UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Escola d'Enginyeria de Barcelona Est

FINAL DEGREE PROJECT

Bachelor's Degree in Biomedical Engineering

Medical Software Module in 3D Slicer for Automatic Segmentation and Trajectory Reconstruction of SEEG Electrodes Using AI and Data Science

Memory and Annexes

Author:	Rocío Ávalos Morillas
Director:	Christian Mata
Co-Director:	Justo Montoya-Gálvez
Co-Director:	Alessandro Principe
Call:	June 2025

Abstract

Epilepsy is a neurological disorder that affects a significant portion of the global population. One of the most widely used methods for its study in patients eligible for surgery is stereoelectroencephalography (SEEG), which involves the implantation of intracranial electrodes to record brain activity in deep regions. These electrodes have multiple contact points that must be precisely located, as their accurate identification is crucial for surgical planning and clinical analysis.

Currently, the localization of these contact points is performed manually by a team of specialists, such as neurosurgeons and neurologists. This process is highly time-consuming, requiring over four hours of collaborative work for a single patient, placing a substantial operational burden on specialized clinical units.

The main objective of this project is to automate the localization of intracranial electrode contacts through the development of an extension module for the 3D Slicer platform. Using an intuitive interface, the module enables end-to-end processing from CT and MRI images to validated electrode coordinates through a novel multi-mask ensemble architecture that generates 38 segmentation variants with intelligent consensus formation.

The proposed approach integrates medical image processing, machine learning, and deep learning within a human-AI collaboration framework that preserves clinical decision-making authority. A conservative confidence scoring system enables graduated clinical decision-making, while the ensemble methodology achieves 98.8% accuracy within the 2mm clinical threshold across the patient cohort, with perfect performance on held-out patient validation.

This system has been successfully deployed at the Epilepsy Unit of Hospital del Mar, reducing electrode localization from over four hours to thirty minutes while maintaining clinical accuracy standards. The dramatic improvement in processing efficiency demonstrates significant potential to enhance patient safety, improve clinical workflows, and support evidence-based surgical planning in epilepsy treatment.

Resumen

La epilepsia es un trastorno neurológico que afecta a una parte considerable de la población mundial. Uno de los métodos más utilizados para su estudio en pacientes candidatos a cirugía es la electroencefalografía estereotáctica (SEEG), la cual consiste en la implantación de electrodos intracraneales para registrar la actividad eléctrica cerebral en regiones profundas. Estos electrodos poseen múltiples puntos de contacto que deben ser localizados con precisión, ya que su correcta identificación es fundamental para la planificación quirúrgica y el análisis clínico.

Actualmente, la localización de estos puntos de contacto se realiza manualmente por parte de un equipo de especialistas, como neurocirujanos y neurólogos. Este proceso es altamente laborioso, requiriendo más de cuatro horas de trabajo colaborativo para un solo paciente, lo que representa una carga operativa considerable para las unidades clínicas especializadas.

El objetivo principal de este proyecto es automatizar la localización de contactos de electrodos intracraneales mediante el desarrollo de un módulo de extensión para la plataforma 3D Slicer. El módulo permite el procesamiento integral desde imágenes de TC y RM hasta coordenadas validadas mediante una arquitectura innovadora de ensemble multi-máscara que genera 38 variantes de segmentación.

El enfoque integra procesamiento de imágenes médicas, aprendizaje automático y aprendizaje profundo dentro de un marco de colaboración humano-IA que preserva la autoridad clínica. La metodología de ensemble alcanza una precisión del 98,8% dentro del umbral clínico de 2mm, con rendimiento perfecto en la validación de pacientes externos.

Este sistema se ha desplegado exitosamente en la Unidad de Epilepsia del Hospital del Mar, reduciendo la localización de electrodos de cuatro horas a treinta minutos. La mejora en eficiencia demuestra potencial significativo para mejorar la seguridad del paciente y optimizar los flujos de trabajo clínicos en el tratamiento de la epilepsia.

Resum

L'epilèpsia és un trastorn neurològic que afecta una part considerable de la població mundial. Un dels mètodes més utilitzats per al seu estudi en pacients candidats a cirurgia és l'electroencefalografia estereotàctica (SEEG), que consisteix en la implantació d'elèctrodes intracranials per registrar l'activitat elèctrica cerebral en regions profundes. Aquests elèctrodes tenen múltiples punts de contacte que han de ser localitzats amb precisió, ja que la seva identificació correcta és fonamental per a la planificació quirúrgica i l'anàlisi clínica.

Actualment, la localització d'aquests punts de contacte es realitza manualment per part d'un equip d'especialistes, com neurocirurgians i neuròlegs. Aquest procés és molt laboriós, requerint més de quatre hores de treball col · laboratiu per a un sol pacient, la qual cosa representa una càrrega operativa considerable per a les unitats clíniques especialitzades.

L'objectiu principal d'aquest projecte és automatitzar la localització dels contactes d'elèctrodes intracranials mitjançant el desenvolupament d'un mòdul d'extensió per a la plataforma 3D Slicer. El mòdul permet el processament integral des d'imatges de TC i RM fins a coordenades validades mitjançant una arquitectura innovadora d'ensemble multi-màscara que genera 38 variants de segmentació.

L'enfocament integra processament d'imatges mèdiques, aprenentatge automàtic i aprenentatge profund dins d'un marc de col · laboració humà-IA que preserva l'autoritat clínica. La metodologia d'ensemble assoleix una precisió del 98,8% dins del llindar clínic de 2mm, amb un rendiment perfecte en la validació de pacients externs.

Aquest sistema s'ha desplegat amb èxit a la Unitat d'Epilèpsia de l'Hospital del Mar, reduint la localització d'elèctrodes de quatre hores a trenta minuts. La millora en eficiència demostra potencial significatiu per millorar la seguretat del pacient i optimitzar els fluxos de treball clínics en el tractament de l'epilèpsia.

Acknowledgements

I would like to express my deepest gratitude to all those who made this work possible and supported me throughout this journey.

First and foremost, I extend my sincere appreciation to my co-directors, Dr. Alessandro Principe and Justo Montoya-Gálvez, for providing me with the opportunity to work on such an exciting and clinically impactful project. Their consistent support, thoughtful guidance, and willingness to address my questions and discuss new proposals have been invaluable throughout this research. I am particularly grateful for their approach in fostering my autonomy while encouraging me to think outside the box—a balance that allowed me to grow as both a researcher and an innovator.

I would also like to thank my director, Professor Christian Mata, for his essential guidance in organizing and articulating the complex methodologies and findings of this work. His expertise in helping me bring together all the technical components into a coherent thesis has been instrumental in its successful completion.

My gratitude extends to the Epilepsy Unit at Hospital del Mar and the Center for Brain and Cognition (CBC) at Universitat Pompeu Fabra (UPF) for providing the collaborative research environment and clinical context that made this work possible. The interdisciplinary nature of this project has been both challenging and rewarding.

Finally, I am profoundly grateful to my family, whose unconditional love, encouragement, and support throughout my life have been the foundation of all my achievements. Their belief in me has been a constant source of strength and motivation.

This work represents not only a technical achievement but also the result of the support, guidance, and encouragement from all these remarkable individuals. Thank you.

Glossary

3D Slicer Open-source platform for medical image visualization and analysis used for viewing CT/MRI data and running segmentation workflows.

CT (Computed Tomography)
X-ray imaging technique providing high-resolution cross-sectional images, essential for visualizing metallic SEEG electrodes post-implantation.

DBSCAN Density-based clustering algorithm used for grouping electrode contacts into coherent trajectories based on spatial proximity.

Dice Coefficient
Similarity metric measuring spatial overlap between predicted and ground truth segmentations, ranging from 0 (no overlap) to 1 (perfect match).

Drug-resistant Epilepsy
Epilepsy condition where seizures persist despite adequate trials of two appropriate antiepileptic medications.

Hounsfield Units (HU)
Standardized scale for CT image intensities where water = 0 HU, air = -1000 HU, and platinum-iridium electrodes = 1500-3000 HU.

LightGBM Gradient boosting machine learning framework used for electrode contact authentication and confidence score prediction.

LOPO (Leave-One-Patient-Out)
Cross-validation strategy where each patient serves as test data while remaining patients train the model.

LPS/RAS Coordinates
Medical coordinate systems: LPS (Left-Posterior-Superior) used by ITK/SimpleITK; RAS (Right-Anterior-Superior) used by 3D Slicer.

MRI (Magnetic Resonance Imaging)
: Non-invasive imaging technique providing detailed soft tissue contrast for anatomical reference and brain extraction.

MONAI
: Medical imaging AI framework used for implementing the 3D U-Net brain segmentation model.

NRRD
: Nearly Raw Raster Data format for storing medical images with spatial metadata, enabling cross-platform compatibility.

PCA (Principal Component Analysis)
: Statistical technique for identifying primary directional patterns in electrode arrangements and trajectory reconstruction.

Random Forest
: Ensemble machine learning algorithm combining multiple decision trees, used for adaptive threshold prediction.

SEEG (Stereoelectroencephalography)
: Minimally invasive technique involving implantation of depth electrodes to localize seizure onset zones in drug-resistant epilepsy.

U-Net
: Convolutional neural network architecture designed for medical image segmentation, implemented here for automated brain extraction from CT scans.

Contents

List of Figures

List of Tables

1 Introduction

1.1 Origin of the project

This project emerged from my internship at Hospital del Mar, where I was introduced to a collaborative initiative between the Hospital del Mar Research Institute and the Center for Brain and Cognition (CBC) at Universitat Pompeu Fabra (UPF).

During my internship, I had the opportunity to explore the development of this project, which was spearheaded by Alessandro Principe, a neurophysiologist specialized in epilepsy, and Justo Montoya-Gálvez, a PhD student in Computational Neuroscience.

Figure 1.1: *Logo of the Universitat Politècnica de Catalunya (UPC), the Center for Brain and Cognition (CBC), the Universitat Pompeu Fabra (UPF), and the Hospital del Mar Research Institute. Source: Own source.*

Their innovative work, which focuses on advancing the understanding of epilepsy through computational methods, greatly influenced my decision to contribute to this research. The main objective of this project is to create a Slicer 3D module for the automatic segmentation of stereotactic intracranial electroencephalography (SEEG) electrodes, which aims to improve the efficiency of electrode localization in patients with epilepsy.

Since the module is intended for clinical use at the Epilepsy Unit of hospitals, it must be user-friendly and intuitive, especially for medical professionals without a background in computer vision or informatics.

1.2 Motivation

The motivation behind this project arises from both academic interests and professional aspirations.

My interest in image processing began during my undergraduate studies, particularly through the course titled Image Processing. This subject introduced me to the wide range of techniques and applications where biomedical engineers can contribute meaningfully to the healthcare field. It fostered a strong curiosity about the role of computational tools in medical imaging and clinical decision-making.

During my internship at Hospital del Mar, I came across a research opportunity at the Center for Brain and Cognition (CBC), focused on the development of a software tool to support epilepsy diagnosis and treatment. The project involved the automatic segmentation of SEEG electrodes from postoperative CT images, using Python-based image analysis and machine learning techniques. This immediately captured my interest, as it aligned closely with my academic background and personal goals.

The proposal emphasized an interdisciplinary approach, combining programming, neuroanatomy, and medical image processing. Following the initial meeting with the research team, I realized the significant potential of the project—not only to improve the precision and efficiency of SEEG electrode localization, but also to allow me to apply advanced image processing methods, machine learning algorithms, and software development skills within the 3D Slicer platform.

This project represents a valuable opportunity to contribute to real-world clinical research while deepening my expertise in biomedical image analysis.

1.3 Background and Rationale

This project addresses a critical need identified by neurosurgeons and neurophysiologists worldwide: the development of a user-friendly module to help automate the localization of contact points on stereo electroencephalography (SEEG) electrodes.

The prevailing clinical approach involves manual identification of electrodes from post-surgical CT scans by medical teams. This methodology places significant de-

mands on specialist time, as it requires coordinated efforts between neurosurgeons and neurophysiologists, often extending to several hours of collaborative analysis per individual patient.

This manual approach creates a cascade of serious clinical and economic consequences that extend far beyond simple workflow inefficiency. Patients with implanted electrodes must remain hospitalized during the lengthy manual analysis process, significantly increasing their exposure to hospital-acquired infections and other complications. The prolonged implantation time of the electrode increases the risk of intracranial bleeding, infection, and other serious complications associated with invasive brain monitoring.

Furthermore, the time-intensive manual process delays critical surgical planning, potentially compromising optimal treatment outcomes for patients already in a vulnerable neurological state.

From a healthcare system perspective, the requirement for several hours represents a substantial misallocation of resources. Specialized neurosurgeons and neurologists whose expertise is critical for patient care are instead tied to time-consuming manual annotation tasks. This creates a significant economic burden through prolonged hospital stays and intensive specialist time that could be redirected to treating additional patients.

Manual processing also creates bottlenecks in the clinical workflow that limit the number of patients who can receive timely SEEG evaluation, further exacerbating waiting times in already strained epilepsy centres. The quality and consistency of manual electrode localization also presents significant challenges.

Inter-observer variability between different specialists can affect treatment planning and surgical decisions, while operator fatigue during extended manual analysis increases the likelihood of human errors, particularly during the later stages of the lengthy process. Manual approaches also vary considerably between institutions and practitioners, limiting reproducibility and comparative analysis across different medical centers.

Although complete automation is neither realistic nor advisable - given that clinical validation by medical experts remains essential and no algorithm achieves reliability 100 %, there is an urgent need for robust automated assistance tools. These

tools should significantly reduce the manual workload while providing reliable initial electrode localization that medical professionals can validate and refine. Even with necessary human oversight, a well-designed automated system could reduce the current process, improving patient safety, clinical efficiency, and healthcare resource utilization.

The present work, conducted in collaboration with the Epilepsy Unit at Hospital del Mar, seeks to address this critical clinical need through the development of an integrated automated module for 3D Slicer that combines advanced image processing techniques, machine learning algorithms, and clinical workflow optimization. By leveraging a dataset of pre- and post-operative imaging data from multiple patients, this research aims to create a practical solution that approaches the ideal automated assistance system while maintaining the clinical reliability and expert oversight essential for patient safety.

1.4 Objectives

The principal objective of this thesis is to develop a clinical decision support tool through a custom module within the 3D Slicer platform [1] for automatic segmentation of stereo electroencephalography (SEEG) electrodes. This work directly addresses a critical clinical bottleneck: the current manual electrode localization process requires over 4 hours of collaborative work between neurosurgeons and neurologists per patient, creating substantial operational burden on specialized epilepsy units and delaying critical surgical decisions for patients with drug-resistant epilepsy. To achieve this aim, the project entails a series of interconnected technical and scientific goals.

First, it is necessary to gain a thorough understanding of the development environment. This includes exploring the internal architecture of *3D Slicer*, its capabilities for Python-based scripting, and its integration with graphical user interface design tools such as *Qt Designer*. In parallel, it is essential to become familiar with the NRRD (Nearly Raw Raster Data) file format, particularly regarding metadata conventions and medical coordinate systems, as it is the standard format used within the platform for volumetric data.

Additionally, a review of the state of the art in SEEG electrode segmentation will

be conducted. This involves exploring a range of methodologies, including:

- Classical image processing techniques for label detection and morphological analysis,
- Machine learning approaches tailored for medical image analysis,
- Deep learning architectures, particularly convolutional neural networks (CNNs),
- Geometric and statistical frameworks designed to reconstruct electrode positions and trajectories.

A major component of the project is the design and implementation of algorithms for the detection of electrode candidates in post-operative CT scans. These algorithms will perform the following:

1. Isolate regions of interest that may correspond to electrodes,
2. Extract spatial and intensity features around candidate centroids,
3. Train classification models to distinguish true electrodes from false positives, using both supervised machine learning and deep learning strategies.

In addition to electrode identification, the thesis focuses on the reconstruction of electrode trajectories. This task requires the development of path estimation algorithms that account for anatomical structures, inter-electrode distances, and angular constraints. The methodology must also accommodate variations in electrode arrangements and address challenges such as overlapping or intersecting paths.

The proposed methods will be evaluated on a multi-patient dataset to ensure their robustness and generalization. Validation will include both qualitative visual assessment and quantitative performance metrics, comparing alternative approaches to determine the most accurate and computationally efficient solutions.

