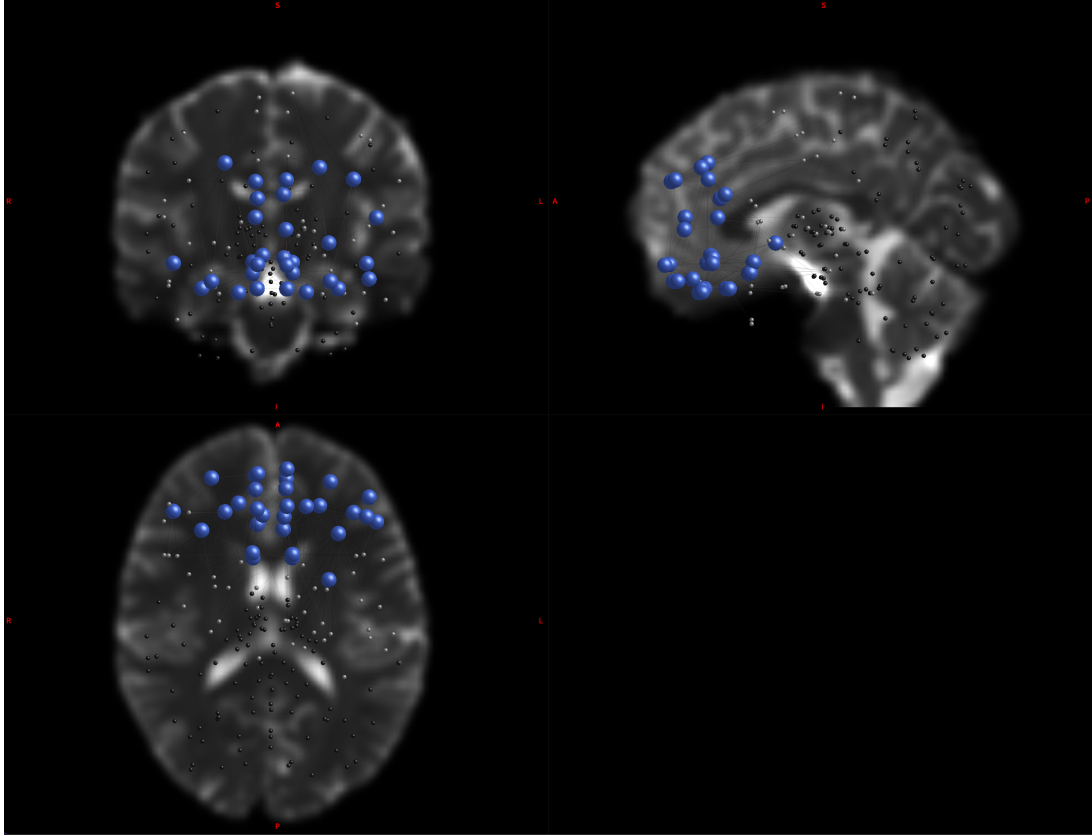


Figure 50: DWI visualization of significant brain regions in the control group. The blue nodes highlight the most important areas, with connections shown in grey. These regions, primarily in the frontal cortex and olfactory areas, reflect stable neural functions in healthy individuals, contrasting with the disruptions seen in TBI patients, as seen in Table [17](#).

The **chronic group** (Figure [51](#)) presents a hybrid configuration. Red nodes point to regions such as the anterior cingulate cortex and thalamus, suggesting ongoing reorganization and incomplete recovery. This intermediate pattern reflects the diversity of trajectories following TBI, encompassing both stable deficits and compensatory adaptations.

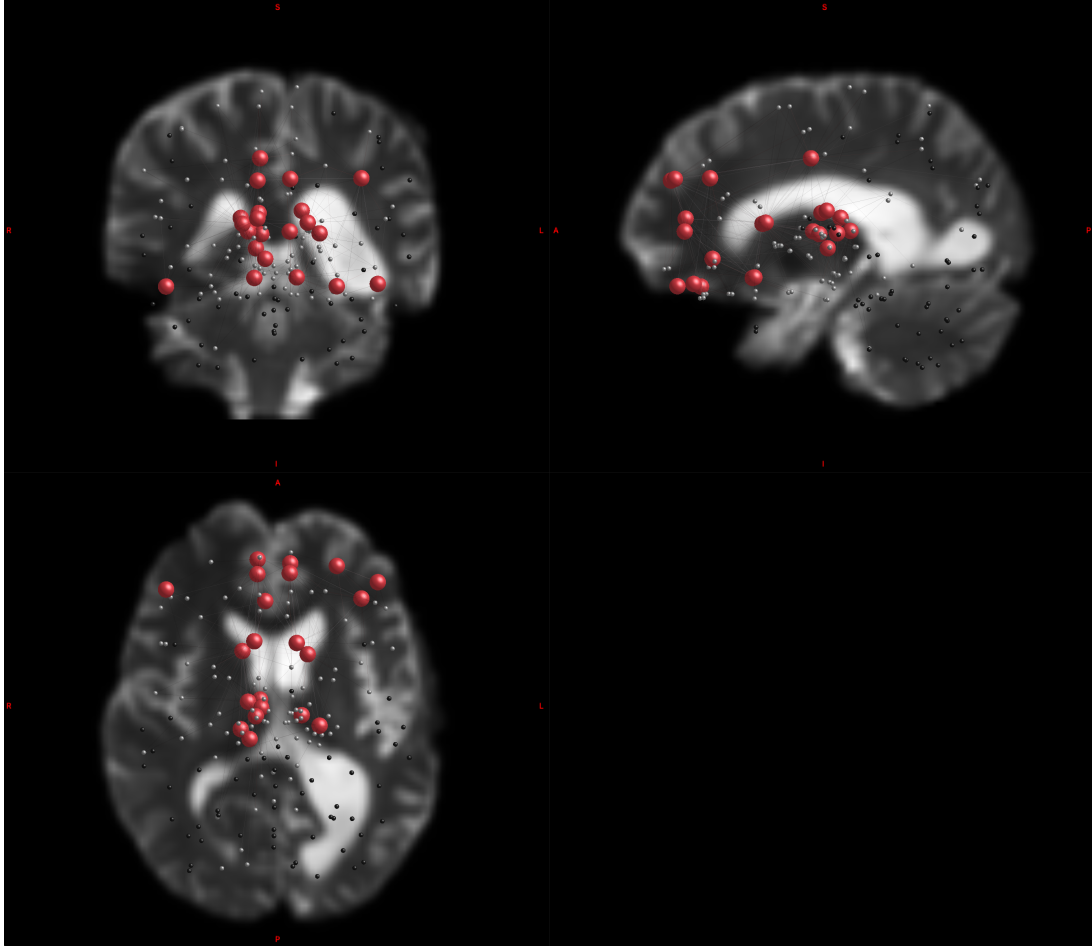


Figure 51: DWI visualization of the most significant brain regions identified in chronic patients without subgroup distinction. The red nodes represent the key regions involved across all chronic patients, highlighting areas primarily in the frontal cortex, thalamus, and brainstem, suggesting persistent disruptions in regions critical for motor control, cognition, and consciousness regulation, as indicated in Table 18.

Figures 52-54 further break down the chronic group into VS, MCS, and recovered consciousness subgroups. These figures show how the spatial distribution of significant nodes reflects the severity of consciousness impairment and the trajectory of recovery.

VS patients (Figure 52) exhibit a narrow set of key regions, mainly subcortical, suggesting profound disconnection.

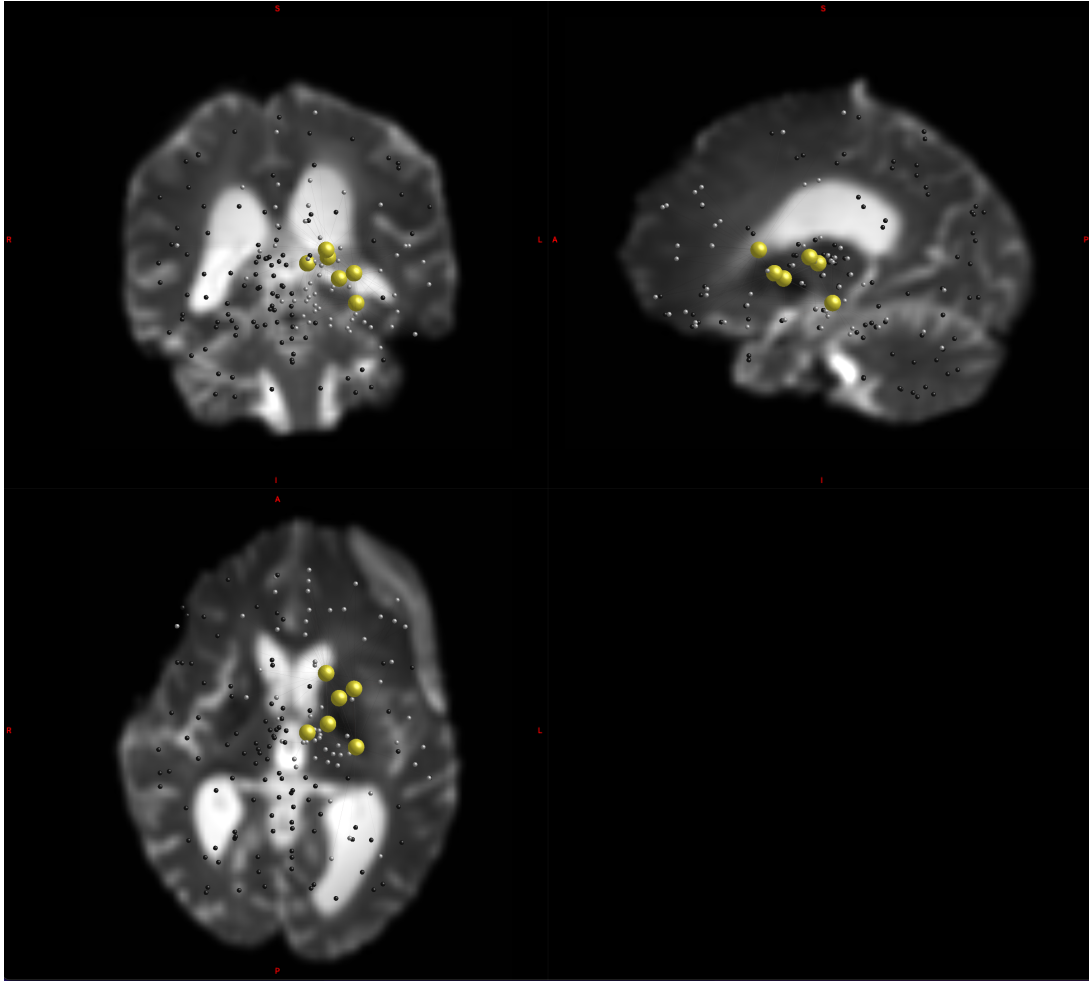


Figure 52: DWI visualization of significant brain regions in patients in a vegetative state (VS) as shown in Table 20. The yellow nodes represent the primary areas involved in sensory integration, motor coordination, and memory processing, with a strong focus on the thalamus, basal ganglia, and hippocampus, which are severely impaired in VS patients.