Finally, the project includes the design of a graphical user interface (GUI) that integrates all components into a coherent and accessible module within *3D Slicer*. The interface will prioritize:

- User-friendly visualization of segmentation and trajectory results,
- Flexibility for clinician and researcher use without requiring deep technical knowledge,

Through this interdisciplinary approach—spanning software development, image processing, machine learning, and neuroscience—this thesis aims to contribute a functional and reliable tool to the domain of SEEG analysis.

1.5 Project Schedule

This chapter presents the project timeline and task distribution throughout the six-month development period from January to June 2025. The project schedule demonstrates the systematic approach taken to develop the automated SEEG electrode localization system, from initial research through clinical deployment.

The development process was structured into six main phases, each with specific objectives and deliverables. Figure 1.2 illustrates the detailed project timeline, showing the distribution of tasks across the development period.

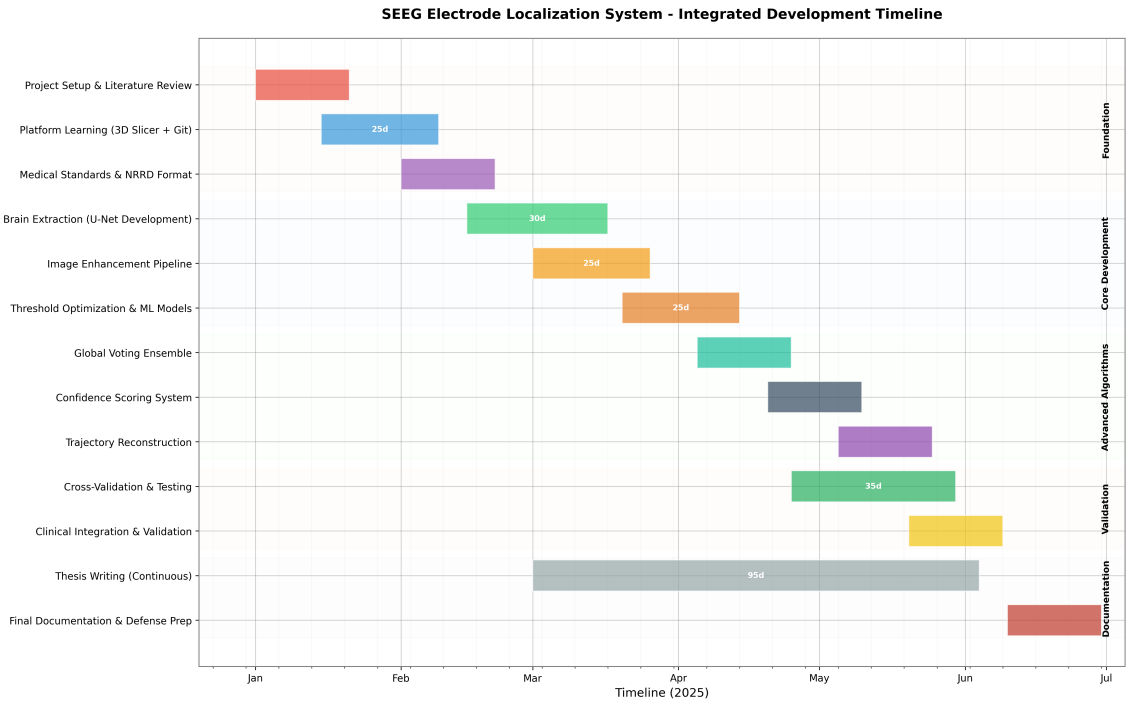


Figure 1.2: *Project timeline showing the six development phases and sixteen major tasks completed during the bachelor thesis development period. Source: Own source.*

The total project duration encompassed approximately 1,000 working hours distributed across literature review, platform learning, algorithm development, validation, and clinical integration. This timeline demonstrates the comprehensive scope of work required to transform a research concept into a clinically deployable solution within the constraints of a bachelor thesis project.

2 Theoretical Framework

Epilepsy is one of the most common neurological disorders worldwide, affecting millions of people [2] and often requiring complex diagnostic and surgical procedures.

While most epilepsy patients respond well to antiepileptic medications, approximately one-third of patients develop drug-resistant epilepsy [3], defined as the failure of adequate trials of two tolerated and appropriately chosen antiepileptic drug schedules [4] to achieve sustained seizure freedom. For these patients with drug-resistant epilepsy, surgical treatment may be considered as the next therapeutic option, but only after precise localization of the seizure onset zone. This critical localization is commonly achieved through stereotactic intracranial electroencephalography (SEEG), an invasive method that involves implanting depth electrodes into the brain to record electrical activity.

Drug-resistant epilepsy represents a particularly challenging clinical scenario where patients continue to experience seizures despite optimal medical management, significantly impacting their quality of life, cognitive function, and psychosocial well-being [4]. For this substantial patient population, representing millions of individuals worldwide, surgical intervention offers the potential for seizure freedom and dramatic improvement in life outcomes. However, the success of epilepsy surgery is critically dependent on the accurate identification and localization of the epileptogenic zone [5], making precise SEEG electrode positioning essential for optimal surgical outcomes.

In this chapter, an overview of the clinical context of epilepsy, the SEEG technique, and the role of medical imaging will be provided. Additionally, a brief introduction to image processing, machine learning, and deep learning methods will be given, as these form the basis of the methodology used in the development of the proposed module.

2.1 Anatomy of the Brain

The human brain is a complex organ composed of multiple interconnected regions that work together to control bodily functions, behavior, and cognition. Anatom-

ically, it can be divided into three main parts: the cerebrum, the cerebellum, and the brainstem.

The cerebrum is the largest and most prominent structure, responsible for higher-order processes such as thought, and voluntary movement. The cerebellum, located posteriorly beneath the cerebrum, plays a critical role in motor coordination and balance. The brainstem, which connects the brain to the spinal cord, regulates vital involuntary functions such as respiration and heartbeat [6]. The thalamus, located within the cerebrum, acts as a relay station for sensory and motor signals, while the hypothalamus regulates body temperature, hunger, and thirst.

The surface of the cerebrum is known as the cerebral cortex, characterized by gyri (ridges) and sulci (grooves), which increase the surface area available for neural processing. The cortex is composed primarily of gray matter, containing neuronal cell bodies, while the inner layers consist of white matter, composed of myelinated axons that enable communication between different brain regions [7].

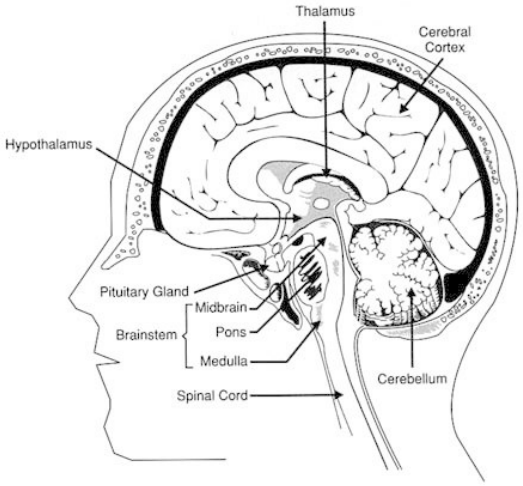


Figure 2.1: *Graphical representation of a sagittal view of the human brain, illustrating key anatomical structures including the cerebral cortex, thalamus, hypothalamus, cerebellum, brainstem, pituitary gland, and spinal cord [8]*

Each cerebral hemisphere is subdivided into four lobes—frontal, temporal, parietal, and occipital—each associated with distinct functional domains. Among these, the temporal and frontal lobes are the most frequent origins of focal epileptic seizures, especially in patients considered for surgical treatment. The temporal lobe, home

to critical structures such as the hippocampus and amygdala, is especially prone to epileptogenic activity [9]. The frontal lobe, is the second most frequent focus of seizure onset.

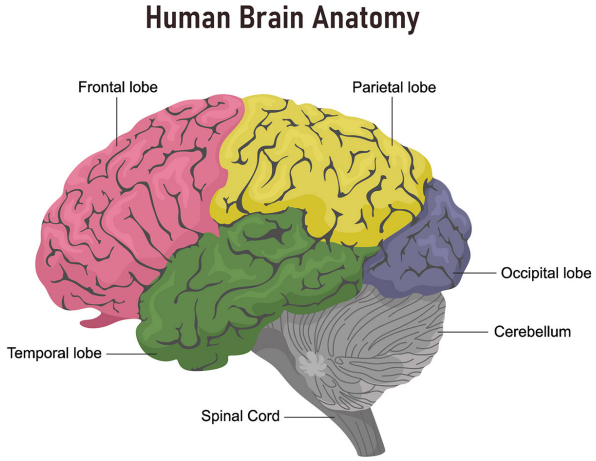


Figure 2.2: *Graphical representation of the lobes of the brain in different colors. Adapted from Sandra Ackerman's Discovering the Brain. [10]*

Given the brain's complex architecture, understanding its anatomical divisions is crucial for localizing seizure activity in epilepsy patients

2.2 Epilepsy

Epilepsy is a chronic neurological disorder characterized by recurrent, unprovoked seizures caused by abnormal electrical discharges in the brain [2]. These seizures may affect a localized region (focal seizures) or involve the entire brain (generalized seizures), often leading to involuntary movements, altered awareness, or even loss of consciousness [11]. The condition can affect individuals of any age, from infancy to old age, and its clinical manifestations vary widely depending on the seizure type and the brain regions involved [2].

Globally, epilepsy affects an estimated 50 million people, making it one of the most prevalent neurological disorders worldwide [2]. Although it can occur at any age, it is most frequently diagnosed in children and older adults, often due to genetic predisposition and age-related neurological changes[12].

The causes of epilepsy are diverse and include genetic mutations, brain injuries (e.g., trauma, stroke), infections (such as neurocysticercosis), tumors, and developmental disorders. However, in approximately 50% of cases, no definitive cause can be identified [13].

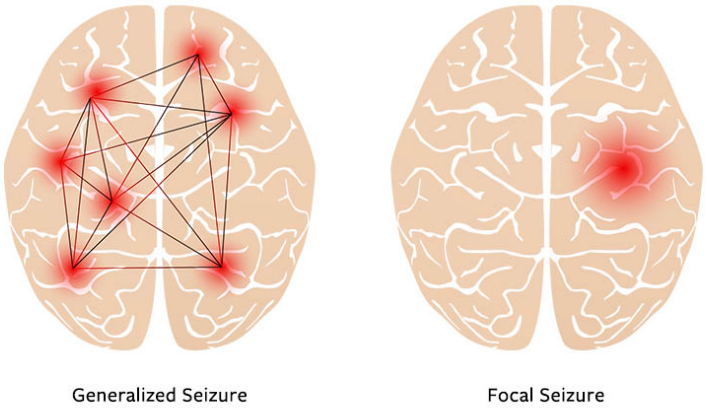


Figure 2.3: *Generalized seizures and focal seizures [14]*

Accurate classification of epilepsy and identification of its underlying cause are essential for effective diagnosis, treatment planning, and prognosis.

2.3 Stereoelectroencephalography (SEEG)

2.3.1 Overview

SEEG is a minimally invasive, highly specialized neurophysiological technique used to diagnose and manage drug-resistant focal epilepsy. It involves the stereotactic implantation of depth electrodes to precisely localize seizure onset zones (SOZ), offering high spatial and temporal resolution across cortical and subcortical structures [15]. Unlike surface recordings such as electrocorticography (ECoG), SEEG provides direct access to deep brain regions, including medial temporal, insular, cingulate, and mesial frontal structures.

SEEG is typically reserved for complex cases where noninvasive methods—such as MRI, PET, or scalp EEG—fail to identify a clear epileptogenic zone. Its development dates back to the 1950s, when Talairach and Bancaud pioneered its use in

Paris, proposing the *anatomo-electro-clinical correlation (AEC)* approach to strategically explore seizure networks [16].

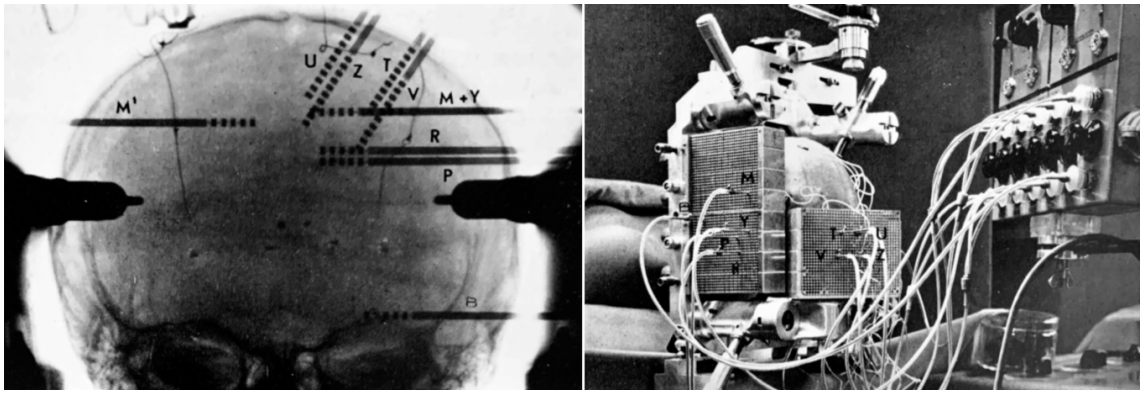


Figure 2.4: *The Talairach stereotaxic frame, developed around 1962, featured a dual-grid system enabling both orthogonal and oblique trajectories for electrode implantation [17].*

2.3.2 Applications of SEEG in Epilepsy

SEEG is increasingly used in the presurgical evaluation of patients with complex focal epilepsies.

The ability to explore deep and bilateral regions, combined with video-EEG monitoring, allows clinicians to construct a precise map of the epileptogenic zone, which is critical for resective or ablative treatment planning.

Electrodes are inserted through small burr holes and externally connected to a recording system for long-term monitoring.

2.3.3 Electrode Localization and Imaging

Accurate localization is essential for interpreting SEEG recordings. Specialized software is used to co-register pre-implantation MRI with post-implantation CT or MRI scans to determine the precise location of each electrode contact. This process helps account for any brain shift that may occur during the implantation procedure.

Pre- and post-implantation imaging involves:

- **MRI**: Isometric T1-weighted images with $\sim 1 \times 1 \times 1$ mm voxel size and no inter-slice gaps
- **3D Reconstructions**: Coronal, sagittal, and axial views used for contact verification

- **Symmetrical Implantation**: Bilateral depth electrodes are sometimes used for comparative analysis [15]

Current Clinical Practice and Limitations: Currently, SEEG contact localization relies heavily on manual annotation by specialized clinical teams, including neurosurgeons and neurophysiologists. This manual process is extremely time-consuming, and represents a significant operational burden on epilepsy units worldwide. The manual approach introduces potential variability between annotators and limits the scalability of SEEG analysis workflows.

Due to metal artifacts and proximity of trajectories, electrode contacts can sometimes appear as merged clusters in imaging. This "contact merging" phenomenon, combined with the need for sub-millimeter precision, makes manual localization particularly challenging and highlights the critical need for automated solutions that can maintain clinical accuracy while dramatically reducing processing time.

SEEG remains a cornerstone in the evaluation of complex epilepsy cases and continues to evolve through integration with advanced imaging, robotics, and computational analysis. The accurate localization of SEEG electrodes is a crucial step, but it's often challenging due to artifacts, the number of electrodes and their proximity. My research aims to address this challenge by developing a medical software module for automatic segmentation of SEEG electrodes, improving the accuracy and efficiency of this critical process

2.4 Medical Imaging in SEEG

SEEG fundamentally relies on the integration of multiple medical imaging modalities to enable precise localization of its electrodes. This section presents the physical principles, clinical relevance, and inherent limitations of the primary imaging techniques—Magnetic Resonance Imaging (MRI) and Computed Tomography (CT)—as well as the role of the NRRD format in supporting standardized, multimodal data representation.

2.4.1 Magnetic Resonance Imaging (MRI)

Physical Principles

Magnetic Resonance Imaging (MRI) exploits the magnetic properties of hydrogen nuclei, which are abundant in biological tissues due to water content. When placed in a strong static magnetic field (B_0, typically 1.5–7 Tesla in clinical settings), these nuclei align with the field. A radiofrequency pulse at the Larmor frequency ($\omega_0 = \gamma B_0$) disturbs this alignment, and as the nuclei return to equilibrium, they emit detectable signals. The relaxation times—T1 (spin-lattice) and T2 (spin-spin)—govern the contrast between different tissue types, enabling detailed visualization of brain anatomy. T1-weighted images excel at distinguishing gray and white matter, while T2-weighted images are sensitive to pathological changes such as lesions. These properties make MRI uniquely suited for high-contrast, high-resolution anatomical imaging of the brain [18, 19, 20].