MCS patients (Figure 53) show a broader range of involvement, especially in thalamic and brainstem structures, indicative of residual awareness.

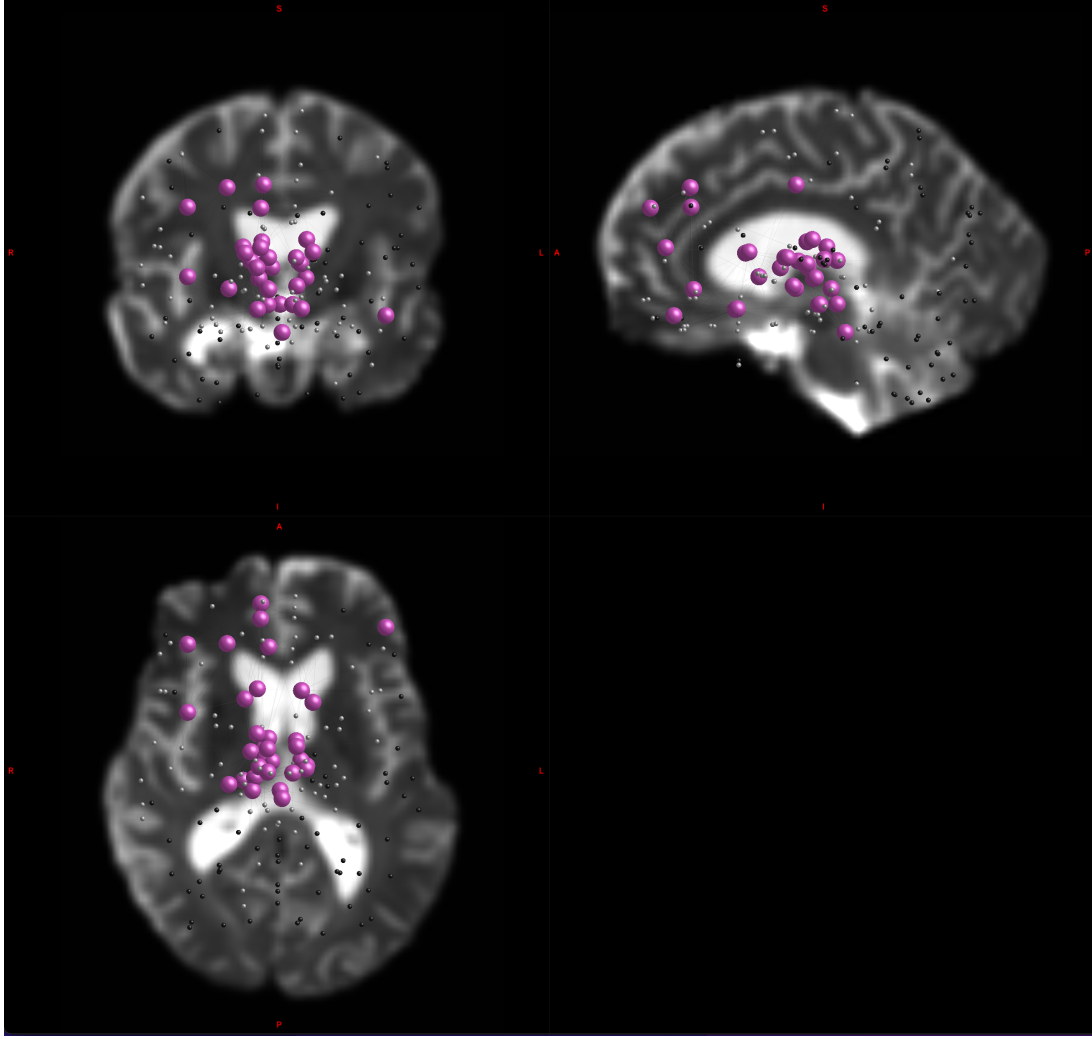


Figure 53: DWI visualization of significant brain regions in patients in a minimally conscious state (MCS). The pink nodes highlight thalamic nuclei, the insula, and brainstem regions, which are more extensively engaged compared to the VS group, reflecting the preserved but minimal awareness in MCS patients, as listed in Table 19.

Patients who regained consciousness (Figure 54) show re-engagement of regions critical for executive and integrative processes, supporting the hypothesis of functional restoration.

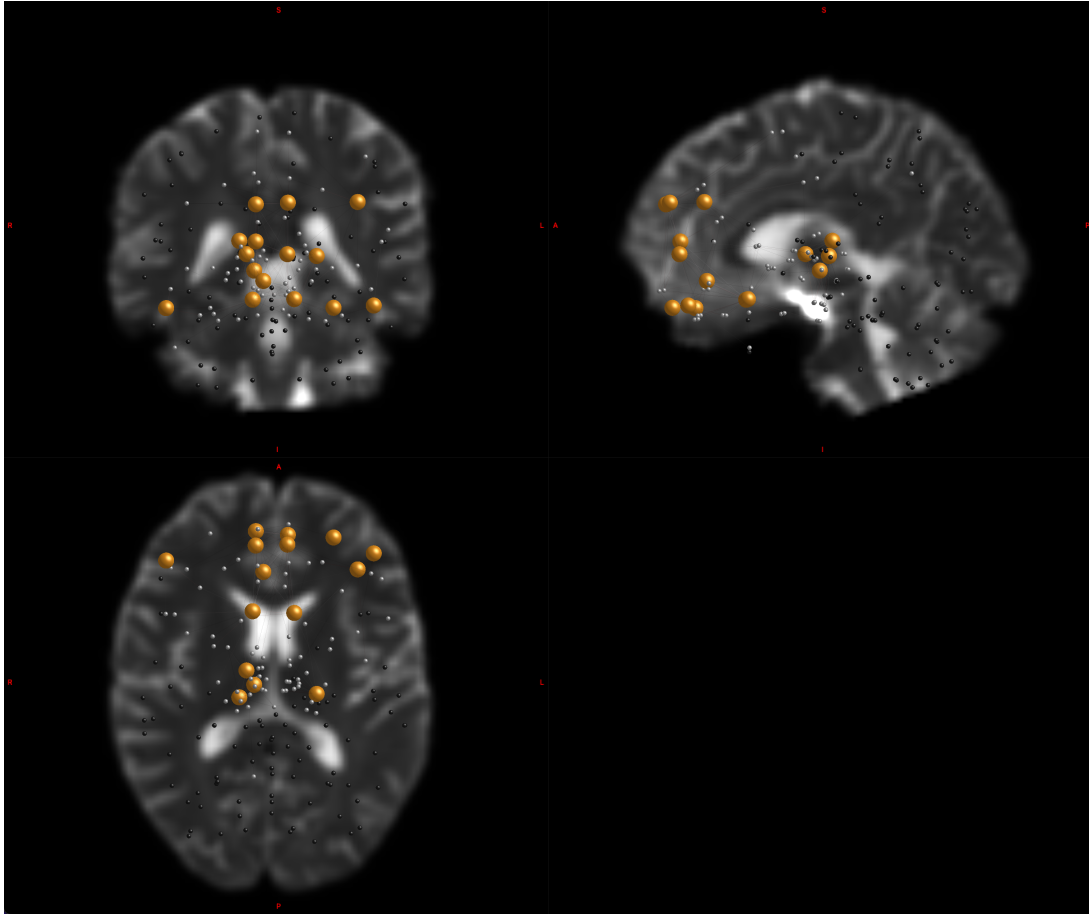


Figure 54: DWI visualization of significant brain regions in chronic patients who regained consciousness after TBI. The orange nodes show areas concentrated in the frontal cortex, thalamus, and anterior cingulate cortex, regions integral to executive functions, sensory processing, and attention regulation, indicating ongoing recovery of neural function in these patients, as outlined in Table [21](#).

Taken together, these results highlight the dynamic nature of brain network changes following TBI and demonstrate how interpretable ML models can provide clinically meaningful insights into neural function and recovery.

Table 16: Grouping of significant brain regions in acute patients

Anatomical Area	Regions
Frontal Cortex	Frontal_Sup_2_R, Frontal_Mid_2_R, Frontal_Inf_Orb_2_L, Frontal_Inf_Orb_2_R, Frontal_Med_Orb_L, Frontal_Med_Orb_R, Rectus_L, Rectus_R, OFCmed_L, OFCmed_R, OFCant_L, OFCpost_L, OFClat_L, OFClat_R
Olfactory	Olfactory_L, Olfactory_R
Insula	Insula_R
Basal Ganglia	Caudate_L, Putamen_R
Temporal Pole	Temporal_Pole_Sup_L, Temporal_Pole_Mid_L, Temporal_Pole_Mid_R
Cerebellum	Cerebellum_7b_L, Cerebellum_7b_R, Cerebellum_10_L, Vermis_1_2, Vermis_9, Vermis_10
Thalamus	Thal_LP_R, Thal_Re_L, Thal_Re_R, Thal_MDI_L
Anterior Cingulate Cortex	ACC_sub_L, ACC_sub_R, ACC_pre_L, ACC_pre_R, ACC_sup_R
Midbrain	VTA_L, LC_L, LC_R

Table 17: Grouping of significant brain regions in control group

Anatomical Area	Regions
Frontal Cortex	Frontal_Sup_2_L, Frontal_Sup_2_R, Frontal_Mid_2_L, Frontal_Inf_Tri_L, Frontal_Inf_Orb_2_L, Frontal_Inf_Orb_2_R, Frontal_Sup_Medial_L, Frontal_Sup_Medial_R, Frontal_Med_Orb_L, Frontal_Med_Orb_R, Rectus_L, Rectus_R, OFCmed_L, OFCmed_R, OFCant_L, OFCant_R, OFCpost_L, OFCpost_R, OFClat_L
Olfactory	Olfactory_L, Olfactory_R
Basal Ganglia	Putamen_L
Anterior Cingulate Cortex	ACC_sub_L, ACC_sub_R, ACC_pre_L, ACC_pre_R, ACC_sup_L, ACC_sup_R
Nucleus Accumbens	N_Acc_L, N_Acc_R

Table 18: Grouping of significant brain regions in chronic patients

Anatomical Area	Regions
Frontal Cortex	Frontal_Mid_2_L, Frontal_Sup_Medial_L, Frontal_Sup_Medial_R
Olfactory	Olfactory_L, Olfactory_R
Orbitofrontal Cortex	OFCant_L, OFClat_L, OFClat_R
Cingulate Cortex	Cingulate_Mid_R
Basal Ganglia	Caudate_L, Caudate_R
Thalamus	Thal_LP_L, Thal_LP_R, Thal_VL_R, Thal_VPL_L, Thal_IL_R, Thal_MDI_R, Thal_PuM_R, Thal_PuL_R
Anterior Cingulate Cortex	ACC_sub_R, ACC_pre_L, ACC_pre_R

Table 19: Significant brain regions in chronic patients (Minimally Conscious State)

Anatomical Area	Regions
Frontal Cortex	Frontal_Sup_2_R, Frontal_Mid_2_R, Frontal_Sup_Medial_R
Olfactory	Olfactory_L, Olfactory_R
Orbitofrontal Cortex	OFClat_L
Insula	Insula_R
Cingulate Cor- tex	Cingulate_Mid_R
Basal Ganglia	Caudate_L, Caudate_R
Thalamus	Thal_AV_L, Thal_AV_R, Thal_LP_L, Thal_LP_R, Thal_VA_R, Thal_VL_R, Thal_IL_L, Thal_IL_R, Thal_Re_L, Thal_Re_R, Thal_MDm_R, Thal_MDl_L, Thal_MDl_R, Thal_LGN_R, Thal_PuM_R, Thal_PuA_R, Thal_PuL_R
Anterior Cingu- late Cortex	ACC_sub_R, ACC_pre_R
Red Nucleus	Red_N_L, Red_N_R
Raphe	Raphe_D, Raphe_M

Table 20: Significant brain regions in chronic patients (Vegetative State)

Anatomical Area	Regions
Hippocampus	Hippocampus_L
Basal Ganglia	Caudate_L, Putamen_L, Pallidum_L
Thalamus	Thal_VL_L, Thal_MDm_L

Table 21: Grouping of significant brain regions in chronic patients who regained consciousness

Anatomical Area	Regions
Frontal Cortex	Frontal_Mid_2_L, Frontal_Sup_Medial_L, Frontal_Sup_Medial_R
Olfactory	Olfactory_L, Olfactory_R
Orbitofrontal Cortex	OFCant_L, OFClat_L, OF- Clat_R
Thalamus	Thal_VL_R, Thal_VPL_L, Thal_IL_R, Thal_PuL_R
Anterior Cingulate Cortex	ACC_sub_R, ACC_pre_L, ACC_pre_R

9.7.0.5 Discussion The outcomes of this study underscore the potential of GCNs as powerful tools for the classification and interpretation of structural brain connectomes in the context of TBI. Through a combination of advanced graph-based modelling and explainability techniques, we were able not only to discriminate among clinical categories, namely acute, chronic, and healthy control subjects, but also to gain insights into the neuroanatomical underpinnings of consciousness disorders.

The classification performance achieved by the proposed GCN model was robust, with an average accuracy of 83.67% across LOOCV iterations. This result is particularly noteworthy considering the complexity and heterogeneity of the clinical data involved. Precision and recall scores of 81.6% and 78%, respectively, highlight the model’s balanced performance in minimizing both false positives and false negatives, thus ensuring reliable and meaningful clinical predictions. From a clinical perspective, such balanced performance is crucial: high precision reduces the risk of misclassifying healthy individuals as impaired, while high recall ensures the inclusion of true cases in the diagnostic process, minimizing the risk of missed diagnoses in critical conditions such as DoC. The LOOCV framework played a central role in ensuring that the reported metrics reflect the model’s generalization capacity across individual patients. By training and testing on all possible partitions, LOOCV simulates real-world scenarios in which new patients with unique connectivity patterns must be classified without prior exposure to similar profiles. The observation that certain folds exhibited lower performance suggests the presence of outlier cases or particularly atypical connectome structures. These deviations are likely indicative of individual differences in injury severity, compensatory plasticity, or comorbid neurological conditions, and they highlight the need for adaptive models capable of capturing this inter-subject variability. A major contribution of this work lies in the integration of explainability methods, specifically Grad-CAM, into the graph-based classification framework. The application of Grad-CAM enabled the identification of the most influential nodes (i.e., brain regions) that guided the model’s decisions. This step is essential for bridging the gap between the black-box nature of neural networks and the need for interpretable clinical tools. Unlike conventional feature importance techniques that rely on aggregate statistical scores, Grad-CAM allows for subject-specific heatmaps that reveal how connectivity patterns within each individual’s brain network relate to their predicted diagnosis.

In the acute cohort, regions such as the thalamus, frontal cortex, and brainstem (including the locus coeruleus and ventral tegmental area) emerged consistently among the top contributors. This finding aligns with current neuroscientific literature highlighting the centrality of thalamo-cortical and brainstem pathways in sustaining consciousness and modulating arousal [204, 205]. The thalamus, acting as a relay station for both sensory and motor signals, is known to support integration across large-scale brain networks, and its disruption has been strongly associated with impaired consciousness in TBI. Similarly, the frontal cortex is implicated in executive functioning, working memory, and attentional control, domains frequently compromised following traumatic injury. The consistent involvement of these regions in acute patients reflects the extensive and diffuse network breakdown typical of early-stage TBI, where structural disconnection interferes with both local and global information integration.

The analysis of chronic patients, in contrast, revealed a more focused pattern of disruptions. While the thalamus and anterior cingulate cortex (ACC) remained prominent, the overall distribution of important nodes suggested a shift toward more localized impairments. This evolution may represent compensatory reorganization processes that emerge over time, as the brain attempts to restore function via plastic adaptation. However, the persistence of dysfunction in these core regions points to the limitations of such compensation, particularly in areas responsible for integrating cognitive and emotional processes. From a therapeutic standpoint, these findings may inform long-term rehabilitation strategies, with a focus on promoting functional connectivity within networks that remain vulnerable to chronic disintegration. The study further distinguished between subgroups within the chronic cohort, offering valuable insights into the neural correlates of distinct clinical outcomes. In patients diagnosed with a MCS, we observed broader engagement of associative regions including the insula, thalamic nuclei, and brainstem components such as the red nucleus and raphe nuclei. These regions are involved in awareness, emotional salience, and regulation of autonomic functions, and their relative preservation may explain the intermittent signs of intentional behaviour observed clinically in MCS [206]. In contrast, the VS subgroup exhibited a markedly sparser distribution of relevant regions, primarily limited to subcortical structures such as the hippocampus, basal ganglia, and medial thalamus. This pattern of restricted engagement likely reflects the profound network disconnection associated with the absence of awareness and volitional behaviour, consistent with the diagnostic criteria for VS [162].

Another crucial observation emerged from the analysis of chronic patients who regained consciousness. In these individuals, nodes such as the frontal cortex, thalamus, and ACC were again highlighted as key contributors, suggesting that reactivation of these regions is associated with recovery trajectories. The restoration of these connectivity hubs may signal the brain's partial reestablishment of large-scale integrative functions, allowing for the emergence of conscious awareness and purposeful behaviour. These results support the hypothesis that successful recovery from DoC involves not merely localized tissue repair but also the re-engagement of distributed network architectures. Consequently, therapeutic approaches that aim to enhance structural and functional connectivity in these circuits, such as neuromodulation or task-based cognitive training, may hold promise for accelerating recovery in selected patients. Visualizations of node frequencies and spatial maps further enriched this interpretation, allowing a group-level comparison of the most consistently involved regions across clinical conditions. In acute patients, widespread and variable patterns suggest an ongoing network collapse and early-stage compensation. In contrast, chronic and recovered patients showed more organized, though still disrupted,

patterns of engagement. These transitions, observed through both Grad-CAM maps and histogram distributions, reflect the dynamic nature of brain network reorganization and provide a framework for studying how different neural trajectories unfold over time. The control group offered a normative baseline, characterized by coherent and spatially concentrated patterns of significant regions, especially within the frontal cortex and olfactory areas. The strong topological organization in healthy individuals highlights the extent of deviation observed in TBI patients and reinforces the utility of graph-based approaches for quantifying such alterations. Moreover, the marked differences in network structure between controls and all patient subgroups underline the relevance of structural connectivity as a biomarker for consciousness-related disorders. Taken together, these findings contribute to the growing body of literature supporting the application of graph theory and deep learning in clinical neuroscience. By uniting high-performing models with explainability tools, this study demonstrates a pathway toward interpretable, individualized predictions based on neuroimaging data. However, several challenges remain. The sample size, while adequate for exploratory analysis, limits the generalizability of the model. Future work should focus on expanding the dataset and incorporating longitudinal data to capture the evolution of brain networks over time. Additionally, integrating multimodal data, such as functional MRI, EEG, and clinical assessments, may enhance the robustness and interpretability of predictive models. In conclusion, this research shows that combining GCN-based classification with explainable AI offers a dual advantage: accurate patient stratification and insight into the neural correlates of TBI and consciousness disorders. The identification of key brain regions implicated in different clinical trajectories opens avenues for targeted interventions and longitudinal monitoring, ultimately contributing to more personalized and effective clinical care for individuals affected by TBI and DoC.