Relevance in SEEG

MRI is fundamental in SEEG planning, providing the anatomical roadmap for safe and effective electrode implantation. High-resolution MRI allows precise identification of cortical and subcortical structures, including the gray–white matter boundary, sulci, gyri, and deep brain regions such as the hippocampus and amygdala. This anatomical detail is crucial for targeting the epileptogenic zone and for avoiding eloquent cortex during electrode placement [19, 20, 21, 22].

Common MRI sequences used in SEEG include T1-weighted (e.g., MP-RAGE) for cortical mapping, T2-weighted (e.g., SPACE) for lesion detection, and susceptibility-weighted imaging (SWI) for visualizing vasculature to avoid during implantation. These sequences together enable comprehensive assessment of both normal anatomy and pathology relevant for SEEG trajectory planning [19, 20].

Role and Importance

In SEEG, MRI serves as the anatomical reference for electrode localization and trajectory optimization. It supports the integration of multimodal data—including seizure semiology, scalp EEG, PET, and SPECT—into a hypothesis-driven implantation plan. Accurate anatomical mapping with MRI is essential for maximizing

the diagnostic yield of SEEG and minimizing surgical risks. MRI-based localization also facilitates post-implantation analysis by allowing precise mapping of electrode contacts relative to brain regions [19, 21, 22, 23].

Limitations

Despite its advantages, MRI is susceptible to geometric distortions, particularly at higher field strengths, and is sensitive to artifacts from metallic objects, such as previously implanted electrodes. Long acquisition times can lead to motion artifacts, affecting image quality. Advanced imaging techniques, including distortion correction algorithms and multi-echo sequences, help mitigate these limitations and improve the accuracy of electrode localization [19, 21].

2.4.2 Computed Tomography (CT) in SEEG Electrode Localization

Physical Principles

CT imaging utilizes X-ray attenuation as described by the Beer-Lambert law:

$$I = I_0 \cdot e^{-\mu_{\text{tot}} x} \tag{2.1}$$

where μ_{tot} represents the linear attenuation coefficients of materials along the X-ray path [24]. Tissue contrast is quantified in Hounsfield Units (HU), defined as:

$$\text{HU} = 1000 \cdot \frac{\mu_{\text{pixel}} - \mu_{\text{water}}}{\mu_{\text{water}}} \tag{2.2}$$

This scale enables clear visualization of platinum-iridium electrodes (1,500-3,000 HU) against biological tissues [24].

Clinical Applications

Postoperative CT serves as the gold standard for electrode localization due to its superior metallic artifact tolerance and submillimeter spatial resolution (typically 512×512 matrices) [24, 25]. Modern protocols combine 120 kVp tube voltage with 0.5-0.625 mm slice thickness to balance radiation dose and spatial accuracy [26]. The rapid acquisition time (<2 minutes) minimizes motion artifacts, particularly

beneficial for postoperative patients [27].

CT imaging acquires a series of thin axial slices, which are then computationally reconstructed into a three-dimensional (3D) volume. This 3D reconstruction is essential for precise localization of SEEG electrodes relative to brain anatomy. However, the use of thinner slices, while improving spatial resolution, can increase image noise due to reduced photon statistics per slice. When these slices are combined for 3D reconstruction, the resulting dataset may exhibit increased graininess, potentially affecting the visibility of small structures. Advanced denoising and iterative reconstruction algorithms are therefore often employed to mitigate noise while preserving anatomical detail [26, 28].

2.4.3 NRRD: Standardized Format for Multimodal Integration

Format Overview

The NRRD format is designed to store and exchange multidimensional medical image data efficiently. It separates metadata from the raw binary image data, with the metadata stored in a simple ASCII header that includes crucial spatial information such as voxel size, orientation (`space directions`), and image origin (`space origin`) [29]. Unlike DICOM, which often consists of many separate files per scan, NRRD typically encapsulates an entire volumetric dataset in a single file, simplifying data management and transfer. Additionally, NRRD files are anonymized by design, excluding sensitive patient information, which facilitates privacy-preserving data sharing [30].

Role in SEEG Workflows

In SEEG workflows, NRRD ensures consistent spatial metadata across different imaging modalities such as MRI and CT. This consistency is essential for accurate multimodal image registration, enabling precise alignment of anatomical images with electrode localization maps. The format's support for storing both raw imaging data and label maps in a unified framework facilitates seamless integration, visualization, and analysis of complex datasets necessary for planning and evaluating SEEG procedures.

2.4.4 Summary

In SEEG, MRI and CT are used in complementary roles: MRI provides high-resolution anatomical reference for surgical planning, while CT confirms electrode placement with high spatial accuracy. Standardized data formats such as NRRD ensure compatibility across imaging and analysis pipelines.

Emerging technologies—including ultra-high-field MRI, photon-counting CT, and AI-enhanced fusion techniques—are expected to further optimize SEEG workflows by reducing artifacts, improving contrast, and automating segmentation and registration.

3 State of the art

Accurate localization of SEEG electrodes in post-implantation CT images is a foundational step in epilepsy surgery planning and analysis. The task is technically challenging due to the small size of electrode contacts, the presence of metal artifacts, anatomical variability, and the frequent bending of electrodes during implantation, also because it requires human intervention. Each electrode has to be visually found and then manually placed in the virtual 3D model. This section provides a comprehensive overview of the theoretical foundations, algorithmic strategies, deep learning approaches, and practical limitations of current methods for SEEG segmentation and localization.

3.1 Theoretical Foundations and Classical Approaches

3.1.1 Image Formation and Challenges

SEEG electrodes are typically visualized in post-implantation CT scans as hyperdense objects, with platinum-iridium contacts and bolts producing high-intensity signals. However, the imaging process introduces several challenges:

- **Metal Artifacts**: Beam hardening and photon starvation can create streaks and shadows, obscuring fine electrode details or adjacent anatomy.
- **Noise**: Thin-slice CT protocols (0.5–0.625 mm) are used for high spatial resolution, but can increase image noise, complicating segmentation.
- **Anatomical Variability**: The trajectory and bending of the electrodes vary between patients and even between electrodes in the same patient.

3.1.2 Thresholding and Connected Component Analysis

The earliest and most widely used approaches for electrode segmentation are based on classical image processing techniques:

- **Thresholding**: A global or adaptive threshold is applied to the CT volume to create a binary mask highlighting hyperdense regions corresponding to metallic components [31].
- **Connected Component Analysis**: The binary mask is analyzed to identify

contiguous regions (components), which are then classified as bolts, shafts, or contacts based on size, shape, and spatial arrangement.

For example, in the SEEG Assistant (SEEGA) extension for 3D Slicer [32], a thresholded CT scan is processed to extract connected components, which are then filtered by size to identify bolts (large components) and contacts (small components). The algorithm requires a user-provided fiducial list specifying the planned entry and target points for each electrode, which is used to guide the assignment of detected contacts to specific electrodes [33], which is not optimal because the plan often changes.

3.1.3 Trajectory Modeling and Bending Compensation

To address the problem of electrode bending, more sophisticated algorithms model the electrode shaft as an elastic rod or spline:

- **Cosserat Rod Theory**: This continuum mechanics approach models the electrode as a flexible rod, providing a physically plausible estimate of its trajectory even in the presence of significant bending [31].
- **Line-of-Best-Fit**: A least-squares line or spline is fitted to the proximal contacts to estimate the trajectory, and remaining contacts are projected onto this line to refine their positions [32].

For instance, Meesters et al. introduced a method that uses position-based dynamics to model both local and global bending, improving the robustness of contact localization, especially when electrodes deviate from the planned trajectory [31].

3.2 Data-Driven and Deep Learning Approaches

3.2.1 U-Net and Hybrid Architectures

Recent years have seen a shift towards data-driven methods, particularly convolutional neural networks (CNNs) such as U-Net.

These networks are trained on large datasets of annotated CT images, sometimes augmented with synthetic data that simulates metal artifacts, anatomical variability, and noise. Dice coefficients for contact segmentation typically range from 0.89 to 0.92, and inference times are suitable for clinical use (seconds to a few minutes per scan) [24, 34].

3.2.2 Gaussian Mixture Models and Clustering

Some pipelines supplement deep learning with statistical clustering methods:

- **Gaussian Mixture Models (GMM)**: Used to group segmented contacts into electrodes based on spatial proximity and intensity features [33].

3.2.3 Synthetic Data and Transfer Learning

To overcome the scarcity of annotated clinical data, several groups have developed synthetic data generation pipelines. These simulate electrode trajectories, bending, and artifacts, allowing for robust training and testing of segmentation algorithms [24]. Transfer learning and cross-subject models are also being explored to improve generalizability across imaging protocols and patient populations.

3.3 Automation, Software Integration, and Accessibility

3.3.1 Manual Initialization and Fiducial Lists

A common feature of most segmentation pipelines is the requirement for manual initialization:

- **Fiducial Lists**: Users are typically required to provide a list of anatomical landmarks (e.g., entry and tip points) for each electrode, either manually or from preoperative planning.
- **Region-of-Interest Selection**: Some methods require users to delineate regions containing electrodes or to select candidate contacts for further analysis.

This user input is critical for guiding the assignment of segmented contacts to specific electrodes, especially in cases with complex trajectories or overlapping electrodes. Making the process more complex and prone to errors.

3.3.2 Public Software and Clinical Usability

While a number of research toolboxes have been developed, most are not fully automatic nor widely integrated into public, clinician-oriented software. Examples include:

- **SEEGA (SEEG Assistant)**: A 3D Slicer extension that provides semi-

automatic contact segmentation and anatomical labeling, but requires manual fiducial files and user supervision [33].

- **BrainQuake**: An open-source Python toolbox that combines classical and statistical methods for electrode segmentation and localization, also requiring user initialization [33].

3.4 Performance and Validation

3.4.1 Validation Protocols

Validation is commonly performed by comparing automatic segmentation results with manual annotations by clinical experts, using metrics such as Euclidean distance, lateral shift, and root mean square deviation (RMSD). Some studies also report inter-rater variability between manual annotators to contextualize the performance of automatic methods [31].

3.5 Current Limitations

3.5.1 Automation and User Independence

Most published methods are semi-automatic, requiring manual initialization or parameter tuning. Fully automatic, end-to-end solutions that do not require any user input remain rare. Increasing automation is a key focus of ongoing research, with efforts directed at:

- Reducing reliance on manual fiducial lists.
- Improving robustness to anatomical variability and artifacts.
- Integrating multimodal imaging (e.g., MRI/CT fusion) for more accurate anatomical localization.

3.5.2 Accessibility and Clinical Integration

The majority of available tools are research prototypes or open-source toolboxes intended for technically proficient users. Integration into public, clinician-friendly software platforms is limited, and most methods are not yet accessible to non-scientists such as neurosurgeons or clinical staff. Improving usability, documentation, and support for clinical workflows is an important area for future development.

3.5.3 Overall limitations

The algorithms and toolboxes described above have provided valuable inspiration and foundational methods for the field. However, there remains a clear need for more user-friendly, robust, and clinically integrated solutions. A fully automatic, intuitive tool that does not require technical expertise or extensive manual input would have substantial impact, facilitating routine SEEG analysis and improving both efficiency and reproducibility in epilepsy surgery planning.

3.6 Scientific literature review

After the literary review previously explained, a chronology has been stablished depicting the works published during the last decade (Table 4.1).

Table 3.1: *Recent publications on SEEG electrode segmentation and localization, with detailed limitations.*

Title	Authors	Year	Limitations
Automatic segmentation of deep intracerebral electrodes in computed tomography scans. [35]	Arnulfo et al.	2015	Requires manual initialization of electrode positions; performance sensitive to threshold and parameter settings; may be affected by CT artifacts and anatomical variability; limited automation for highly curved or overlapping electrodes; validation on a single-center dataset; integration into clinical workflow may require additional steps.

Table 3.1 – continued from previous page

Title	Authors	Year	Limitations
Automatic segmentation of stereoelectroencephalography (SEEG) electrodes post-implantation considering bending [31]	Granados et al.	2018	Rule-based, parameter-driven algorithm; performance sensitive to parameter tuning and imaging protocol; assumes uniform bending, degrades with severe curvature; may miss contacts obscured by artifacts; limited generalizability across scanners; requires user-provided fiducial list for electrode assignment.
3D U-Net Electrode Detection in CT [34]	Vlasov et al.	2021	High GPU memory and computational requirements; reduced accuracy for contacts at image boundaries or near the skull; model performance may drop with unseen artifact patterns; limited validation on diverse clinical datasets.
BrainQuake: GMM-based Contact Clustering [33]	Arnulfo et al.	2022	Clustering sensitive to initialization and parameter selection; difficulty distinguishing closely spaced or overlapping contacts; requires manual region-of-interest selection; robustness to severe artifacts not fully established.

Table 3.1 – continued from previous page

Title	Authors	Year	Limitations
Hybrid 2D/3D U-Net for Contact Segmentation [36]	Pantović et al.	2022	Limited detection of edge contacts (near skull); synthetic artifact augmentation may not fully reflect real CT noise/artifacts; requires large annotated datasets for training; inference speed depends on hardware; generalizability to new scanners may require retraining.

The literature review reveals significant advances in SEEG electrode localization, with methods ranging from classical thresholding approaches to sophisticated deep learning architectures. However, a critical gap exists between research developments and clinical implementation. Most published algorithms require manual initialization, technical expertise for deployment, or are not publicly available for clinical use. The semi-automatic nature of existing tools, combined with their sensitivity to parameter tuning and limited integration into standard clinical workflows, creates barriers for routine adoption in hospital environments.

Given these limitations and the clinical need for automated electrode localization, this thesis presents an original approach designed specifically for seamless integration into existing medical imaging infrastructure, requiring no manual initialization while maintaining the accuracy and reliability demanded by clinical practice.

4 Materials

This study retrospectively analysed imaging data obtained from a cohort of patients with drug-resistant epilepsy who underwent pre-surgical evaluation at the Epilepsy Unit of Hospital del Mar (Barcelona, Spain) between 2013 and 2019. All procedures were conducted in accordance with the ethical standards of the 25th Declaration of Helsinki, and written informed consent was obtained from all participants prior to data collection and analysis.

4.1 Experimental Data

4.1.1 Patients

The dataset comprises imaging and annotation data from **eight patients** undergoing presurgical evaluation for epilepsy.

Electrode specifications**:** Intracranial electrodes (manufactured by Dixi Médical, *Besançon*, France) with a diameter of 0.8 mm, and composed of 5 to 15 cylindrical contacts (each 2 mm in length, spaced 1.5 mm apart) were used.

Data Composition

Each patient dataset includes the following components:

- **Pre-surgical CT scan:** High-resolution computed tomography (CT) scan acquired before electrode implantation.
- **Post-surgical CT scan:** CT scan acquired after the implantation of SEEG electrodes, used to localize electrode contacts.
- **Magnetic Resonance Imaging (MRI):** Pre-implantation MRI providing detailed anatomical reference for electrode localization and registration.
- **Manual Annotations:** Ground truth electrode contact locations and trajectories, manually marked by expert clinicians using Slicer's markup tools.

Imaging Specifications

- **CT Image Dimensions:** Most CT volumes have a matrix size of $256 \times 256 \times 256$ voxels.

- **Voxel Spacing:** Isotropic spacing of $1 \times 1 \times 1$ mm, ensuring consistent resolution across all axes.
- **Metadata:** Each scan includes a direction matrix, image origin, and other metadata necessary for precise spatial registration and analysis.

Summary Table

Modality	Use	Resolution	Annotations
CT (before surgery)	Baseline scan	256^3, 1 mm	None
CT (after surgery)	Electrode localization	256^3, 1 mm	Electrode positions
MRI	Anatomy reference	Variable, high-res	None
Manual Markups	Ground truth	N/A	Electrode contacts

Table 4.1: *Summary of imaging data and annotations contained in the dataset.*

4.2 Software Tools, Programming Languages, and Libraries

This project made use of a combination of open-source neuroimaging software, programming languages, and computational environments for data processing, analysis, visualization, and extension development.

4.2.1 Software Tools

- **3D Slicer**: Open-source platform for medical image analysis; used for CT/MRI visualization, segmentation, and running Python scripts.
- **Qt Designer**: Used to design GUIs for Slicer modules via a visual drag-and-drop interface.
- **Google Colab**: Cloud-based notebook environment for prototyping and GPU-accelerated computation.