10 Conclusion and Future Directions

This doctoral thesis has introduced and validated DicomOS, a Linux-based operating system specifically designed for the management, analysis, and advanced processing of medical imaging data. The system was conceived to address the limitations of current software tools in clinical and research settings, providing a unified, open source, and extensible platform to handle DICOM images and integrate AI tools within a cohesive and user-friendly environment. Its architecture is tailored to support both command-line operations and graphical interfaces, enabling seamless access for developers, researchers, and clinicians alike.

At the heart of DicomOS lies the ambition to lower the technical barrier that often separates the clinical world from the computational one. By embedding preconfigured modules for image visualization, conversion, annotation, segmentation, and preprocessing, along with native DICOM handling and customizable pipelines, the system positions itself as a bridge between daily radiological practice and cutting-edge AI research. This thesis not only introduced DicomOS as a standalone operating system but also demonstrated its practical versatility through a series of integrated AI applications, each addressing different clinical challenges. These applications were designed, implemented, and tested within the DicomOS environment, showcasing the platform’s adaptability and its ability to host advanced workflows, from ML model training to explainability and visual analytics. In the oncological domain, the thesis explored the use of ViT for the classification of skin lesions and breast cancer, incorporating data augmentation strategies, both geometric and generative (e.g., diffusion models), to improve robustness. These experiments were carried out on publicly available datasets such as ISIC 2017 and CBIS-DDSM, confirming the viability of transformer-based architectures for medical image classification, even in data-limited regimes. The design and evaluation of these models also contributed to a better understanding of how XAI techniques can be integrated into deep learning frameworks to ensure interpretability and clinical trust.

In the neuroscience and connectomics section, DicomOS served as the foundation for a novel study on TBI using GNNs. Here, DWI-derived brain connectivity matrices were modelled as graphs, with GCNs used to classify patients in acute, chronic, and control cohorts. The use of node-level topological features (e.g., degree, PageRank, clustering coefficient) and advanced explainability tools such as SHAP and Grad-CAM enabled both accurate predictions and biologically meaningful insights into the structural reorganization of brain networks following trauma. Through these experiments, the thesis illustrated how deep learning models, if properly designed and interpreted, can support both diagnostic accuracy and neuroscientific discovery. From a broader perspective, the integration of XAI methods across all applications represents a unifying thread in the thesis. Techniques such as SHAP and Grad-CAM were applied not only to highlight the inner workings of the models, but also to uncover clinically relevant patterns, whether in the detection of aggressive skin lesions, the identification of significant breast tissue areas or the location of brain regions disrupted by TBI. This commitment to transparency and interpretability reflects a fundamental principle of responsible AI in healthcare, particularly in high-stakes environments where decisions must be both accurate and explainable. The contributions of this thesis are therefore threefold. First, it delivers a novel, lightweight operating system, DicomOS, explicitly built for the needs of radiology and medical imaging research. Second, it implements a series of AI-driven applications, grounded in clinical relevance and aligned with the principles of explainability and reproducibility. Third, it proposes a vision

of AI integration that is modular, interpretable, and embedded within the infrastructure, rather than outside it.

In conclusion, this work demonstrates that by uniting operating system design with AI and XAI, it is possible to construct tools that are not only technologically sophisticated but also aligned with clinical workflows and scientific inquiry. DicomOS is not just an OS, it is a framework for innovation, built to evolve alongside the needs of medicine and research, and to serve as a catalyst for the responsible integration of AI into real-world healthcare applications.

References

- [1] S. Möller, H. N. Krabbenhöft, A. Tille, D. Paleino, *et al.*, “Community-driven computational biology with debian linux,” *BMC Bioinformatics*, vol. 11, no. 12, p. S5, 2010.
- [2] M. J. Cardoso, W. Li, R. Brown, N. Ma, E. Kerfoot, Y. Wang, B. Murrey, A. Myronenko, C. Zhao, D. Yang, *et al.*, “Monai: An open-source framework for deep learning in healthcare,” *arXiv preprint arXiv:2211.02701*, 2022.
- [3] T. Currier, O. Gambino, R. Pirrone, and S. Vitabile, “Dicomos: A preliminary study on a linux-based operating system tailored for medical imaging and enhanced interoperability in radiology workflows,” *Electronics*, vol. 14, no. 2, p. 330, 2025.
- [4] Y. O. Halchenko and M. Hanke, “Open is not enough. let’s take the next step: an integrated, community-driven computing platform for neuroscience,” *Frontiers in neuroinformatics*, vol. 6, p. 22, 2012.
- [5] K. Nemoto, I. Dan, C. Rorden, T. Ohnishi, D. Tsuzuki, M. Okamoto, F. Yamashita, and T. Asada, “Lin4neuro: a customized linux distribution ready for neuroimaging analysis,” *BMC medical imaging*, vol. 11, pp. 1–7, 2011.
- [6] O. S. Pianykh, *Digital imaging and communications in medicine (DICOM): a practical introduction and survival guide*, vol. 10. Springer, 2012.
- [7] U. Bick and H. Lenzen, “Pacs: the silent revolution,” *European Radiology*, vol. 9, pp. 1152–1160, 1999.
- [8] C. E. Kahn, J. A. Carrino, M. J. Flynn, D. J. Peck, and S. C. Horii, “Dicom and radiology: past, present, and future,” *Journal of the American College of Radiology*, vol. 4, no. 9, pp. 652–657, 2007.
- [9] H. Pereira, L. Romero, and P. Miguel Faria, “Web-based dicom viewers: A survey and a performance classification,” *Journal of Imaging Informatics in Medicine*, pp. 1–19, 2024.
- [10] S. Pieper, M. Halle, and R. Kikinis, “3d slicer,” in *2004 2nd IEEE international symposium on biomedical imaging: nano to macro (IEEE Cat No. 04EX821)*, pp. 632–635, IEEE, 2004.
- [11] O. Ratib and A. Rosset, “Open-source software in medical imaging: development of osirix,” *International Journal of Computer Assisted Radiology and Surgery*, vol. 1, pp. 187–196, 2006.

- [12] N. Hasani, F. Farhadi, M. A. Morris, M. Nikpanah, A. Rhamim, Y. Xu, A. Pariser, M. T. Collins, R. M. Summers, E. Jones, *et al.*, “Artificial intelligence in medical imaging and its impact on the rare disease community: threats, challenges and opportunities,” *PET clinics*, vol. 17, no. 1, p. 13, 2022.
- [13] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, “A survey on deep learning in medical image analysis,” *Medical image analysis*, vol. 42, pp. 60–88, 2017.
- [14] A. Esteva, A. Robicquet, B. Ramsundar, V. Kuleshov, M. DePristo, K. Chou, C. Cui, G. Corrado, S. Thrun, and J. Dean, “A guide to deep learning in healthcare,” *Nature medicine*, vol. 25, no. 1, pp. 24–29, 2019.
- [15] E. J. Topol, “High-performance medicine: the convergence of human and artificial intelligence,” *Nature medicine*, vol. 25, no. 1, pp. 44–56, 2019.
- [16] A. Holzinger, C. Biemann, C. S. Pattichis, and D. B. Kell, “What do we need to build explainable ai systems for the medical domain?,” *arXiv preprint arXiv:1712.09923*, 2017.
- [17] E. A. Krupinski and K. S. Berbaum, “Improving human–computer interaction in radiology: exploring new input modalities and their potential impact on clinical practice,” *Journal of the American College of Radiology*, vol. 7, no. 9, pp. 663–668, 2010.
- [18] C. M. Johnson, T. R. Johnson, and J. Zhang, “Designing and evaluating clinical user interfaces: a systematic literature review,” *Journal of Biomedical Informatics*, vol. 116, p. 103717, 2021.
- [19] F. Ehrler and C. Lovis, “Human-centered design for health information technology: development of an interactive dicom browser for radiology education,” *JMIR Medical Informatics*, vol. 7, no. 3, p. e12301, 2019.
- [20] B. A. Myers, “A brief history of human-computer interaction technology,” *interactions*, vol. 5, no. 2, pp. 44–54, 1998.
- [21] M. Gampe, S. Biffl, and S. Stöckl, “User interface design patterns and software quality: a systematic literature review,” *Software Quality Journal*, vol. 28, no. 3, pp. 961–1004, 2020.
- [22] J. A. Carrino, P. J. Unkel, A. J. Duerinckx, P. A. Johnson, and G. McGuinness, “American college of radiology white paper on teleradiology,” *Journal of the American College of Radiology*, vol. 1, no. 2, pp. 124–131, 2004.
- [23] “Ubuntu: The official ubuntu documentation.” <https://help.ubuntu.com/>.
- [24] C. M. Johnson, T. R. Johnson, and J. Zhang, “A user-centered framework for redesigning health care interfaces,” *Journal of biomedical informatics*, vol. 38, no. 1, pp. 75–87, 2005.
- [25] A. Fedorov, R. Beichel, J. Kalpathy-Cramer, J. Finet, J.-C. Fillion-Robin, S. Pujol, C. Bauer, D. Jennings, F. Fennessy, M. Sonka, *et al.*, “3d slicer as an image computing platform for the quantitative imaging network,” *Magnetic resonance imaging*, vol. 30, no. 9, pp. 1323–1341, 2012.