4.2.2 Programming Languages

- **Python**: The main programming language used for data processing, analysis, scripting, and extension development in both 3D Slicer (version 3.9.10) and Google Colab (version 3.10) environments.

4.2.3 Key Libraries and Packages

This project utilized various Python libraries for data processing, visualization, and machine learning:

- **NumPy**, **SciPy**, and **pandas**: For numerical operations and data manipulation.
- **Matplotlib** and **Seaborn**: For visualization.
- **NiBabel**, **pynrrd**, and **SimpleITK**: For handling neuroimaging data and image processing.
- **VTK** and **OpenCV**: For 3D and general image processing.
- **PyTorch**, **MONAI**, **scikit-learn**, **XGBoost**, and **LightGBM**: For deep learning and machine learning tasks.

Table 4.2 lists the main library versions used in Google Colab (June 2025).

Table 4.2: *Library versions used in Google Colab (June 2025)*

Library	Version
NumPy	2.0.2
SciPy	1.15.3
pandas	2.2.2
Matplotlib	3.10.0
NiBabel	5.3.2
pynrrd	1.1.3
scikit-image	0.25.2
MONAI	1.4.0

4.2.4 Reproducibility

All code and workflows were developed using open-source tools to ensure reproducibility and transparency. The implementation is available in a public GitHub repository [37].

5 Methods

5.1 Overview

5.1.1 System Architecture and Clinical Workflow

Post-operative SEEG electrode localization presents three fundamental challenges: (1) distinguishing genuine electrode signatures from extensive metal artifacts, (2) handling patient-specific anatomical and imaging variations, and (3) maintaining seamless clinical workflow integration. The developed methodology addresses these interconnected challenges through a pipeline structured into six sequential stages, each targeting specific clinical and technical requirements of electrode identification via complementary technical approaches.

5.1.2 Pipeline Architecture

Stage 1: Brain Region Extraction (Section 5.2) A MONAI-based 3D U-Net was developed to automatically segment brain tissue from post-operative CT scans. This brain mask serves as an anatomical constraint, preventing false positive detections from extra-cranial metallic artifacts and reducing computational complexity by focusing analysis on relevant brain regions.

Stage 2: Multi-Modal Image Enhancement (Section 5.3) Seven enhancement strategies were implemented to improve electrode visibility by addressing metal artifacts, beam hardening effects, and intensity variations. These approaches include Gaussian filtering, gamma correction, wavelet denoising, and morphological processing. Each method targets different aspects of the electrode detection challenge to handle diverse imaging conditions.

Stage 3: Threshold Prediction Model (Section 5.4) A Random Forest regression model was developed to predict optimal binarization thresholds for each enhanced volume using 38 statistical and morphological features extracted from intensity histograms. This adaptive approach eliminates manual parameter tuning while maintaining performance across different patients and imaging protocols.

Stage 4: Global Voting Ensemble (Section 5.6) The approach generates binary

mask variants per patient and implements a consensus voting mechanism rather than relying on a single segmentation method. This ensemble approach creates redundant coverage around electrode locations, ensuring detection when individual methods fail due to artifacts or challenging anatomy.

Stage 5: Contact Authentication and Localization (Section 5.7) A machine learning pipeline was implemented to distinguish genuine electrode contacts from imaging artifacts using a multi-dimensional feature space encompassing spatial coordinates, intensity characteristics, neighborhood density, and trajectory alignment. The system outputs confidence scores rather than binary classifications, enabling clinicians to exercise judgment about prediction reliability.

Stage 6: Trajectory Reconstruction (Section 5.9) Complete electrode pathways are reconstructed using a hybrid clustering approach that combines DBSCAN spatial analysis with Louvain community detection and adaptive parameter optimization. This multi-algorithm consensus framework identifies linear electrode arrangements while accommodating anatomical constraints and electrode bending. The adaptive parameter system adjusts clustering parameters based on cluster count, noise levels, and electrode size distribution, providing complete trajectory information from cortical entry points to deep brain targets.

5.1.3 Clinical Integration and Human-Machine Collaboration

Conservative Confidence Design: The confidence scoring system was designed with aggressive penalization to prevent overconfidence while preserving clinical utility across the prediction spectrum. Supporting a tiered clinical workflow where physicians can adapt acceptance criteria based on patient-specific factors, as detailed in Section 5.7.4.

3D Slicer Integration: The complete pipeline was implemented as custom extensions within 3D Slicer, enabling integration into existing clinical workflows as described in Chapter 7. The modular design supports both automated processing and interactive refinement.

Validation Strategy: Leave-One-Patient-Out (LOPO) cross-validation was implemented across eight patients to ensure robust generalization assessment without

patient-specific overfitting.

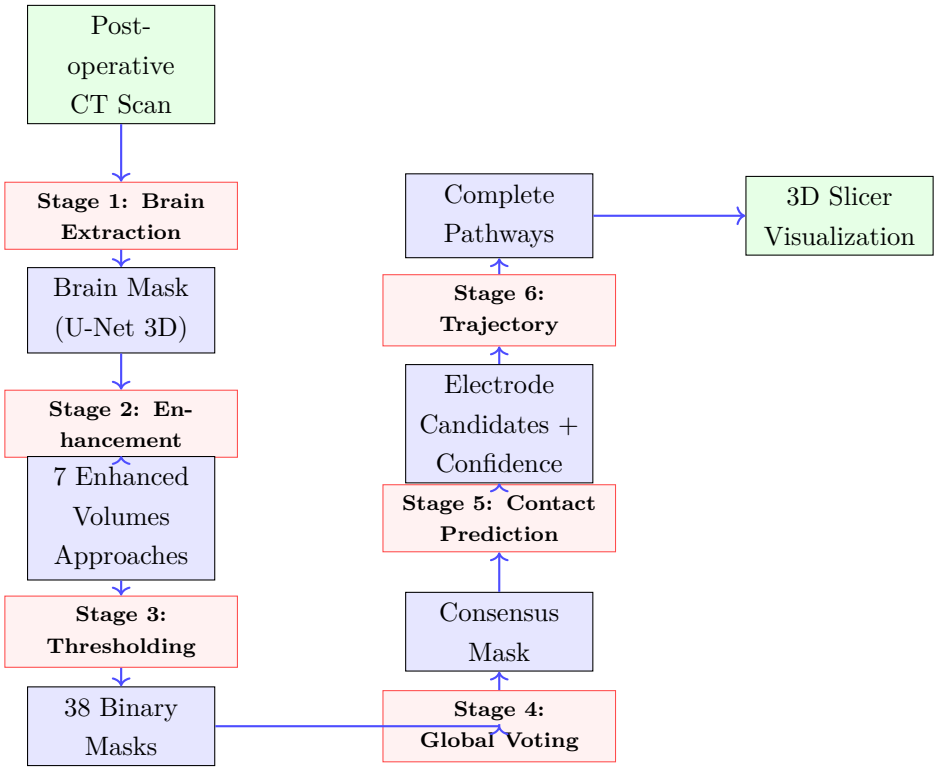


Figure 5.1: *SEEG electrode localization pipeline methodology. The pipeline implements six sequential stages: (1) automated brain tissue segmentation using 3D U-Net, (2) multimodal image enhancement with seven different approaches, (3) adaptive threshold prediction using Random Forest regression, (4) consensus mask generation through global voting of 38 binary variants, (5) electrode contact authentication using machine learning classification, and (6) trajectory reconstruction through hybrid clustering algorithms. Processing time represents computational requirements per patient. Source: Own source.*

The following sections provide detailed technical implementation of each pipeline stage, mathematical formulations, and validation results.

5.2 Brain Region Extraction

5.2.1 MRI-Based Brain Extraction Pipeline

When structural MRI data are available, brain masking is performed through a multi-stage image processing pipeline to isolate the brain region and serve as spatial

constraints for subsequent electrode detection algorithms.

The pipeline consists of three sequential steps: (1) spatial smoothing using a 3D Gaussian kernel ($\sigma = 2$) to reduce noise while preserving anatomical edges, (2) global thresholding via Otsu's method [38] to differentiate brain tissue from non-brain regions through intra-class variance minimization, and (3) morphological refinement using 3D binary closing and hole filling operations to ensure anatomical consistency.

The mathematical formulation follows:

$$I_{\text{smooth}} = G_\sigma * I_{\text{raw}}, \quad T_{\text{otsu}} = \arg\min_T \left[\omega_1(T)\sigma_1^2(T) + \omega_2(T)\sigma_2^2(T)\right] \tag{5.1}$$

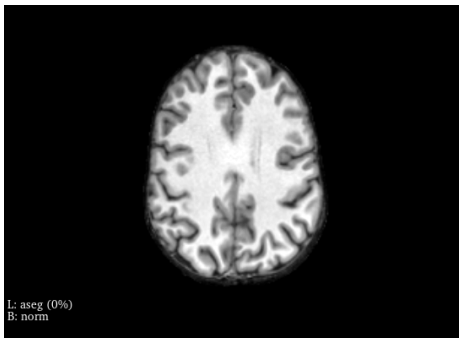
where G_σ represents a 3D Gaussian kernel with standard deviation σ, I_{raw} is the original MRI volume, $\omega_i(T)$ and $\sigma_i^2(T)$ represent the probabilities and variances of brain and non-brain classes respectively.

Morphological operations eliminate gaps and ensure contiguous segmentation:

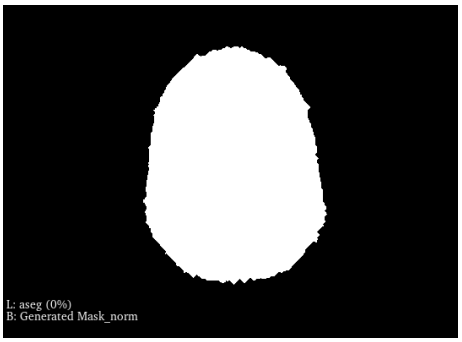
$$M_{\text{final}} = \mathcal{F}(\text{Close}(M_{\text{thresh}}, B)) \tag{5.2}$$

where M_{thresh} is the thresholded binary mask, B is a spherical structuring element with radius one voxel, and $\mathcal{F}$ represents the hole filling operation.

While this MRI-based approach provides accurate brain segmentation when available, clinical SEEG workflows typically rely solely on post-operative CT imaging. The following deep learning framework addresses this primary requirement.



(a) *Axial view of the brain MRI from Patient 1 in the dataset*



(b) *Axial view of the binarized brain region from Patient 1 in the dataset*

Figure 5.2: *Brain extraction results for Patient 1: (a) original MRI and (b) extracted brain mask. Source: Own source.*

5.2.2 Deep Learning-Based CT Brain Segmentation

Given the limitations of classical approaches on post-operative CT scans with electrode artifacts, a deep learning framework was developed to provide brain segmentation directly from CT imaging. This approach forms the primary brain extraction method for the clinical pipeline.

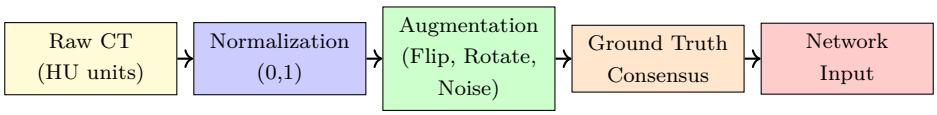


Figure 5.3: *Data preprocessing pipeline for 3D U-Net brain segmentation showing the complete workflow from raw CT input to network-ready data, including the multi-threshold consensus strategy for robust ground truth generation.Source: Own source*

Dataset Characteristics and Preprocessing

The dataset used in this study consists of CT scans and corresponding brain mask annotations for multiple patients. A key feature of this dataset is the multi-threshold approach used for ground truth mask generation. For each patient, seven different mask variants were created, denoted as *mask_1* through *mask_7*, each representing brain segmentations at different threshold levels.

These multiple threshold variants serve several important purposes:

- **Uncertainty Representation:** The different thresholds aim to capture the inherent uncertainty in defining exact brain boundaries in CT images.
- **Consensus Building:** Multiple threshold variants intend to enable the generation of consensus masks through averaging or voting mechanisms.
- **Robustness Enhancement:** Training with multiple variants per patient pursues the improvement of model's ability to generalize across different segmentation criteria.
- **Boundary Precision:** The range of thresholds seeks varying degrees of conservativeness in boundary delineation.

During dataset preparation, CT scans and their corresponding brain mask variants were automatically paired through a systematic file matching process. The pairing algorithm identified each patient's CT scan and collected all seven associated mask variants, creating a dataset structure where each patient entry consisted of one

CT scan linked to seven mask variants. For model training, these variants were processed to generate a consensus ground truth through averaging followed by binary thresholding.

Architecture and Implementation

A deep learning framework was developed for the automatic segmentation of brain regions from CT scans containing SEEG electrodes. The architecture is based on a 3D U-Net model, implemented using the MONAI framework [39] , a specialized toolkit for medical image analysis. The network transforms a single-channel 3D CT volume into a binary brain mask through a series of encoding and decoding operations:

$$M : \mathbb{R}^{1\times D\times H\times W} \rightarrow \mathbb{R}^{1\times D\times H\times W} \tag{5.3}$$

where D represents the depth (number of axial slices), H the height (coronal dimension), and W the width (sagittal dimension) of the 3D CT volume.

The implemented 3D U-Net architecture features:

1. **Encoder Path:** Five hierarchical levels with progressively increasing feature channels (32, 64, 128, 256, 512)
2. **Decoder Path:** Symmetric upsampling operations with skip connections to preserve spatial information
3. **Residual Units:** Two residual units per encoding/decoding level to facilitate gradient flow and improve convergence characteristics
4. **Regularization:** Dropout with a rate of 0.5 applied to deep feature maps to mitigate overfitting
5. **Normalization:** Instance normalization applied after convolutional layers to stabilize training across varying input intensity distributions
6. **Activation:** LeakyReLU activation functions with a negative slope of 0.01 to handle negative values while preventing vanishing gradients

Data Preprocessing and Augmentation

CT volumes undergo a standardized preprocessing pipeline to ensure consistent and normalized network inputs. The normalization procedure is defined as:

$$I_{\text{normalized}} = \text{clip}\left(\frac{I_{\text{raw}} - (-1000)}{2000}, 0, 1\right) \tag{5.4}$$

where I_{raw} represents the original CT volume in Hounsfield Units (HU), $I_{\text{normalized}}$ is the normalized output volume, and the clip function constrains values to the range [0, 1]. This transformation maps HU from the clinically relevant range of [-1000, 1000] to the normalized [0, 1] interval suitable for deep learning models.

To enhance model generalization and robustness, a strategic data augmentation pipeline was implemented:

1. **Spatial Padding:** Volumes are padded to uniform dimensions (multiples of 16) for efficient GPU-based computation
2. **Random Flipping:** Applied with 60% probability along each spatial axis, introducing orientation invariance
3. **Random Rotation:** 90° rotations (up to 270°) with 60% probability to simulate different head orientations
4. **Gaussian Noise:** Added with 30% probability ($\mu = 0, \sigma = 0.1$) to simulate imaging artifacts and scanner variations

This augmentation strategy aims to potentially increase the training set size and improve model generalization.

Training Strategy and Optimization

Loss Function Formulation A composite loss function was designed to address the specific challenges of brain segmentation in CT imaging:

$$\mathcal{L} = 0.8 \times \mathcal{L}_{\text{Dice}} + 0.2 \times \mathcal{L}_{\text{BCE}} \tag{5.5}$$

This weighted combination leverages both the Dice loss ($\mathcal{L}_{\text{Dice}}$) and binary cross-entropy loss ($\mathcal{L}_{\text{BCE}}$). The Dice component prioritizes accurate region-based segmentation and addresses class imbalance, while the BCE component provides pixel-wise supervision. The weighting 0.8/0.2 was empirically determined to optimize performance metrics.

Optimization Parameters

The network was optimized using:

- **Optimizer:** AdamW [40]with 1×10^{-5} weight decay.
- **Learning Rate:** 1×10^{-4}, reduced on plateau (factor 0.5, patience 5).
- **Gradient Clipping:** Max norm of 1.0.
- **Early Stopping:** Stops after 10 epochs without validation improvement.