- [26] Oracle Corporation, “Oracle vm virtualbox version 6.1.” <https://www.virtualbox.org/>, 2023.
- [27] C. L. Ventola, “Mobile devices and apps for health care professionals: uses and benefits,” *Pharmacy and Therapeutics*, vol. 39, no. 5, p. 356, 2014.
- [28] “Plymouth theming guide.” <https://www.freedesktop.org/wiki/Software/Plymouth/Theming/>. Accessed: October 10, 2023.
- [29] P. A. Yushkevich, Y. Gao, and G. Gerig, “Itk-snap: An interactive tool for semi-automatic segmentation of multi-modality biomedical images,” in *2016 38th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, pp. 3342–3345, IEEE, 2016.
- [30] R. Kikinis, S. D. Pieper, and K. G. Vosburgh, “3d slicer: A platform for subject-specific image analysis, visualization, and clinical support,” 2014.
- [31] P. A. Yushkevich, A. Pashchinskiy, I. Oguz, S. Mohan, J. E. Schmitt, J. M. Stein, D. Zukić, J. Vicory, M. McCormick, N. Yushkevich, *et al.*, “User-guided segmentation of multi-modality medical imaging datasets with itk-snap,” *Neuroinformatics*, vol. 17, pp. 83–102, 2019.
- [32] A. W. Kushniruk and V. L. Patel, “Cognitive and usability engineering methods for the evaluation of clinical information systems,” *Journal of biomedical informatics*, vol. 37, no. 1, pp. 56–76, 2004.
- [33] F. Lundh, “An introduction to tkinter,” *URL: www.pythonware.com/library/tkinter/introduction/index.htm*, vol. 539, p. 540, 1999.
- [34] E. Bisong and E. Bisong, “Matplotlib and seaborn,” *Building machine learning and deep learning models on google cloud platform: A comprehensive guide for beginners*, pp. 151–165, 2019.
- [35] T. pydicom community, *pydicom: DICOM handling in Python*.
- [36] P. D. Team, “Pyinstaller: Package python applications,” 2024. Version 5.13, accessed on December 6, 2024.
- [37] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, *et al.*, “The fair guiding principles for scientific data management and stewardship,” *Scientific data*, vol. 3, no. 1, pp. 1–9, 2016.
- [38] D. Brunet, E. R. Vrscay, and Z. Wang, “On the mathematical properties of the structural similarity index,” *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 1488–1499, 2011.
- [39] R. Bammer, “Basic principles of diffusion-weighted imaging,” *European journal of radiology*, vol. 45, no. 3, pp. 169–184, 2003.
- [40] X. Zhang, J. Huang, Y. Yang, X. He, R. Liu, and N. Zhong, “Applying python in brain science education,” in *2019 International Joint Conference on Information, Media and Engineering (IJCIME)*, pp. 396–400, IEEE, 2019.

- [41] G. Bradski, A. Kaehler, *et al.*, “Opencv,” *Dr. Dobb’s journal of software tools*, vol. 3, no. 2, 2000.
- [42] B. C. Lowekamp, D. T. Chen, L. Ibáñez, and D. Blezek, “The design of simpleitk,” *Frontiers in neuroinformatics*, vol. 7, p. 45, 2013.
- [43] M. Kerrisk, *The Linux programming interface: a Linux and UNIX system programming handbook*. No Starch Press, 2010.
- [44] A. Casey, E. Davidson, M. Poon, H. Dong, D. Duma, A. Grivas, C. Grover, V. Suárez-Paniagua, R. Tobin, W. Whiteley, *et al.*, “A systematic review of natural language processing applied to radiology reports,” *BMC medical informatics and decision making*, vol. 21, no. 1, p. 179, 2021.
- [45] Y. Singh, J. B. Andersen, Q. Hathaway, S. K. Venkatesh, G. J. Gores, and B. Erickson, “Deep learning-based uncertainty quantification for quality assurance in hepatobiliary imaging-based techniques,” *Oncotarget*, vol. 16, p. 249, 2025.
- [46] L. Emma, “Ai in radiology education: Training the next generation of radiologists,” 2025.
- [47] M. T. Contaldo, G. Pasceri, G. Vignati, L. Bracchi, S. Triggiani, and G. Carrafiello, “Ai in radiology: navigating medical responsibility,” *Diagnostics*, vol. 14, no. 14, p. 1506, 2024.
- [48] T. Ciecierski-Holmes, R. Singh, M. Axt, S. Brenner, and S. Barteit, “Artificial intelligence for strengthening healthcare systems in low-and middle-income countries: a systematic scoping review,” *NPJ digital medicine*, vol. 5, no. 1, p. 162, 2022.
- [49] S. M. Shah, R. A. Khan, S. Arif, and U. Sajid, “Artificial intelligence for breast cancer analysis: Trends & directions,” *Computers in Biology and Medicine*, vol. 142, p. 105221, 2022.
- [50] C. Martin-Isla, V. M. Campello, C. Izquierdo, Z. Raisi-Estabragh, B. Baeßler, S. E. Petersen, and K. Lekadir, “Image-based cardiac diagnosis with machine learning: a review,” *Frontiers in cardiovascular medicine*, vol. 7, p. 1, 2020.
- [51] X. Liang, X. Yu, and T. Gao, “Machine learning with magnetic resonance imaging for prediction of response to neoadjuvant chemotherapy in breast cancer: A systematic review and meta-analysis,” *European Journal of Radiology*, vol. 150, p. 110247, 2022.
- [52] A. Hosny, C. Parmar, J. Quackenbush, L. H. Schwartz, and H. J. Aerts, “Artificial intelligence in radiology,” *Nature Reviews Cancer*, vol. 18, no. 8, pp. 500–510, 2018.
- [53] M. H. Rezazade Mehrizi, P. van Ooijen, and M. Homan, “Applications of artificial intelligence (ai) in diagnostic radiology: a technography study,” *European radiology*, vol. 31, pp. 1805–1811, 2021.
- [54] Y. Xu, Y. Zhou, P. Sekula, and L. Ding, “Machine learning in construction: From shallow to deep learning,” *Developments in the built environment*, vol. 6, p. 100045, 2021.

- [55] P. Lambin, E. Rios-Velazquez, R. Leijenaar, S. Carvalho, R. G. Van Stiphout, P. Granton, C. M. Zegers, R. Gillies, R. Boellard, A. Dekker, *et al.*, “Radiomics: extracting more information from medical images using advanced feature analysis,” *European journal of cancer*, vol. 48, no. 4, pp. 441–446, 2012.
- [56] R. J. Gillies, P. E. Kinahan, and H. Hricak, “Radiomics: images are more than pictures, they are data,” *Radiology*, vol. 278, no. 2, pp. 563–577, 2016.
- [57] B. J. Erickson, P. Korfiatis, Z. Akkus, and T. L. Kline, “Machine learning for medical imaging,” *Radiographics*, vol. 37, no. 2, pp. 505–515, 2017.
- [58] C. Militello, L. Rundo, M. Dimarco, A. Orlando, I. D’Angelo, V. Conti, and T. Bartolotta, “Robustness analysis of dce-mri-derived radiomic features in breast masses: Assessing quantization levels and segmentation agreement,” *Applied Sciences*, vol. 12, no. 11, p. 5512, 2022.
- [59] G. Luo, “A review of automatic selection methods for machine learning algorithms and hyper-parameter values,” *Network Modeling Analysis in Health Informatics and Bioinformatics*, vol. 5, pp. 1–16, 2016.
- [60] P. Probst, A.-L. Boulesteix, and B. Bischl, “Tunability: Importance of hyperparameters of machine learning algorithms,” *The Journal of Machine Learning Research*, vol. 20, no. 1, pp. 1934–1965, 2019.
- [61] R. Reed and R. J. MarksII, *Neural smithing: supervised learning in feedforward artificial neural networks*. Mit Press, 1999.
- [62] K. J. Nam, H. Park, E. S. Ko, Y. Lim, H.-H. Cho, and J. E. Lee, “Radiomics signature on 3t dynamic contrast-enhanced magnetic resonance imaging for estrogen receptor-positive invasive breast cancers: Preliminary results for correlation with oncotype dx recurrence scores,” *Medicine*, vol. 98, no. 23, 2019.
- [63] J. R. F. Junior, M. Koenigkam-Santos, F. E. G. Cipriano, A. T. Fabro, and P. M. de Azevedo-Marques, “Radiomics-based features for pattern recognition of lung cancer histopathology and metastases,” *Computer methods and programs in biomedicine*, vol. 159, pp. 23–30, 2018.
- [64] S.-H. Lee, H. Park, and E. S. Ko, “Radiomics in breast imaging from techniques to clinical applications: a review,” *Korean journal of radiology*, vol. 21, no. 7, p. 779, 2020.
- [65] M. Liu, N. Mao, H. Ma, J. Dong, K. Zhang, K. Che, S. Duan, X. Zhang, Y. Shi, and H. Xie, “Pharmacokinetic parameters and radiomics model based on dynamic contrast enhanced mri for the preoperative prediction of sentinel lymph node metastasis in breast cancer,” *Cancer Imaging*, vol. 20, pp. 1–8, 2020.
- [66] J. Zhou, Y. Zhang, K.-T. Chang, K. E. Lee, O. Wang, J. Li, Y. Lin, Z. Pan, P. Chang, D. Chow, *et al.*, “Diagnosis of benign and malignant breast lesions on dce-mri by using radiomics and deep learning with consideration of peritumor tissue,” *Journal of Magnetic Resonance Imaging*, vol. 51, no. 3, pp. 798–809, 2020.

- [67] B. Kocak, B. Baessler, S. Bakas, R. Cuocolo, A. Fedorov, L. Maier-Hein, N. Mercaldo, H. Müller, F. Orlhac, D. Pinto dos Santos, *et al.*, “Checklist for evaluation of radiomics research (clear): a step-by-step reporting guideline for authors and reviewers endorsed by esr and eusomii,” *Insights into Imaging*, vol. 14, no. 1, pp. 1–13, 2023.
- [68] J. Mongan, L. Moy, and C. E. Kahn Jr, “Checklist for artificial intelligence in medical imaging (claim): a guide for authors and reviewers,” 2020.
- [69] S. Santhosh Baboo and M. Amirthapriya, “Comparison of machine learning techniques on twitter emotions classification,” *SN Computer Science*, vol. 3, pp. 1–8, 2022.
- [70] M. Vallieres, E. Kay-Rivest, L. J. Perrin, X. Liem, C. Furstoss, H. J. Aerts, N. Khaouam, P. F. Nguyen-Tan, C.-S. Wang, K. Sultanem, *et al.*, “Radiomics strategies for risk assessment of tumour failure in head-and-neck cancer,” *Scientific reports*, vol. 7, no. 1, p. 10117, 2017.
- [71] M. Wang, Z. Feng, L. Zhou, L. Zhang, X. Hao, and J. Zhai, “Computed-tomography-based radiomics model for predicting the malignant potential of gastrointestinal stromal tumors preoperatively: a multi-classifier and multicenter study,” *Frontiers in Oncology*, vol. 11, p. 582847, 2021.
- [72] S. Wu, J. Zheng, Y. Li, Z. Wu, S. Shi, M. Huang, H. Yu, W. Dong, J. Huang, and T. Lin, “Development and validation of an mri-based radiomics signature for the preoperative prediction of lymph node metastasis in bladder cancer,” *EBioMedicine*, vol. 34, pp. 76–84, 2018.
- [73] C. Cortes and V. Vapnik, “Support-vector networks,” *Machine learning*, vol. 20, pp. 273–297, 1995.
- [74] T. Joachims, *Learning to classify text using support vector machines*, vol. 668. Springer Science & Business Media, 2002.
- [75] L. Rokach, “Ensemble-based classifiers,” *Artificial intelligence review*, vol. 33, pp. 1–39, 2010.
- [76] V. Podgorelec, P. Kokol, B. Stiglic, and I. Rozman, “Decision trees: an overview and their use in medicine,” *Journal of medical systems*, vol. 26, pp. 445–463, 2002.
- [77] M. M. Ghiasi and S. Zendehboudi, “Application of decision tree-based ensemble learning in the classification of breast cancer,” *Computers in biology and medicine*, vol. 128, p. 104089, 2021.
- [78] L. Breiman, “Random forests,” *Machine learning*, vol. 45, pp. 5–32, 2001.
- [79] T. M. Oshiro, P. S. Perez, and J. A. Baranauskas, “How many trees in a random forest?,” in *Machine Learning and Data Mining in Pattern Recognition: 8th International Conference, MLDM 2012, Berlin, Germany, July 13-20, 2012. Proceedings* 8, pp. 154–168, Springer, 2012.

- [80] P. Probst and A.-L. Boulesteix, “To tune or not to tune the number of trees in random forest,” *The Journal of Machine Learning Research*, vol. 18, no. 1, pp. 6673–6690, 2017.
- [81] J. H. Friedman, “Stochastic gradient boosting,” *Computational statistics & data analysis*, vol. 38, no. 4, pp. 367–378, 2002.
- [82] T. Chen and C. Guestrin, “Xgboost: A scalable tree boosting system,” in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785–794, 2016.
- [83] R. P. Sheridan, W. M. Wang, A. Liaw, J. Ma, and E. M. Gifford, “Extreme gradient boosting as a method for quantitative structure–activity relationships,” *Journal of chemical information and modeling*, vol. 56, no. 12, pp. 2353–2360, 2016.
- [84] T. Davies, J. C. Y. Louie, R. Ndanuko, S. Barbieri, O. Perez-Concha, and J. H. Wu, “A machine learning approach to predict the added-sugar content of packaged foods,” *The Journal of Nutrition*, vol. 152, no. 1, pp. 343–349, 2022.
- [85] H. A. Abu Alfeilat, A. B. Hassanat, O. Lasassmeh, A. S. Tarawneh, M. B. Alhasanat, H. S. Eyal Salman, and V. S. Prasath, “Effects of distance measure choice on k-nearest neighbor classifier performance: a review,” *Big data*, vol. 7, no. 4, pp. 221–248, 2019.
- [86] L. Deng and D. Yu, “Deep learning: Methods and applications.,” *Found. Trends Signal Process.*, vol. 7, no. 3-4, pp. 197–387, 2014.
- [87] W. S. McCulloch and W. Pitts, “A logical calculus of the ideas immanent in nervous activity,” *The bulletin of mathematical biophysics*, vol. 5, pp. 115–133, 1943.
- [88] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, “A survey of convolutional neural networks: analysis, applications, and prospects,” *IEEE transactions on neural networks and learning systems*, 2021.
- [89] A. Apicella, F. Donnarumma, F. Isgrò, and R. Prevete, “A survey on modern trainable activation functions,” *Neural Networks*, vol. 138, pp. 14–32, 2021.
- [90] K. Hornik, “Approximation capabilities of multilayer feedforward networks,” *Neural networks*, vol. 4, no. 2, pp. 251–257, 1991.
- [91] G. Cybenko, “Approximation by superpositions of a sigmoidal function,” *Mathematics of control, signals and systems*, vol. 2, no. 4, pp. 303–314, 1989.
- [92] F. Prinzi, C. Militello, V. Conti, and S. Vitabile, “Impact of wavelet kernels on predictive capability of radiomic features: A case study on covid-19 chest x-ray images,” *Journal of Imaging*, vol. 9, no. 2, p. 32, 2023.
- [93] S. M. Anwar, M. Majid, A. Qayyum, M. Awais, M. Alnowami, and M. K. Khan, “Medical image analysis using convolutional neural networks: A review,” *Journal of medical systems*, vol. 42, p. 226, October 2018.
- [94] J. Premaladha and K. Ravichandran, “Novel approaches for diagnosing melanoma skin lesions through supervised and deep learning algorithms,” *Journal of medical systems*, vol. 40, pp. 1–12, 2016.

- [95] P. Kharazmi, J. Zheng, H. Lui, Z. Jane Wang, and T. K. Lee, “A computer-aided decision support system for detection and localization of cutaneous vasculature in dermoscopy images via deep feature learning,” *Journal of medical systems*, vol. 42, p. 33, January 2018.
- [96] J. Coady, A. O’Riordan, G. Dooly, T. Newe, and D. Toal, “An overview of popular digital image processing filtering operations,” in *2019 13th International conference on sensing technology (ICST)*, pp. 1–5, IEEE, 2019.
- [97] R. Nirthika, S. Manivannan, A. Ramanan, and R. Wang, “Pooling in convolutional neural networks for medical image analysis: a survey and an empirical study,” *Neural Computing and Applications*, vol. 34, no. 7, pp. 5321–5347, 2022.
- [98] D. H. Wolpert and W. G. Macready, “No free lunch theorems for optimization,” *IEEE transactions on evolutionary computation*, vol. 1, no. 1, pp. 67–82, 1997.
- [99] M. Kamal, A. R. Pratap, M. Naved, A. S. Zamani, P. Nancy, M. Ritonga, S. K. Shukla, F. Sammy, *et al.*, “Machine learning and image processing enabled evolutionary framework for brain mri analysis for alzheimer’s disease detection,” *Computational Intelligence and Neuroscience*, vol. 2022, 2022.
- [100] R. Safdari, A. Deghatipour, M. Gholamzadeh, and K. Maghooli, “Applying data mining techniques to classify patients with suspected hepatitis c virus infection,” *Intelligent Medicine*, vol. 2, no. 04, pp. 193–198, 2022.
- [101] S. Kabiraj, M. Raihan, N. Alvi, M. Afrin, L. Akter, S. A. Sohagi, and E. Podder, “Breast cancer risk prediction using xgboost and random forest algorithm,” in *2020 11th international conference on computing, communication and networking technologies (ICCCNT)*, pp. 1–4, IEEE, 2020.
- [102] A. Adadi and M. Berrada, “Peeking inside the black-box: a survey on explainable artificial intelligence (xai),” *IEEE access*, vol. 6, pp. 52138–52160, 2018.
- [103] D. Gunning and D. Aha, “Darpa’s explainable artificial intelligence (xai) program,” *AI magazine*, vol. 40, no. 2, pp. 44–58, 2019.
- [104] A. B. Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. García, S. Gil-López, D. Molina, R. Benjamins, *et al.*, “Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai,” *Information fusion*, vol. 58, pp. 82–115, 2020.
- [105] M. T. Ribeiro, S. Singh, and C. Guestrin, “” 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*, pp. 1135–1144, 2016.
- [106] S. M. Lundberg and S.-I. Lee, “A unified approach to interpreting model predictions,” *Advances in neural information processing systems*, vol. 30, 2017.
- [107] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, “Grad-cam: Visual explanations from deep networks via gradient-based localization,” in *Proceedings of the IEEE international conference on computer vision*, pp. 618–626, 2017.

- [108] M. Sundararajan, A. Taly, and Q. Yan, “Axiomatic attribution for deep networks,” in *International conference on machine learning*, pp. 3319–3328, PMLR, 2017.
- [109] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, *et al.*, “An image is worth 16x16 words: Transformers for image recognition at scale,” *arXiv preprint arXiv:2010.11929*, 2020.
- [110] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, “Attention is all you need,” *Advances in neural information processing systems*, vol. 30, 2017.
- [111] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, “An image is worth 16x16 words: Transformers for image recognition at scale,” 2021.
- [112] D. Hendrycks and K. Gimpel, “Gaussian error linear units (gelus).” arxiv. doi: 10.48550, *arXiv preprint arXiv.1606.08415*, 2020.
- [113] C. Shorten and T. M. Khoshgoftaar, “A survey on image data augmentation for deep learning,” *Journal of big data*, vol. 6, no. 1, pp. 1–48, 2019.
- [114] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers, “Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2097–2106, 2017.
- [115] N. Tajbakhsh, L. Jeyaseelan, Q. Li, J. N. Chiang, Z. Wu, and X. Ding, “Embracing imperfect datasets: A review of deep learning solutions for medical image segmentation,” *Medical image analysis*, vol. 63, p. 101693, 2020.
- [116] Z.-Q. Zhao, P. Zheng, S.-t. Xu, and X. Wu, “Object detection with deep learning: A review,” *IEEE transactions on neural networks and learning systems*, vol. 30, no. 11, pp. 3212–3232, 2019.
- [117] T. DeVries and G. W. Taylor, “Improved regularization of convolutional neural networks with cutout,” *arXiv preprint arXiv:1708.04552*, 2017.
- [118] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, “mixup: Beyond empirical risk minimization,” *arXiv preprint arXiv:1710.09412*, 2017.
- [119] D. P. Kingma and M. Welling, “Auto-encoding variational bayes,” 2022.
- [120] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial networks,” 2014.
- [121] P. Dhariwal and A. Nichol, “Diffusion models beat gans on image synthesis,” 2021.
- [122] J. Ho, A. Jain, and P. Abbeel, “Denoising diffusion probabilistic models,” 2020.
- [123] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” 2015.