Leave-One-Out Cross-Validation (LOOCV) Given the limited availability of annotated medical images, a LOOCV strategy was implemented:

$$\text{LOOCV}_i = \text{train}(\{P_j | j \neq i\}), \text{test}(P_i) \tag{5.6}$$

For each fold i, one patient dataset P_i was held out for testing while the model was trained on all remaining patients.

Training for each fold continued for up to 50 epochs with an early stopping mechanism (patience=10) to prevent overfitting. The model checkpoint with the highest Dice score for the validation patient was retained for each fold.

5.2.3 Inference Pipeline and Post-processing

Model Inference

During inference, CT volumes not used during training undergo identical preprocessing steps to ensure consistency with training data.

The binary segmentation mask is generated from the raw network output using:

$$M_{\text{binary}} = \mathbb{I}(\sigma(M_{\text{raw}}) > 0.5) \tag{5.7}$$

where M_{raw} is the raw network output, σ represents the sigmoid activation function that converts logits to probabilities, 0.5 is the binary classification threshold, $\mathbb{I}$ is the indicator function, and M_{binary} is the final binary brain mask.

Post-processing Refinement

To improve anatomical accuracy, the binary prediction is refined by:

1. **Removing small components** (<100 voxels) to reduce noise.
2. **Filling internal holes** for mask continuity.
3. **Applying morphological smoothing** (optional) to refine boundaries.

These refinement steps significantly improve the quality of the final segmentation, especially in cases with imaging artifacts or noise.

5.2.4 Evaluation Methodology

Quantitative Metrics

Model performance was primarily evaluated using the Dice similarity coefficient:

$$\text{DSC} = \frac{2|X \cap Y|}{|X| + |Y|} = \frac{2TP}{2TP + FP + FN} \tag{5.8}$$

where X and Y represent the predicted and ground truth masks respectively. This metric quantifies the spatial overlap between the predicted segmentation and the manual annotation, ranging from 0 (no overlap) to 1 (perfect overlap).

Multi-planar Visualization

A visualization pipeline was developed to facilitate qualitative assessment of segmentation results across orthogonal planes. For each patient, the following visualizations were generated:

- Central axial, coronal, and sagittal slices with predicted and ground truth mask overlays
- Direct visual comparisons between predictions and ground truth
- Overlay visualizations to assess anatomical correspondence

These visualizations enable efficient qualitative assessment and identification of systematic patterns in model performance.

Final Model Training Strategy

Following cross-validation analysis, a final deployment model is trained using the complete dataset. This approach leverages all available training data while incorpo-

rating validated hyperparameters from the LOOCV evaluation:

- **Reduced learning rate:** 1×10^{-5} to enable fine-tuning and prevent overfitting on the complete dataset
- **Extended training duration:** 20 epochs to allow convergence with the lower learning rate
- **Weight initialization:** Pretrained weights from the best-performing LOOCV fold to leverage proven feature representations

This training strategy represents the optimal configuration for clinical deployment, combining maximal utilization of training data with validated hyperparameters.

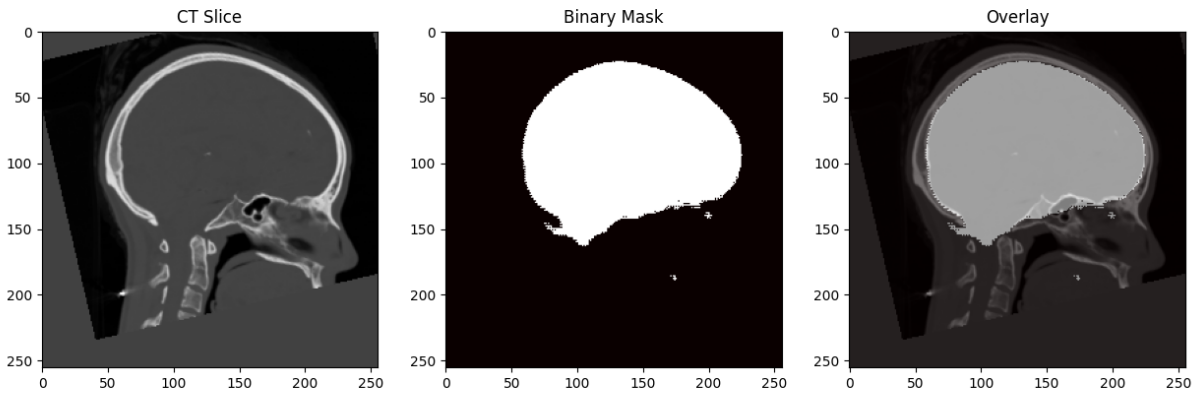


Figure 5.4: *Deep learning brain segmentation pipeline: (left) input CT scan with SEEG electrodes, (right) generated brain mask. Source: Own source.*

This deep learning-based brain segmentation provides the anatomical constraints required for Stage 1 of the pipeline. The resulting brain masks enable subsequent electrode detection algorithms to focus on anatomically plausible regions while eliminating false positives from extra-cranial artifacts. Having established reliable brain boundaries, the pipeline proceeds to Stage 2: Multi-modal image enhancement for electrode visibility optimization.

5.3 Image Processing Pipeline

Detecting electrodes in CT scans is hindered by metal artifacts and imaging noise, which can closely resemble electrode signatures. This challenge is addressed by an image processing pipeline consisting of three sequential stages: (1) the brain region is isolated to focus computational resources, (2) electrode visibility is enhanced through seven complementary methods, and (3) adaptive thresholds are selected based on each patient's imaging characteristics.

5.3.1 Brain Extraction and ROI Processing

The pipeline begins with precise brain region isolation to reduce computational complexity and eliminate extra-cranial artifacts. Given an input CT volume $V_{orig} \in \mathbb{R}^{N_x \times N_y \times N_z}$ and binary brain mask M_{brain}, the region of interest (ROI) is extracted through element-wise multiplication:

$$V_{ROI} = V_{orig} \odot M_{brain} \tag{5.9}$$

where $\odot$ denotes the Hadamard product. The mask is derived through either co-registered MRI segmentation or manual annotation. Robust morphological processing ensures mask integrity through:

$$M_{processed} = \phi_{close}(\phi_{fill}(M_{brain}), B_{sphere}(r)) \tag{5.10}$$

where ϕ_{fill} performs hole-filling, ϕ_{close} is morphological closing with a spherical structuring element B_{sphere}, and r is the structuring element radius.

5.3.2 Multi-Modal Enhancement Approaches

Post-operative CT scans present a complex imaging environment where genuine electrode signatures must be distinguished from extensive artifacts and noise. The challenge stems from multiple confounding factors that characterize post-surgical imaging:

- **Metal streak artifacts** from the electrodes themselves
- **Beam hardening effects** that distort intensity values
- **Patient movement artifacts** during acquisition

- **CT reconstruction noise** inherent to the imaging process

These factors create scenarios where authentic electrode signatures can be obscured by noise or mimicked by spurious high-intensity regions throughout the scan.

Enhancement Strategy Design

To address these challenges systematically, seven distinct enhancement strategies were implemented and evaluated, each designed to tackle specific aspects of the noise-signal separation problem (Table 5.1). The overarching methodology follows an iterative approach combining complementary image processing operations:

Noise Reduction: Gaussian filtering and wavelet denoising reduce imaging noise while preserving critical electrode edge information.

Contrast Enhancement: Gamma correction amplifies high-intensity metallic regions, improving electrode conspicuity relative to background tissue.

Edge Enhancement: Morphological operations and sharpening filters accentuate electrode boundaries, making metallic contacts more distinct from surrounding tissue.

Spatial Constraint: ROI masking eliminates extra-cranial artifacts and constrains processing to anatomically relevant brain regions.

Table 5.1: *Enhancement Approaches and Processing Sequences*

Approach	Processing Sequence
Original CTP	Gaussian → Gamma → Sharpening
ROI + Gamma	Masking → Gamma → CLAHE/Wavelet
ROI Only	Direct ROI processing
ROI + Gamma After	Masking → Gaussian → Gamma
Wavelet ROI	Wavelet denoising on ROI
Original Idea	Gaussian → Gamma → Top-hat
First Try	Top-hat → Gaussian → Morphological

Implementation Approach

The seven strategies represent different philosophies for balancing competing requirements in medical image processing: preserving genuine electrode signals while suppressing artifacts, enhancing edges without amplifying noise, and maintaining

computational efficiency for clinical deployment. Table 5.1 shows how each approach sequences these operations to target different imaging scenarios.

Each enhanced volume undergoes patient-specific adaptive binarization optimized for individual scan characteristics, with threshold values determined through the machine learning framework described in Section 5.4.

Detailed mathematical formulations and parameter specifications are provided in Appendix A.2.

5.3.3 Histogram-Based Feature Analysis

The seven enhancement approaches 5.3.2 generate multiple candidate volumes per patient, each potentially optimal under different imaging conditions

To address this challenge and create a predictive thresholding model, an analysis of intensity histograms was performed across all 38 enhanced volumes per patient. This approach recognizes that optimal electrode detection depends on both enhancement technique selection and adaptive threshold determination, which account for patient-specific intensity distributions and imaging characteristics.

From each enhanced volume, 23 quantitative features were extracted, specifically designed to capture the statistical and structural properties of intensity distributions that correlate with electrode visibility and segmentation success. These features serve as the foundation for threshold prediction, enabling the system to automatically adapt to varying imaging conditions while maintaining consistent electrode detection performance.

The extracted features encompass three complementary categories:

1. **Intensity statistics (15 features):** mean, variance, standard deviation, percentiles (5th, 10th, 25th, 50th, 75th, 90th, 95th, 99th, 99.5th, 99.9th), entropy, minimum, maximum, range, median
2. **Distribution metrics (8 features):** skewness, kurtosis, modality, coefficient of variation, interquartile range, etc.
3. **High-intensity features (15 features):** count above various thresholds (99th, 99.5th, 99.9th percentile), peak characteristics, contrast ratios, etc.

This feature-driven approach transforms the subjective task of threshold selection into an objective, data-driven process that can generalize across different patients,

imaging protocols, and electrode configurations.

Entropy is computed as:

$$\text{Entropy} = -\sum_{k=0}^{255} p(k) \log_2 p(k), \quad p(k) = \frac{h(k)}{\sum h} \tag{5.11}$$

where $h(k)$ is the histogram count at intensity k and $p(k)$ is the normalized histogram. This entropy measure quantifies the information content of intensity distributions, with higher values indicating greater complexity in electrode visibility patterns.

5.3.4 Optimal Percentile Estimation for Repeated Measurements

While machine learning models provide intelligent threshold prediction based on histogram features, robust clinical deployment requires fallback mechanisms for cases where predicted thresholds fall outside the valid intensity range of a given volume. To address this challenge, an optimal percentile estimation framework was developed to provide reliable backup thresholds when model predictions exceed histogram bounds.

To estimate an optimal percentile from repeated measurements of a specific subject or identifier, a multi-metric strategy was applied. This approach accounts for sample distribution characteristics, robustness to outliers, and statistical confidence.

Data Preparation and Filtering

For each repeated instance (e.g., same identifier across multiple rows), valid percentile values were extracted using the following formula:

$$\text{percentile} = \frac{|\{x_i \in [\min, \max] \mid x_i \leq \text{threshold}\}|}{|\{x_i\}|} \cdot 100 \tag{5.12}$$

where x_i represents uniformly sampled values between the minimum and maximum bounds. Instances where max $\leq$ min were excluded due to invalid range definitions.

Summary Statistics

Descriptive statistics were calculated, including:

- **Mean** (μ), **Median** ($\tilde{x}$), and **Trimmed Mean** (10%) for central tendency.
- **Standard Deviation** (σ) and **Coefficient of Variation (CV)**:

$$\text{CV} = \frac{\sigma}{\mu} \cdot 100 \tag{5.13}$$

  for dispersion assessment.
- **Interquartile Range (IQR)**, used to filter outliers via the 1.5×IQR rule. From the remaining values, a *robust mean* and *robust median* were derived.

Confidence Score Calculation

A confidence score was assigned to reflect the reliability of the selection:

$$\text{Confidence Score} = \min\left(100, 60 \cdot \text{Consistency Score} + 40 \cdot \text{Sample Size Factor}\right) \tag{5.14}$$

where the Consistency Score equals $1/(1 + \text{CV}/100)$ with CV representing the coefficient of variation, the Sample Size Factor is computed as $\min(1, n/5)$ with n denoting the sample size, and the weighting factors 60 and 40 represent the relative importance of consistency versus sample size in the final confidence assessment.

Method Selection Logic

A decision rule was applied to determine the most reliable statistical estimate by weighting each method (mean (μ), median ($\tilde{x}$), trimmed mean, robust mean) .

Sample size considerations favor median-based estimators for smaller datasets, as they provide more stable estimates with limited observations. When the coefficient of variation indicates high dispersion and potential instability in the data, robust metrics are preferred over traditional measures. Additionally, the presence of outliers, detected through the 1.5×IQR rule, shifts preference toward trimmed or robust statistical approaches that minimize the influence of extreme values.

Final Output

The final output included the selected method, the optimal percentile value, a textual justification of the selection, the confidence score (from 0 to 100), and a dictionary with supporting statistics (mean, median, trimmed mean, robust mean, standard deviation, CV, sample size, confidence interval, and number of outliers detected).

These feature extraction and percentile estimation methods provide the foundation for Stage 3 of the pipeline: Adaptive threshold prediction using machine learning models trained on histogram features (Section 5.4) to automatically determine optimal binarization parameters for each enhanced volume, with percentile-based fallback mechanisms ensuring robust operation across all imaging conditions.

5.4 Threshold Binarization Model

Analysis of optimal threshold values across the seven enhancement approaches revealed substantial patient-specific variability, with optimal thresholds ranging from 203 HU to 2886 HU depending on imaging conditions and electrode visibility. This variability made fixed thresholding approaches inadequate for robust clinical deployment. To address this challenge, a machine learning regressor was developed to predict patient-specific optimal thresholds based on intensity distribution characteristics, enabling automated binarization while maintaining segmentation quality across diverse imaging conditions.

5.4.1 Adaptive Feature Engineering Strategy

Beyond the 23 base histogram-derived metrics, 15 additional features were engineered to capture intensity distribution relationships correlating with optimal thresholds. This strategy recognizes that optimal electrode detection depends on relationships between different parts of the intensity distribution, not just absolute values

Category	Feature	Formula
Range Features	Intensity range	$\texttt{range} = \texttt{max} - \texttt{min}$
	Mean-centered min	$\texttt{mean\_centered\_min} = \texttt{min} - \texttt{mean}$
	Mean-centered max	$\texttt{mean\_centered\_max} = \texttt{max} - \texttt{mean}$
Interquartile Metrics	IQR	$\texttt{p75} - \texttt{p25}$
	IQR-to-STD ratio	$\texttt{iqr}/(\texttt{std} + 1e{-}5)$
	Range-to-IQR ratio	$\texttt{range}/(\texttt{iqr} + 1e{-}5)$
Contrast & Sharpness	Contrast per IQR	$\texttt{contrast\_ratio}/(\texttt{iqr} + 1e{-}5)$
	Peak-to-height ratio	$\texttt{non\_zero\_peak2\_value}/(\texttt{non\_zero\_peak2\_height} + 1e{-}5)$
Entropy-Related	Entropy per range	$\texttt{entropy}/(\texttt{range} + 1e{-}5)$
	Entropy-IQR interaction	$\texttt{entropy} \times \texttt{iqr}$
Nonlinear Transforms	Squared skewness	$\texttt{skewness}^2$
	Log-transformed kurtosis	$\log(1 + \texttt{kurtosis} - \min(\texttt{kurtosis}))$

Table 5.2: *Engineered Features for Threshold Binarization Model*

5.4.2 Model Architecture and Methodological Rationale

The threshold prediction model was designed to handle the specific challenges of medical imaging data: intensity variations across scanners, patient-specific anatomical differences, and imaging artifacts that create outliers in feature distributions. The architecture emphasizes robustness through consensus-based outlier detection and ensemble learning approaches.

Data Preparation and Quality Control

Features were filtered to retain only numerical predictors relevant to threshold prediction. Quasi-constant features (>99% identical values) and highly correlated features ($|r| > 0.95$) were eliminated to reduce redundancy and improve model interpretability. The final dataset was partitioned using an 80/20 train-test split.

Feature Analysis and Validation

Statistical analysis was performed on the 38-dimensional feature space to assess data quality and inform model selection. Descriptive statistics highlighted feature distributions and potential skewness. Correlation analysis identified key intensity features related to threshold values. Visualization with distribution plots and heatmaps revealed nonlinear patterns and important feature interactions for accurate threshold prediction.