- [124] R. Corvi, D. Cozzolino, G. Zingarini, G. Poggi, K. Nagano, and L. Verdoliva, "On the detection of synthetic images generated by diffusion models," in *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1–5, IEEE, 2023.
- [125] N. C. F. Codella, D. Gutman, M. E. Celebi, B. Helba, M. A. Marchetti, S. W. Dusza, A. Kalloo, K. Liopyris, N. Mishra, H. Kittler, and A. Halpern, "Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (isbi), hosted by the international skin imaging collaboration (isic)," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pp. 168–172, 2018.
- [126] M. Hayati, K. Muchtar, Roslidar, N. Maulina, I. Syamsuddin, G. N. Elwirehardja, and B. Pardamean, "Impact of clahe-based image enhancement for diabetic retinopathy classification through deep learning," *Procedia Computer Science*, vol. 216, pp. 57–66, 2023. 7th International Conference on Computer Science and Computational Intelligence 2022.
- [127] G. Cirrincione, S. Cannata, G. Cicceri, F. Prinzi, T. Currieri, M. Lovino, C. Militello, E. Pasero, and S. Vitabile, "Transformer-based approach to melanoma detection," *Sensors*, vol. 23, no. 12, p. 5677, 2023.
- [128] E. D. Cubuk, B. Zoph, J. Shlens, and Q. V. Le, "RandAugment: Practical automated data augmentation with a reduced search space," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pp. 702–703, 2020.
- [129] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255, 2009.
- [130] Y. Gulzar and S. A. Khan, "Skin lesion segmentation based on vision transformers and convolutional neural networks—a comparative study," *Applied Sciences*, vol. 12, no. 12, p. 5990, 2022.
- [131] X. He, E.-L. Tan, H. Bi, X. Zhang, S. Zhao, and B. Lei, "Fully transformer network for skin lesion analysis," *Medical Image Analysis*, vol. 77, p. 102357, 2022.
- [132] R. Azad, M. Asadi-Aghbolaghi, M. Fathy, and S. Escalera, "Attention deeplabv3+: Multi-level context attention mechanism for skin lesion segmentation," in *Computer Vision—ECCV 2020 Workshops: Glasgow, UK, August 23–28, 2020, Proceedings, Part I 16*, pp. 251–266, Springer, 2020.
- [133] W. Wu, S. Mehta, S. Nofallah, S. Knezevich, C. J. May, O. H. Chang, J. G. Elmore, and L. G. Shapiro, "Scale-aware transformers for diagnosing melanocytic lesions," *IEEE Access*, vol. 9, pp. 163526–163541, 2021.
- [134] S. Aladhadh, M. Alsanea, M. Aloraini, T. Khan, S. Habib, and M. Islam, "An effective skin cancer classification mechanism via medical vision transformer," *Sensors*, vol. 22, no. 11, p. 4008, 2022.
- [135] S. K. Datta, M. A. Shaikh, S. N. Srihari, and M. Gao, "Soft-attention improves skin cancer classification performance," *medRxiv*, 2021.

- [136] B. Zhang, S. Jin, Y. Xia, Y. Huang, and Z. Xiong, "Attention mechanism enhanced kernel prediction networks for denoising of burst images," *arXiv preprint arXiv:1910.08313*, 2019.
- [137] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7132–7141, 2018.
- [138] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, and X. Tang, "Residual attention network for image classification," 2017.
- [139] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015.
- [140] R. S. Lee, F. Gimenez, A. Hoogi, K. K. Miyake, M. Gorovoy, and D. L. Rubin, "A curated mammography data set for use in computer-aided detection and diagnosis research," *Scientific data*, vol. 4, p. 170177, December 2017.
- [141] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255, 2009.
- [142] W. Zhu, Q. Lou, Y. S. Vang, and X. Xie, "Deep multi-instance networks with sparse label assignment for whole mammogram classification," in *International conference on medical image computing and computer-assisted intervention*, pp. 603–611, Springer, 2017.
- [143] X. Shu, L. Zhang, Z. Wang, Q. Lv, and Z. Yi, "Deep neural networks with region-based pooling structures for mammographic image classification," *IEEE transactions on medical imaging*, vol. 39, no. 6, pp. 2246–2255, 2020.
- [144] M. C. Dewan, A. Rattani, S. Gupta, R. E. Baticulon, Y.-C. Hung, M. Punchak, A. Agrawal, A. O. Adeleye, M. G. Shrimel, A. M. Rubiano, *et al.*, "Estimating the global incidence of traumatic brain injury," *Journal of neurosurgery*, vol. 130, no. 4, pp. 1080–1097, 2018.
- [145] A. A. Hyder, C. A. Wunderlich, P. Puvanachandra, G. Gururaj, and O. C. Kobusingye, "The impact of traumatic brain injuries: a global perspective," *NeuroRehabilitation*, vol. 22, no. 5, pp. 341–353, 2007.
- [146] S. O'Brien, K. Metcalf, and J. Batchelor, "An examination of the heterogeneity of cognitive outcome following severe to extremely severe traumatic brain injury," *The Clinical Neuropsychologist*, vol. 34, no. 1, pp. 120–139, 2020.
- [147] A. L. Lee, "Advanced imaging of traumatic brain injury," *Korean Journal of Neurotrauma*, vol. 16, no. 1, p. 3, 2020.
- [148] A. Irimia, S.-Y. Goh, C. Torgerson, P. Vespa, and J. Van Horn, "Structural and connectomic neuroimaging for the personalized study of longitudinal alterations in cortical shape, thickness and connectivity after traumatic brain injury," *Journal of neurosurgical sciences*, vol. 58, no. 3, p. 129, 2014.

- [149] F. Prinzi, T. Currieri, S. Gaglio, and S. Vitabile, “Shallow and deep learning classifiers in medical image analysis,” *European Radiology Experimental*, vol. 8, no. 1, p. 26, 2024.
- [150] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini, “The graph neural network model,” *IEEE transactions on neural networks*, vol. 20, no. 1, pp. 61–80, 2008.
- [151] X. Li, N. C. Dvornek, Y. Zhou, J. Zhuang, P. Ventola, and J. S. Duncan, “Graph neural network for interpreting task-fMRI biomarkers,” in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part V 22*, pp. 485–493, Springer, 2019.
- [152] S. Han, Z. Sun, K. Zhao, F. Duan, C. F. Caiafa, Y. Zhang, and J. Solé-Casals, “Early prediction of dementia using fMRI data with a graph convolutional network approach,” *Journal of Neural Engineering*, vol. 21, no. 1, p. 016013, 2024.
- [153] U. Saha, I. U. Ahamed, I. U. Ahamed, and A.-A. Hossain, “Graph convolutional network-based approach for parkinson’s disease classification using euclidean distance graphs,” *2024 7th International Conference on Informatics and Computational Sciences (ICICoS)*, pp. 532–537, 2024.
- [154] E. Chen, B. Barile, F. Durand-Dubief, T. Grenier, and D. Sappey-Mariniér, “Multiple sclerosis clinical forms classification with graph convolutional networks based on brain morphological connectivity,” *Frontiers in Neuroscience*, vol. 17, p. 1268860, 2024.
- [155] P. E. Pope, S. Kolouri, M. Rostami, C. E. Martin, and H. Hoffmann, “Explainability methods for graph convolutional neural networks,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10772–10781, 2019.
- [156] H. Zhang and K. Ogasawara, “Grad-cam-based explainable artificial intelligence related to medical text processing,” *Bioengineering*, vol. 10, no. 9, p. 1070, 2023.
- [157] F. Ferri, M. Cannariato, L. Pallante, E. A. Zizzi, and M. A. Deriu, “Explainable machine learning and deep learning models for predicting tas2r-bitter molecule interactions,” *arXiv preprint arXiv:2406.15039*, 2024.
- [158] J. Zhou, G. Cui, S. Hu, Z. Zhang, C. Yang, Z. Liu, L. Wang, C. Li, and M. Sun, “Graph neural networks: A review of methods and applications,” *AI open*, vol. 1, pp. 57–81, 2020.
- [159] S. B. Snider, Y. G. Bodien, A. Frau-Pascual, M. Bianciardi, A. S. Foulkes, and B. L. Edlow, “Ascending arousal network connectivity during recovery from traumatic coma,” *NeuroImage: Clinical*, vol. 28, p. 102503, 2020.
- [160] S. B. Snider, Y. G. Bodien, M. Bianciardi, E. N. Brown, O. Wu, and B. L. Edlow, “Disruption of the ascending arousal network in acute traumatic disorders of consciousness,” *Neurology*, vol. 93, no. 13, pp. e1281–e1287, 2019.