Complete preprocessing specifications, mathematical formulations (including skewness calculation, correlation coefficients, and ANOVA F-statistics), and feature analysis details are provided in Appendix A.3.

5.4.3 Consensus-Based Outlier Detection

Medical imaging data contains systematic outliers from scanner calibration drift, patient movement artifacts, and genuine anatomical variants that appear statistically anomalous but represent valid biological diversity. SEEG electrode imaging compounds these challenges through metal artifacts and beam hardening effects creating spurious high-intensity regions.

To ensure robustness against these multifaceted outlier sources, a consensus-based detection strategy was implemented combining four complementary algorithms: Isolation Forest, Elliptic Envelope, Local Outlier Factor, and One-Class SVM. Each method produces binary outlier classifications, with final consensus requiring agreement from at least three of the four detectors (Figure 5.5).

This conservative voting approach enhances robustness compared to individual detector methods while avoiding false flagging of legitimate anatomical variants. Data preprocessing included robust scaling and Yeo-Johnson transformation to stabilize variance before outlier analysis.

Complete mathematical formulations and algorithm implementations are provided in Appendix A.3.3.

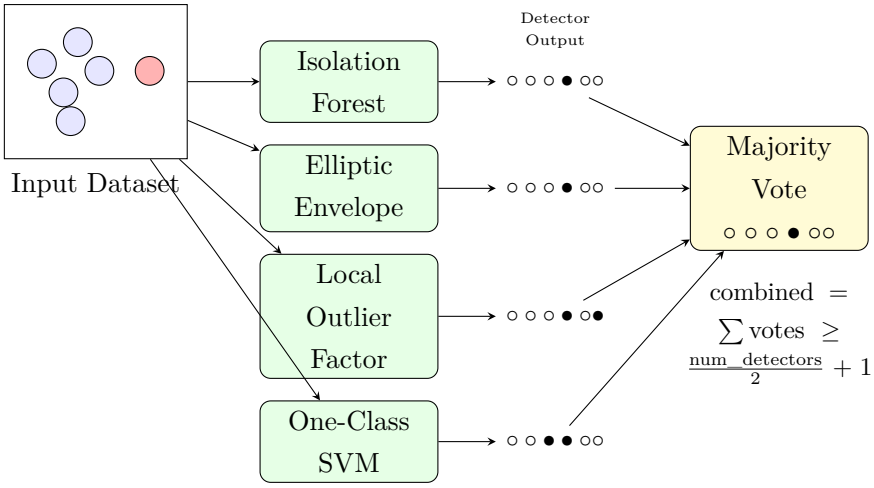


Figure 5.5: *Consensus-based outlier detection pipeline combining four complementary algorithms. Input data points are processed through Isolation Forest, Elliptic Envelope, Local Outlier Factor, and One-Class SVM detectors, each producing binary classifications (blue dots = normal, red dots = outlier). The majority voting mechanism classifies points as outliers (red) only when at least three of the four methods agree, enhancing robustness compared to individual detector approaches. The mathematical criterion* $\sum votes \geq \frac{num\_detectors}{2} + 1$ *ensures conservative outlier identification. Source: Own source.*

5.4.4 Random Forest Regression and Hyperparameter Optimization

Random Forest regression was selected for its proven effectiveness in biomedical applications: ability to model complex nonlinear relationships, natural resilience to overfitting, and robust handling of mixed feature types [41]. The ensemble approach combines multiple decision trees trained on bootstrap samples, with final predictions computed as the arithmetic mean across all trees (Figure 5.6).

Hyperparameter optimization employed Bayesian methods with Tree-structured Parzen Estimator (TPE) [42] to search the parameter space. The optimization objective balanced overall performance with outlier robustness through a composite loss function. A five-fold cross-validation was performed.

Complete mathematical formulations, hyperparameter specifications, and optimization details are provided in Appendix A.3.4.

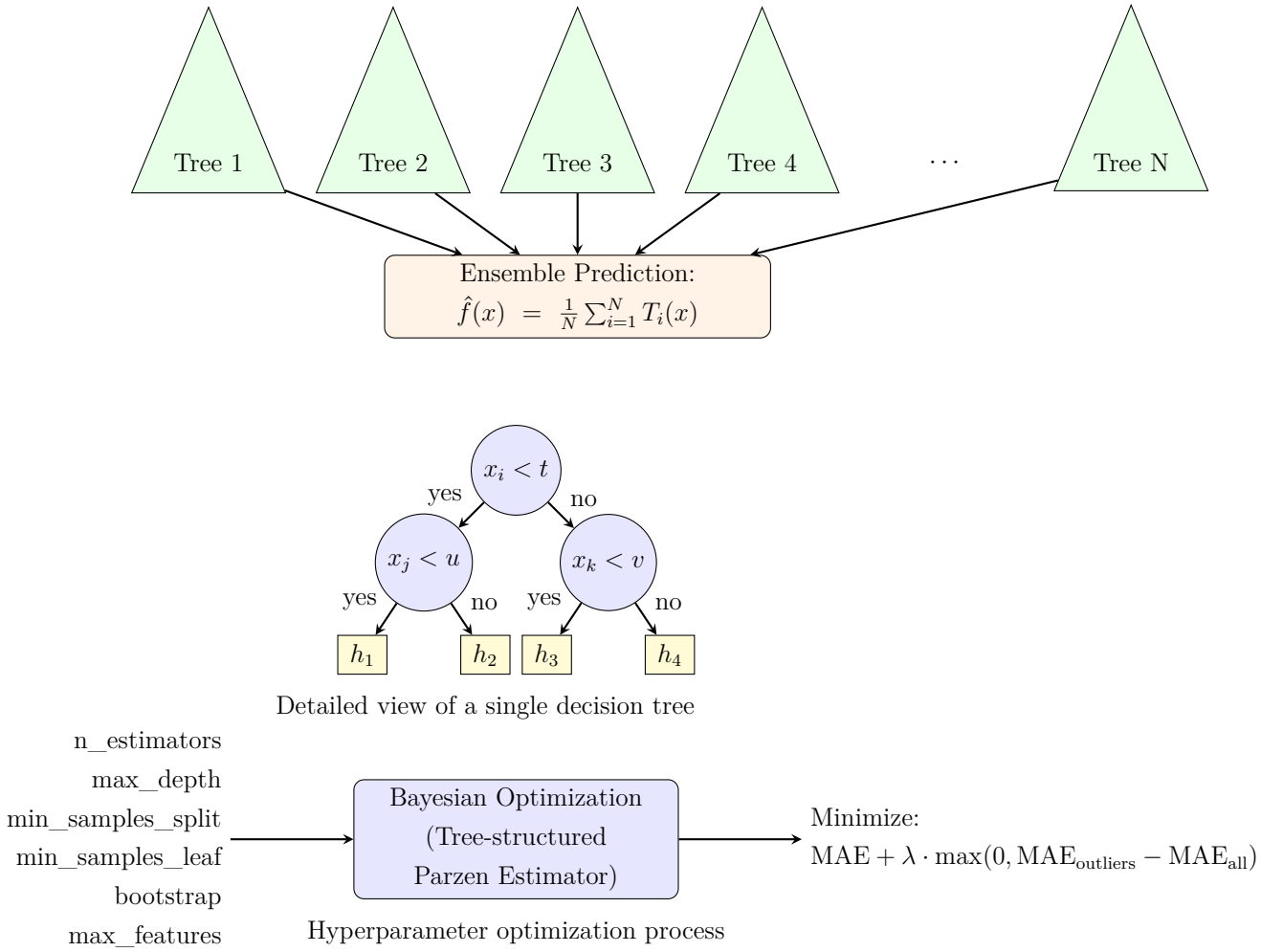


Figure 5.6: *Random Forest architecture and hyperparameter optimization. The model combines predictions from multiple decision trees, with hyperparameters optimized using Bayesian methods to balance performance on normal samples and outliers. Source: Own source.*

5.4.5 Model Evaluation

Model performance was assessed with three complementary metric categories: error quantification, goodness-of-fit assessment, and outlier-specific performance analysis (Figure 5.7).

Performance was evaluated using error quantification (RMSE, MAE), goodness-of-fit (R^2), and outlier-specific metrics, with visualization providing qualitative insights into model behavior (Figure 5.7)

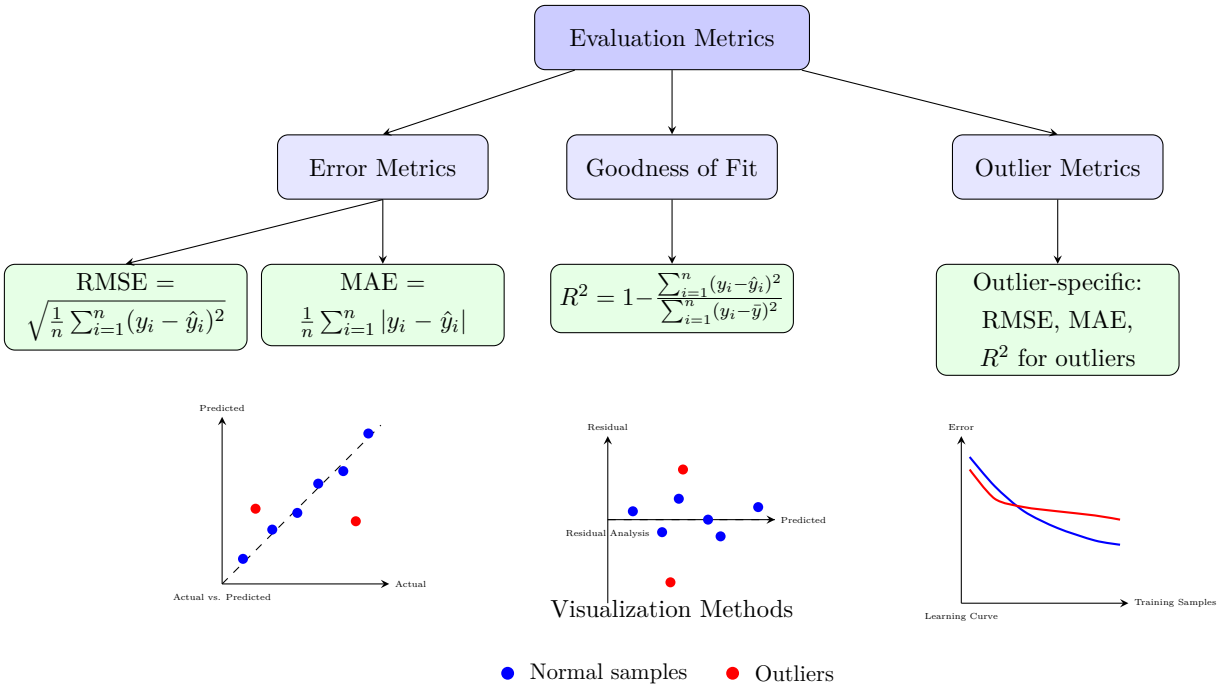


Figure 5.7: *Evaluation metrics and visualization methods used to assess model performance. The framework includes standard error measurements, goodness of fit metrics, and specialized metrics for outlier performance. Source: Own source.*

Mathematical formulations and detailed metric specifications are provided in Appendix A.3.5.

5.5 Model Application in CTP Binarization

5.5.1 Cross-Validation Strategy

The threshold prediction model was validated using leave-one-patient-out cross-validation.

5.5.2 Pipeline Architecture

The binarization framework comprises three key stages (Pipeline 5.8):

1. **Image Enhancement** Seven enhancement strategies combining various filtering, contrast enhancement, and morphological operations (detailed in Table 5.1).

2. **Feature Extraction**
   The final feature set of 38 statistical descriptors combines the base 23 histogram-derived metrics (Section 5.3.3) with the 15 additional engineered features (Table 5.2). This feature space captures both core intensity distribution characteristics and complex nonlinear interactions essential for robust threshold prediction across diverse imaging conditions.

$$\mathcal{F} = \left\{ \mu, \tilde{x}, \sigma^2, P_{25}, P_{75}, P_{99}, H_{\text{shannon}}, \gamma_1, \gamma_2, \ldots \right\} \tag{5.15}$$

3. **Threshold Prediction**
   Constrained random forest prediction:

$$\tau_{\text{pred}} = \max\left(\min\left(\mathcal{M}(\mathcal{F}), \tau_{\max}\right), \tau_{\min}\right) \tag{5.16}$$

   where τ_{pred} is the final predicted threshold value, $\mathcal{M}(\mathcal{F})$ represents the trained Random Forest model prediction given feature vector $\mathcal{F}$, $\tau_{\max}$ and $\tau_{\min}$ are the maximum and minimum allowable threshold bounds, and the nested min-max operations ensure the prediction falls within physiologically reasonable limits.

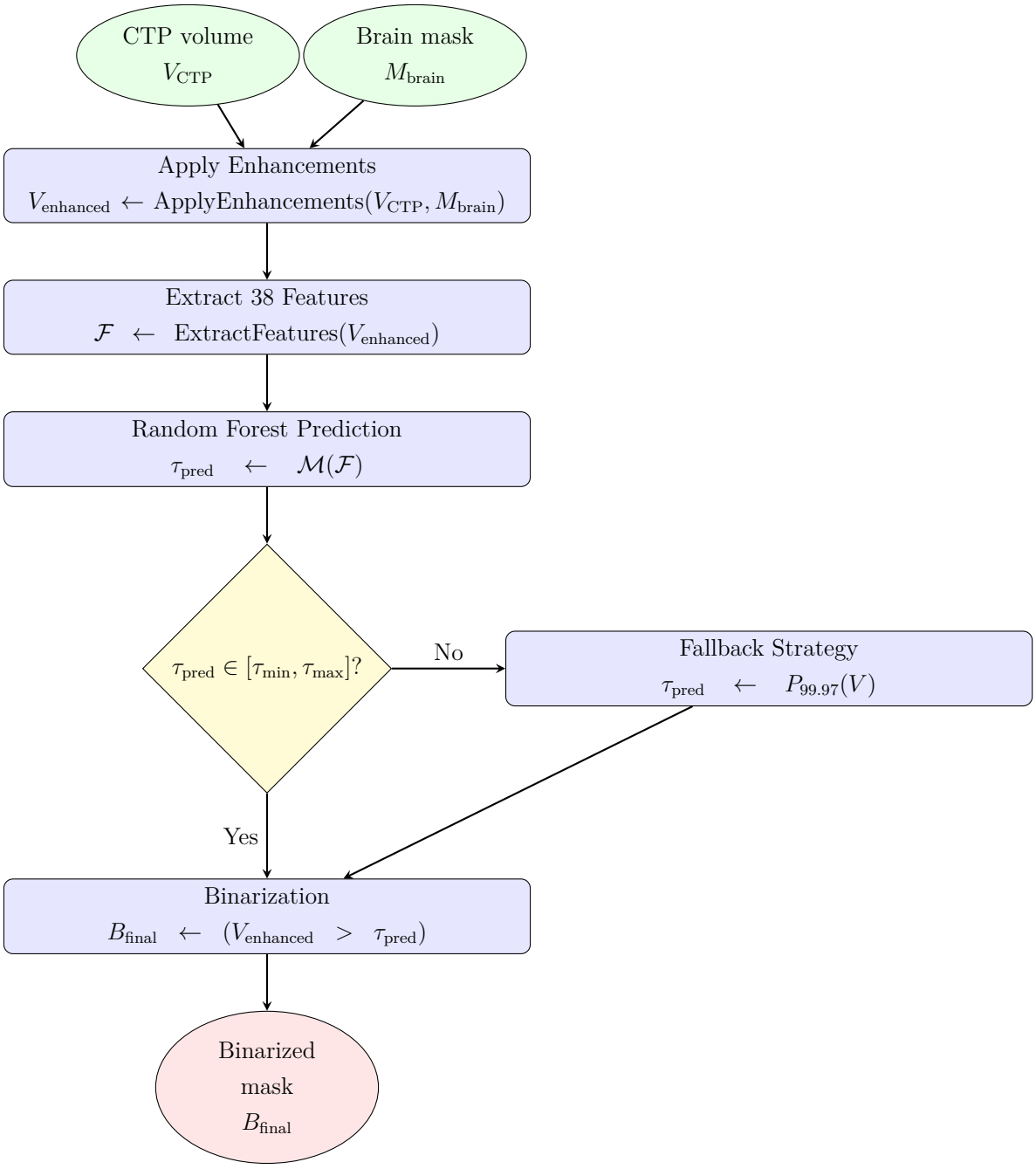


Figure 5.8: *Model-Based Binarization Pipeline showing the complete workflow from input volumes to final binary mask generation through enhancement, feature extraction, threshold prediction with Random Forest model, and constrained binarization with fallback strategy. Source: Own source.*

5.5.3 Validation Methodology

The binarization pipeline was evaluated using established medical image segmentation metrics assessing both volumetric overlap and spatial accuracy. Three complementary measures provided segmentation quality assessment.

The Dice Similarity Coefficient (DSC) quantifies spatial agreement between predicted and reference segmentations, providing robust overall accuracy assessment widely used in medical imaging validation. The Hausdorff Distance (HD) captures maximum spatial deviation between boundaries, identifying worst-case localization errors impacting clinical interpretation. The Volumetric Overlap Error (VOE) measures relative volume of incorrectly classified regions, offering an alternative accuracy perspective.

Performance evaluation metrics:

$$\text{DSC} = \frac{2|X \cap Y|}{|X| + |Y|} \tag{5.17}$$

$$\text{HD} = \max\left(\sup_{x \in X} \inf_{y \in Y} d(x, y), \sup_{y \in Y} \inf_{x \in X} d(x, y)\right) \tag{5.18}$$

$$\text{VOE} = 1 - \frac{|X \cap Y|}{|X \cup Y|} \tag{5.19}$$

where X and Y represent predicted and ground truth binary masks respectively, and $d(x, y)$ is the Euclidean distance between points. DSC ranges from 0 (no overlap) to 1 (perfect agreement), HD provides distance measurements in millimeters with lower values indicating better boundary accuracy, and VOE ranges from 0 to 1 with smaller values indicating superior performance.

5.6 Global Voting Mechanism for Mask Ensemble

Single segmentation methods fail unpredictably in SEEG electrode detection due to metal artifacts, intensity variations, and imaging noise, manifesting as missed contacts, inconsistent boundaries, and false positives. A global voting mechanism addresses these limitations by combining multiple independent segmentation approaches into a robust consensus framework.

In this thesis, rather than relying on any single segmentation method, multiple

binary masks are generated per patient using diverse enhancement strategies and threshold selection approaches. A voting system identifies regions where multiple methods agree, creating consensus boundaries that individual method failures cannot compromise.

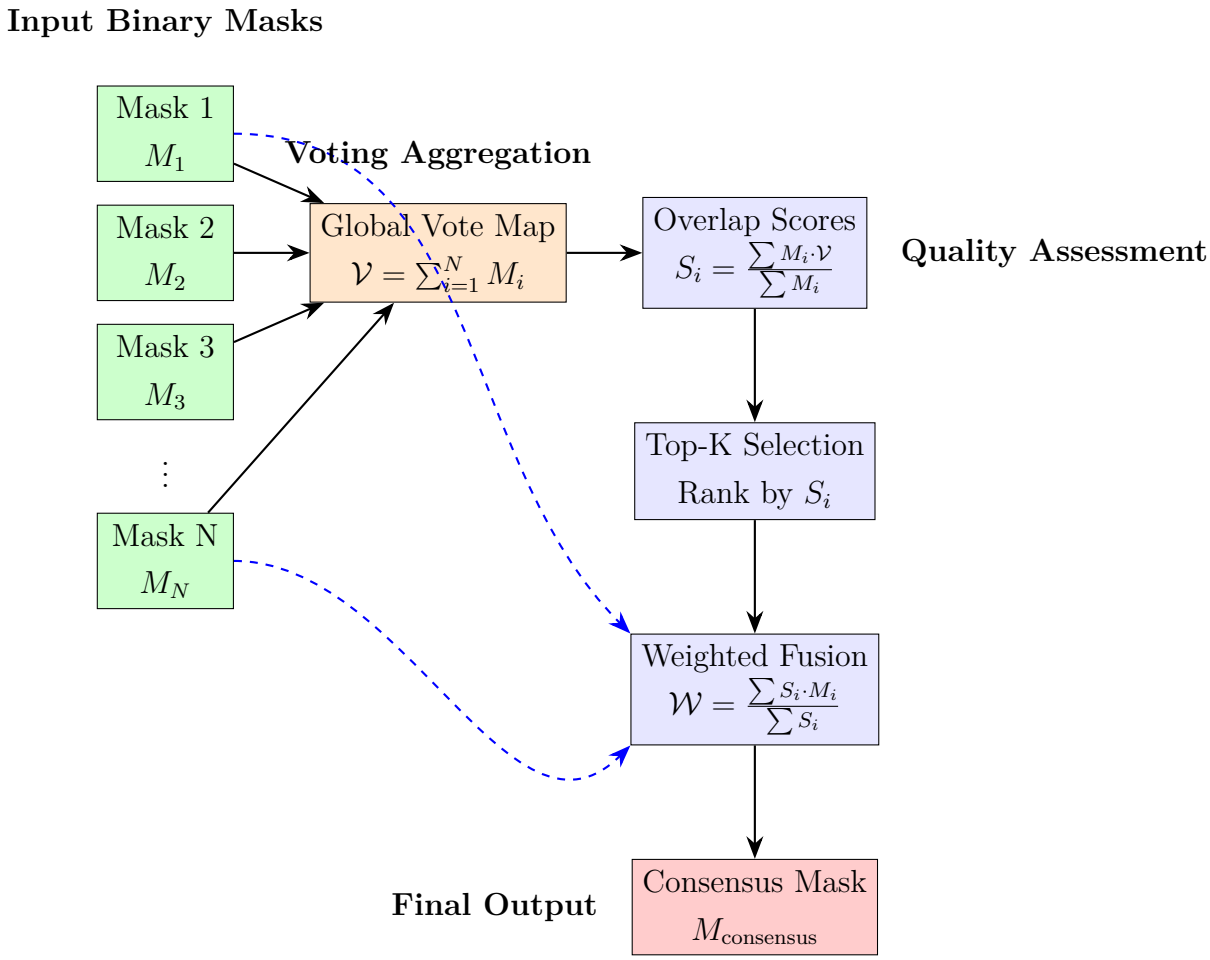


Figure 5.9: *Multi-stage ensemble fusion pipeline for robust SEEG electrode segmentation. The methodology combines N binary masks through: (1) global voting aggregation to identify consensus regions, (2) overlap scoring to assess individual mask quality, (3) top-K selection to identify best-performing masks, and (4) weighted fusion to generate the final consensus mask. Color coding: green indicates input binary masks, orange represents voting aggregation, blue denotes quality assessment and processing steps, and red highlights the final consensus output. Dashed arrows show the flow of selected masks to the weighted fusion stage. Source: Own source.*

5.6.1 Ensemble Architecture

The ensemble combines multiple binary masks through voting to achieve electrode detection. Masks are generated using model-based thresholding (Section 5.4) and empirical percentile approaches. A spatial vote map quantifies agreement across methods, where high-vote regions indicate reliable electrode positions and isolated detections represent artifacts. Quality-weighted selection emphasizes top-performing masks based on overlap scoring, while progressive fusion analysis determines the optimal ensemble size for reliable segmentation with minimal computational cost.

5.6.2 Clinical Integration

The voting mechanism transforms unreliable individual segmentations into clinical tools through systematic redundancy. Multiple independent methods ensure that challenging imaging conditions cannot compromise overall detection reliability, while providing spatial confidence estimates for clinical assessment 5.10. *Mathematical formulations and implementation details: Appendix A.3.6*

5.6.3 Coordinate System Handling

Accurate electrode localization requires precise coordinate transformations between computational environments. The pipeline implements dual transformation architectures supporting 3D Slicer (VTK-based) and external platforms (SimpleITK-based), with automated conversion between indexing conventions and LPS-to-RAS coordinate systems while maintaining sub-millimeter accuracy (Appendix A.1).

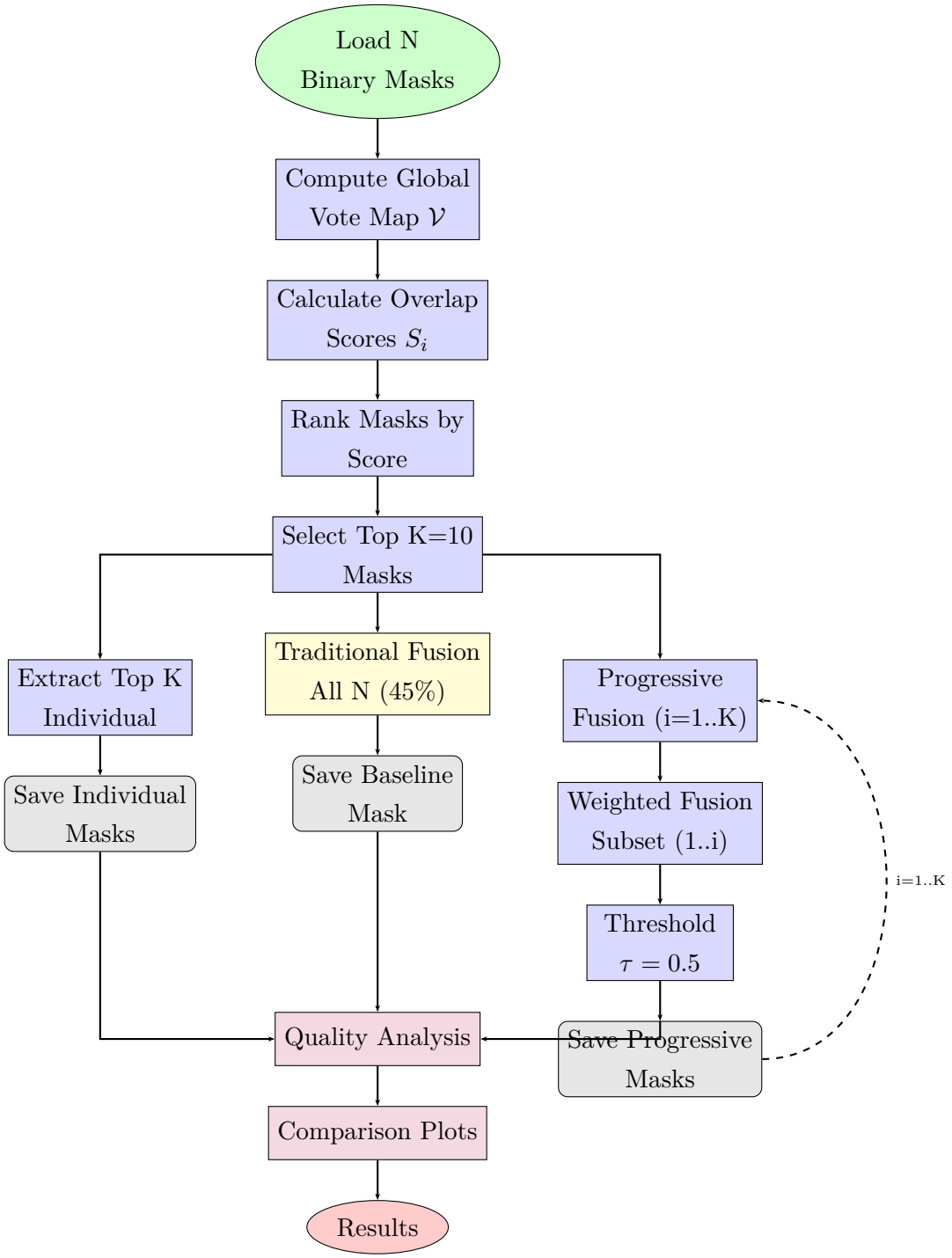


Figure 5.10: *Mask fusion pipeline implementing three parallel strategies for SEEG brain segmentation: individual top-K extraction, traditional baseline fusion (all N masks, 45% threshold), and progressive weighted fusion (incremental subsets i=1..K). The workflow enables systematic comparison of ensemble approaches through analysis and performance evaluation. Color coding: green indicates input data, blue represents processing steps, yellow highlights the baseline comparison method, purple denotes the progressive fusion pathway, and gray shows output storage operations. Solid arrows indicate sequential workflow steps, while the dashed arrow represents the iterative loop (i=1..K) in the progressive fusion strategy. Source: Own source.*

5.7 Points of contact predictor model

Following the global voting mechanism and mask ensemble generation, a critical challenge remains: distinguishing genuine SEEG electrode contacts from the thousands of false candidates generated by imaging artifacts. A novel regression-based authentication system was developed to address this fundamental bottleneck while preserving clinical decision-making flexibility through probabilistic confidence scoring

5.7.1 Clinical Problem and Methodological Foundation

The Authentication Challenge

The automated identification of electrode contacts represents a critical bottleneck where thousands of candidate points must be evaluated per patient, with only 10-15% representing authentic electrode positions. In our cohort, this translates to evaluating approximately 4,000-6,000 candidates per patient to identify 80-120 genuine electrode contacts, representing a 95% false positive challenge that manual review cannot practically address

Metal artifacts from beam hardening and photon starvation effects generate extensive streaking patterns that mimic electrode intensity signatures across multiple voxels. Simultaneously, anatomical structures including calcifications, bone fragments, and high-density vascular components exhibit similar Hounsfield unit values (1,500-3,000 HU) to platinum-iridium electrodes, creating visually indistinguishable intensity profiles. The challenge was further complicated by partial volume effects, where sub-voxel electrode contacts (0.8 mm diameter) within $1\times1\times1$ mm resolution voxels create complex intensity distributions that defy traditional segmentation approaches. Additionally, variations in imaging protocols, reconstruction algorithms, and electrode manufacturers produce diverse artifact patterns requiring robust discrimination strategies across different clinical environments.

Proposal: Confidence-Based Authentication

A regression-based authentication framework transforms the contact validation challenge from binary classification into graduated clinical decision support. Instead of definitive contact/non-contact decisions, the system generates continuous confidence

scores $P(c_i \in \mathcal{C}) \in [0, 1]$ for each candidate point c_i, where $\mathcal{C}$ represents the set of authentic electrode contacts.

This design preserves clinical autonomy by providing guidance rather than decisions, enabling physicians to exercise judgment based on patient-specific factors. The confidence scoring employs patient-specific feature analysis to handle varying imaging conditions and anatomical configurations, with flexible acceptance criteria that adapt to different clinical scenarios.

The framework supports a tiered workflow: high-confidence predictions enable direct clinical use, moderate-confidence cases prompt clinical review, and low-confidence detections require manual validation. This human-machine collaboration preserves clinical decision-making authority while providing intelligent automation support.

5.7.2 Data Preparation and Validation Pipeline

To enable robust confidence-based authentication of electrode candidates identified through global voting, a data preparation and validation framework was developed to systematically extract, validate, and standardize electrode contact coordinates from the ensemble of segmentation masks. This preprocessing stage ensures data integrity and establishes the foundation for machine learning model development through rigorous quality control and clinical validation procedures.

Coordinate Extraction and Standardization

The data preparation pipeline operates on the collection of binary masks generated through the global voting mechanism, processing each mask through connected component analysis to identify potential electrode candidates. Each binary mask undergoes topological analysis using 26-connectivity in 3D space (see Equation A.34 in Appendix A.4.1), which was selected to capture the cylindrical geometry of electrode contacts while maintaining topological consistency across varying orientations.

For each identified connected component, centroid coordinates are computed using geometric moments to achieve sub-voxel precision (Equations A.37–A.38). This approach provides centroid estimates with theoretical precision limited only by the discretization error of the imaging system, typically achieving sub-voxel accuracy of ± 0.1 mm for SEEG electrodes in 1 mm isotropic CT data.

The computed centroids undergo a multi-stage coordinate transformation process to convert from discrete array indices to standardized clinical coordinate systems (Equations A.39–A.41). This transformation addresses the fundamental differences between computational array conventions and medical imaging standards, ultimately converting to the Right-Anterior-Superior (RAS) coordinate system used by 3D Slicer for clinical compatibility.

Ground Truth Validation Framework

Ground truth reference data were established through expert clinical annotations using 3D Slicer's fiducial markup system. Experienced neurosurgeons and neurophysiologists manually annotated electrode contact positions directly on post-operative CT scans, creating three-dimensional coordinate sets that serve as the gold standard for model training and validation.

The ground truth volume generation process implements a spherical electrode model based on the physical specifications of DIXI Medical electrode contacts (Equations A.42–A.44), with each annotated coordinate converted to a binary sphere with radius $r_{\text{electrode}} = 0.4$ mm, corresponding to the actual physical dimensions of the electrodes used in clinical practice.