- [161] D. Kondziella, A. Bender, K. Diserens, W. van Erp, A. Estraneo, R. Formisano, S. Laureys, L. Naccache, S. Ozturk, B. Rohaut, *et al.*, “European academy of neurology guideline on the diagnosis of coma and other disorders of consciousness,” *European journal of neurology*, vol. 27, no. 5, pp. 741–756, 2020.
- [162] A. Magliacano, P. Liuzzi, R. Formisano, A. Grippo, E. Angelakis, A. Thibaut, O. Gosseries, G. Lamberti, E. No  , S. Bagnato, *et al.*, “Predicting long-term recovery of consciousness in prolonged disorders of consciousness based on coma recovery scale-revised subscores: validation of a machine learning-based prognostic index,” *Brain Sciences*, vol. 13, no. 1, p. 51, 2022.
- [163] E. Landsness, M.-A. Bruno, Q. Noirhomme, B. Riedner, O. Gosseries, C. Schnakers, M. Massimini, S. Laureys, G. Tononi, and M. Boly, “Electrophysiological correlates of behavioural changes in vigilance in vegetative state and minimally conscious state,” *Brain*, vol. 134, no. 8, pp. 2222–2232, 2011.
- [164] M.-A. Bruno, A. Vanhaudenhuyse, A. Thibaut, G. Moonen, and S. Laureys, “From unresponsive wakefulness to minimally conscious plus and functional locked-in syndromes: recent advances in our understanding of disorders of consciousness,” *Journal of neurology*, vol. 258, pp. 1373–1384, 2011.
- [165] S. M. Smith, M. Jenkinson, M. W. Woolrich, C. F. Beckmann, T. E. Behrens, H. Johansen-Berg, P. R. Bannister, M. De Luca, I. Drobnjak, D. E. Flitney, *et al.*, “Advances in functional and structural mr image analysis and implementation as fsl,” *Neuroimage*, vol. 23, pp. S208–S219, 2004.
- [166] J.-D. Tournier, R. Smith, D. Raffelt, R. Tabbara, T. Dhollander, M. Pietsch, D. Christiaens, B. Jeurissen, C.-H. Yeh, and A. Connelly, “Mrtrix3: A fast, flexible and open software framework for medical image processing and visualisation,” *Neuroimage*, vol. 202, p. 116137, 2019.
- [167] E. Garyfallidis, M. Brett, B. Amirbekian, A. Rokem, S. Van Der Walt, M. Descoteaux, I. Nimmo-Smith, and D. Contributors, “Dipy, a library for the analysis of diffusion mri data,” *Frontiers in neuroinformatics*, vol. 8, p. 8, 2014.
- [168] M. Jenkinson, “Bet2: Mr-based estimation of brain, skull and scalp surfaces,” in *Eleventh Annual Meeting of the Organization for Human Brain Mapping, 2005*, 2005.
- [169] J. L. Andersson, S. Skare, and J. Ashburner, “How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging,” *Neuroimage*, vol. 20, no. 2, pp. 870–888, 2003.
- [170] J. L. Andersson and S. N. Sotiropoulos, “An integrated approach to correction for off-resonance effects and subject movement in diffusion mr imaging,” *Neuroimage*, vol. 125, pp. 1063–1078, 2016.
- [171] E. Kellner, B. Dhital, V. G. Kiselev, and M. Reiser, “Gibbs-ringing artifact removal based on local subvoxel-shifts,” *Magnetic resonance in medicine*, vol. 76, no. 5, pp. 1574–1581, 2016.

- [172] J. Veraart, E. Fieremans, and D. S. Novikov, “Diffusion mri noise mapping using random matrix theory,” *Magnetic resonance in medicine*, vol. 76, no. 5, pp. 1582–1593, 2016.
- [173] N. J. Tustison, B. B. Avants, P. A. Cook, Y. Zheng, A. Egan, P. A. Yushkevich, and J. C. Gee, “N4itk: improved n3 bias correction,” *IEEE transactions on medical imaging*, vol. 29, no. 6, pp. 1310–1320, 2010.
- [174] J. Falcó-Roget, A. Cacciola, F. Sambataro, and A. Crimi, “Functional and structural reorganization in brain tumors: a machine learning approach using desynchronized functional oscillations,” *Communications Biology*, vol. 7, no. 1, p. 419, 2024.
- [175] E. T. Rolls, C.-C. Huang, C.-P. Lin, J. Feng, and M. Joliot, “Automated anatomical labelling atlas 3,” *NeuroImage*, vol. 206, p. 116189, 2020.
- [176] T. Dhollander, D. Raffelt, and A. Connelly, “Unsupervised 3-tissue response function estimation from single-shell or multi-shell diffusion mr data without a co-registered t1 image,” in *ISMRM workshop on breaking the barriers of diffusion MRI*, vol. 5, Lisbon, Portugal, 2016.
- [177] T. Dhollander, R. Mito, and A. Connelly, “3-tissue compositional data analysis of developing hcp (dhcp) diffusion mri data,” *Hum. Brain Mapp*, vol. 25, p. T498, 2019.
- [178] R. E. Smith, J.-D. Tournier, F. Calamante, and A. Connelly, “Anatomically-constrained tractography: improved diffusion mri streamlines tractography through effective use of anatomical information,” *Neuroimage*, vol. 62, no. 3, pp. 1924–1938, 2012.
- [179] R. E. Smith, J.-D. Tournier, F. Calamante, and A. Connelly, “Sift2: Enabling dense quantitative assessment of brain white matter connectivity using streamlines tractography,” *Neuroimage*, vol. 119, pp. 338–351, 2015.
- [180] D. S. Bassett and O. Sporns, “Network neuroscience,” *Nature neuroscience*, vol. 20, no. 3, pp. 353–364, 2017.
- [181] O. Sporns, *The human connectome: a structural description of the human brain*, vol. 1. 2011.
- [182] M. Rubinov and O. Sporns, “Complex network measures of brain connectivity: uses and interpretations,” *Neuroimage*, vol. 52, no. 3, pp. 1059–1069, 2010.
- [183] T. N. Kipf and M. Welling, “Semi-supervised classification with graph convolutional networks,” *arXiv preprint arXiv:1609.02907*, 2016.
- [184] M. M. Bronstein, J. Bruna, Y. LeCun, A. Szlam, and P. Vandergheynst, “Geometric deep learning: going beyond euclidean data,” *IEEE Signal Processing Magazine*, vol. 34, no. 4, pp. 18–42, 2017.
- [185] M. Rubinov and O. Sporns, “Complex network measures of brain connectivity: Uses and interpretations,” *NeuroImage*, vol. 52, no. 3, pp. 1059–1069, 2010. Computational Models of the Brain.

- [186] H. Khalili, M. Rismani, M. A. Nematollahi, M. S. Masoudi, A. Asadollahi, R. Taheri, H. Pourmontaseri, A. Valibeygi, M. Roshanzamir, R. Alizadehsani, *et al.*, “Prognosis prediction in traumatic brain injury patients using machine learning algorithms,” *Scientific reports*, vol. 13, no. 1, p. 960, 2023.
- [187] G. Turcato, A. Zaboli, N. Pfeifer, A. Maccagnani, A. Tenci, A. Giudiceandrea, M. Zannoni, G. Ricci, A. Bonora, and F. Brigo, “Decision tree analysis to predict the risk of intracranial haemorrhage after mild traumatic brain injury in patients taking doacs,” *The American Journal of Emergency Medicine*, vol. 50, pp. 388–393, 2021.
- [188] S. Mazurek, R. Blanco, J. Falcó-Roget, and A. Crimi, “Explainable graph neural networks for eeg classification and seizure detection in epileptic patients,” in *2024 IEEE International Symposium on Biomedical Imaging (ISBI)*, IEEE, 2024.
- [189] J. Adebayo, J. Gilmer, M. Muelly, I. Goodfellow, M. Hardt, and B. Kim, “Sanity checks for saliency maps,” in *Advances in Neural Information Processing Systems (NeurIPS)*, pp. 9505–9515, 2018.
- [190] P.-J. Kindermans, S. Hooker, J. Adebayo, M. Alber, K. T. Schütt, S. Dähne, D. Erhan, and B. Kim, “The (un)reliability of saliency methods,” *arXiv preprint arXiv:1711.00867*, 2019.
- [191] M. Newman, *Networks: An Introduction*. Oxford University Press, 03 2010.
- [192] O. Tanglay, I. M. Young, N. B. Dadario, H. Taylor, P. Nicholas, S. Doyen, and M. E. Sughrue, “Eigenvector pagerank difference as a measure to reveal topological characteristics of the brain connectome for neurosurgery,” *Journal of Neuro-Oncology*, vol. 157, pp. 49 – 61, 2022.
- [193] L. Freeman, “A set of measures of centrality based on betweenness,” *Sociometry*, 1977.
- [194] V. Latora and M. Marchiori, “Efficient behavior of small-world networks,” *Physical review letters*, vol. 87, no. 19, p. 198701, 2001.
- [195] D. Chen and H. Su, “Identification of influential nodes in complex networks with degree and average neighbor degree,” *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, vol. 13, pp. 734–742, 2023.
- [196] A. Barrat, M. Barthélemy, R. Pastor-Satorras, and A. Vespignani, “The architecture of complex weighted networks,” *Proceedings of the national academy of sciences*, vol. 101, no. 11, pp. 3747–3752, 2004.
- [197] P. Bonacich, “Power and centrality: A family of measures,” *American Journal of Sociology*, vol. 92, pp. 1170 – 1182, 1987.
- [198] L. Katz, “A new status index derived from sociometric analysis,” *Psychometrika*, vol. 18, no. 1, pp. 39–43, 1953.
- [199] D. S. Bassett and E. T. Bullmore, “Human brain networks in health and disease,” *Current opinion in neurology*, vol. 22, no. 4, pp. 340–347, 2009.

- [200] M. P. Van Den Heuvel and O. Sporns, “Rich-club organization of the human connectome,” *Journal of Neuroscience*, vol. 31, no. 44, pp. 15775–15786, 2011.
- [201] R. P. Vertes and S. B. Linley, “Efferent and afferent connections of the dorsal and median raphe nuclei in the rat,” in *Serotonin and sleep: molecular, functional and clinical aspects*, pp. 69–102, Springer, 2008.
- [202] S. J. Sara and S. Bouret, “Orienting and reorienting: the locus coeruleus mediates cognition through arousal,” *Neuron*, vol. 76, no. 1, pp. 130–141, 2012.
- [203] B. L. Jacobs and E. C. Azmitia, “Structure and function of the brain serotonin system,” *Physiological reviews*, vol. 72, no. 1, pp. 165–229, 1992.
- [204] S. M. Sherman and R. W. Guillery, *Functional connections of cortical areas: a new view from the thalamus*. MIT press, 2013.
- [205] B. L. Edlow, M. Olchanyi, H. J. Freeman, J. Li, C. Maffei, S. B. Snider, L. Zöllei, J. E. Iglesias, J. Augustinack, Y. G. Bodien, *et al.*, “Sustaining wakefulness: brain-stem connectivity in human consciousness,” *bioRxiv*, 2023.
- [206] J. T. Giacino, S. Ashwal, N. Childs, R. Cranford, B. Jennett, D. I. Katz, J. P. Kelly, J. H. Rosenberg, J. Whyte, R. D. Zafonte, *et al.*, “The minimally conscious state: definition and diagnostic criteria,” *Neurology*, vol. 58, no. 3, pp. 349–353, 2002.