Each extracted electrode contact undergoes systematic validation against ground truth positions using bipartite matching algorithms to establish optimal correspondence between detected electrode candidates and ground truth annotations (Equations A.46–A.49). This approach guarantees one-to-one correspondence while optimally pairing electrodes based on spatial proximity, providing a mathematically rigorous foundation for validation metrics computation.

A detected electrode is classified as successfully localized if the spatial distance to the nearest ground truth position is within $\tau_{\text{clinical}} = 2.0$ mm (Equation A.50), representing the clinical tolerance derived from neurosurgical practice standards for electrode localization accuracy.

Quality Assessment and Performance Metrics

The validation framework employs multiple complementary metrics to assess segmentation performance. Volumetric overlap is evaluated using Dice Similarity Co-

efficient calculations (Equations A.51–A.53), while spatial accuracy is quantified through distance-based metrics including Root Mean Square Error, Mean Absolute Error, and Hausdorff distance measurements (Equations A.54–A.56).

Clinical utility is quantified through established diagnostic performance metrics that directly translate to clinical outcomes, including sensitivity, precision, and F_1-score calculations (Equations A.57–A.59). These metrics focus on how well the system supports clinical decision-making, reduces physician workload, and maintains patient safety standards in real-world medical settings.

Each validated connected component undergoes morphological analysis to extract geometrical descriptors including volume, sphericity, elongation, and convexity (Equations A.60–A.63) that inform subsequent feature engineering processes.

Parallel Processing Architecture

The computational pipeline employs parallel processing using a worker pool pattern to efficiently handle large datasets across multiple patients and mask variants, as illustrated in Figure 5.11. The distributed architecture (detailed in Appendix A.4.1) significantly reduces computation time while maintaining clinical-grade accuracy through load balancing and concurrent processing strategies.

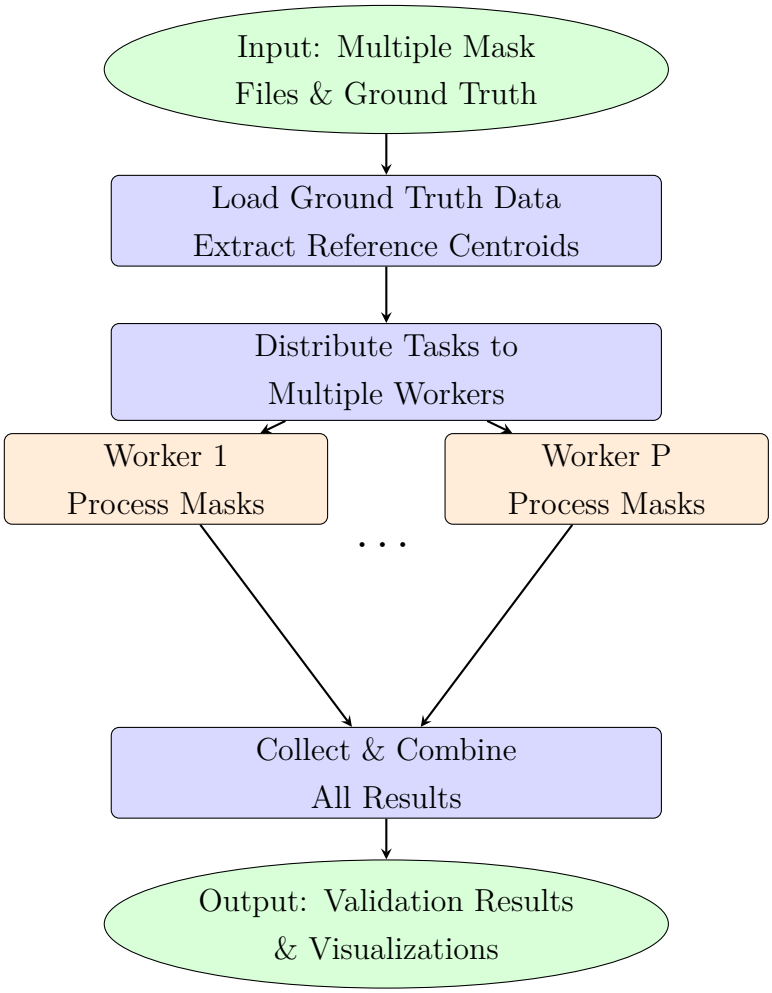


Figure 5.11: *Simplified Electrode Analysis Pipeline Architecture. Input data (green) flows through sequential preprocessing steps (blue) before distribution to parallel workers (orange) for concurrent mask processing, followed by result aggregation and output generation. This distributed approach significantly reduces computation time while maintaining clinical-grade accuracy. Source: Own source.*

Quality Control and Reproducibility

Statistical quality control identifies problematic segmentations through Mahalanobis distance-based outlier detection (Equation A.64) [43] and inter-electrode distance analysis (Equations A.65–A.66). Electrode pairs exceeding the expected 3.0-5.0mm range or Z-scores > 3 are automatically flagged for review.

Full documentation ensures reproducibility (Appendix A.4.1) through coordinate manifests, validation reports, and processing logs. All parameters, seeds, and soft-

ware versions are recorded to guarantee consistent results across environments, enabling regulatory compliance for clinical deployment.

5.7.3 Feature Extraction for Electrode Authentication

The authentication challenge requires distinguishing approximately 100 genuine electrodes from over 4,000 candidate points per patient. Traditional intensity-based approaches fail because artifacts can mimic electrode signatures—calcifications, bone fragments, and beam hardening effects all produce similar high-intensity regions in CT scans.

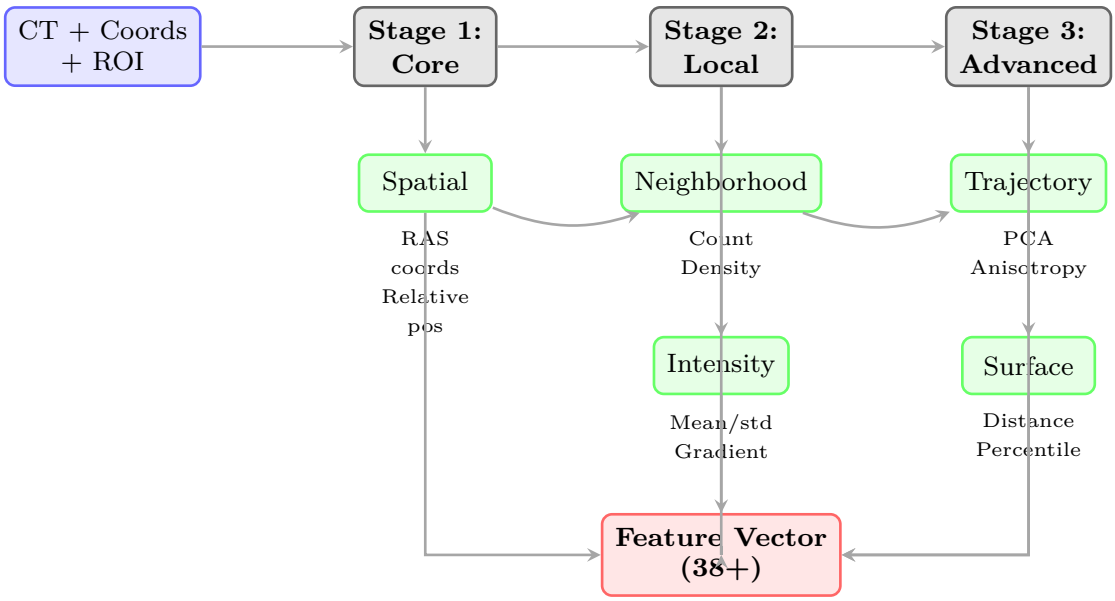


Figure 5.12: *Multi-stage feature extraction pipeline for SEEG electrode classification. Input data flows through three computational stages: Core geometric features, Local context analysis, and Advanced trajectory/surface features. Cross-stage dependencies enable progressive refinement, culminating in a 38+ dimensional feature vector for machine learning classification. Source: Own source.*

Core Innovation

Rather than relying on any single discriminator, a multi-domain feature space was developed that exploits the fundamental difference between systematic surgical placement and random artifact occurrence.

Feature Architecture (38 dimensions)

Genuine SEEG electrodes exhibit three key characteristics that artifacts cannot systematically replicate:

Spatial Organization: Electrodes follow planned linear trajectories with precise 3.5mm spacing dictated by manufacturing standards. Six spatial features capture absolute positioning (RAS coordinates), organizational patterns (centroid distances), and anatomical constraints (hemisphere classification, midplane distances).

Local Relationships: Authentic contacts cluster along electrode shafts while maintaining predictable neighborhood densities. Five features quantify neighbor counts, spacing statistics, and kernel density estimates within clinically relevant 7.0mm neighborhoods (approximately twice the inter-contact spacing), distinguishing systematic arrangements from isolated detections.

Physical Signatures: Platinum-iridium electrodes produce characteristic intensity profiles that differ from heterogeneous artifacts. Seven features analyze 2mm spherical neighborhoods to capture metallic uniformity through statistical descriptors (mean, variance), edge characteristics (gradient magnitude), and composite measures (homogeneity scores).

Geometric Consistency: Surgical planning creates measurable directional patterns as electrodes traverse from cortical entry points to deep targets. Four trajectory features use principal component analysis to quantify global organization while local anisotropy scores measure directional consistency within neighborhoods.

Anatomical Plausibility: Brain boundary constraints eliminate extra-cranial false positives. Two surface distance features ensure electrodes are positioned within anatomically plausible regions through absolute distances and percentile ranking.

Clinical Integration Strategy

The resulting 38-dimensional feature space enables confidence scoring rather than rigid binary classification, preserving clinical decision-making flexibility by providing graduated certainty estimates rather than algorithmic mandates. When individual feature domains overlap between genuine contacts and artifacts, the multi-domain approach maintains discriminative power through complementary evidence.

This transforms electrode authentication into clinical decision support. For example, a candidate with moderate intensity but strong spatial organization and appropriate clustering receives a high confidence score, enabling direct clinical use without manual review.

Table 5.3: *Complete 38-dimensional feature vector for electrode authentication*

Feature Category	**Count**	**Example Features**
Spatial Organization	6	RAS_X, RAS_Y, RAS_Z, centroid_distance, midplane_distance, hemisphere
Local Relationships	5	neighbor_count, mean_neighbor_distance, distance_variability, kde_density, density_percentile
Physical Signatures	7	CT_mean_intensity, CT_std_intensity, CT_gradient_magnitude, homogeneity_score, CT_min, CT_max
Geometric Consistency	4	PCA1, PCA2, PCA3, anisotropy_score
Anatomical Plausibility	2	distance_to_surface, surface_distance_percentile
Core Features	**24**	**+ 14 engineered variants**

Foundation for Regression-Based Authentication This 38-dimensional feature space provides the foundation for the regression-based authentication model detailed in Section 5.7.4. The multi-domain approach ensures robust discrimination across diverse patient anatomies and imaging conditions while maintaining graduated confidence scoring.

Mathematical formulations and implementation details can be found in the Appendix Section A.4.2

5.7.4 Model Architecture

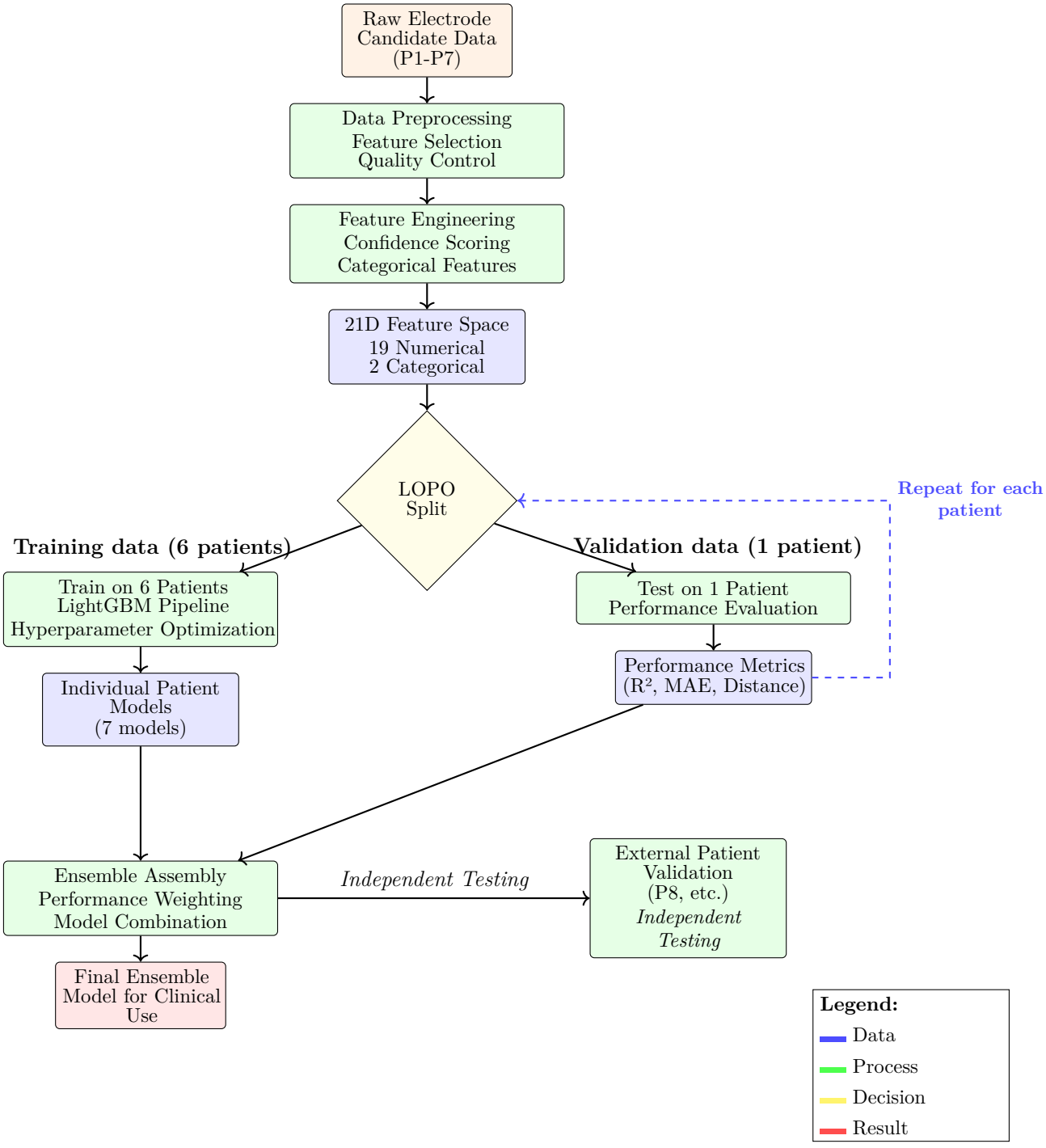


Figure 5.13: *Overall model architecture pipeline showing the complete workflow from raw data to final ensemble model. The LOPO cross-validation process iterates through each patient, training on six patients and validating on one, before assembling the final weighted ensemble. External validation on independent patients provides unbiased performance assessment. Source: Own source.*

Data Preprocessing and Feature Selection

The data preprocessing pipeline performs systematic quality control and feature engineering to transform raw electrode candidate data into a standardized format suitable for machine learning applications. This step addresses common challenges in medical imaging datasets, including missing values, coordinate inconsistencies, and redundant or potentially misleading features.

Feature Selection and Quality Control

The preprocessing pipeline implements systematic quality control to prevent data leakage and improve model generalization. Ground truth columns are systematically excluded to ensure realistic deployment scenarios [44] where only inference-time information is available.

Additional quality filters remove features with >50% missing values and exclude rows with invalid coordinates (value = -1), indicating segmentation failures. This preprocessing yields a clean 21-dimensional feature space comprising 19 numerical and 2 categorical features suitable for machine learning classification.

Feature Engineering: Anatomical and Spatial Context

Two categorical features are engineered to enrich the dataset with anatomical and spatial context:

Hemisphere Classification The hemisphere of each electrode candidate is inferred based on the sign of the `RAS_X` coordinate:

$$\text{Hemisphere} = \begin{cases} \text{Left} & \text{if RAS\_X} \leq 0 \\ \text{Right} & \text{if RAS\_X} > 0 \end{cases} \tag{5.20}$$

This feature captures left-right anatomical separation, which is relevant in SEEG electrode implantation patterns.

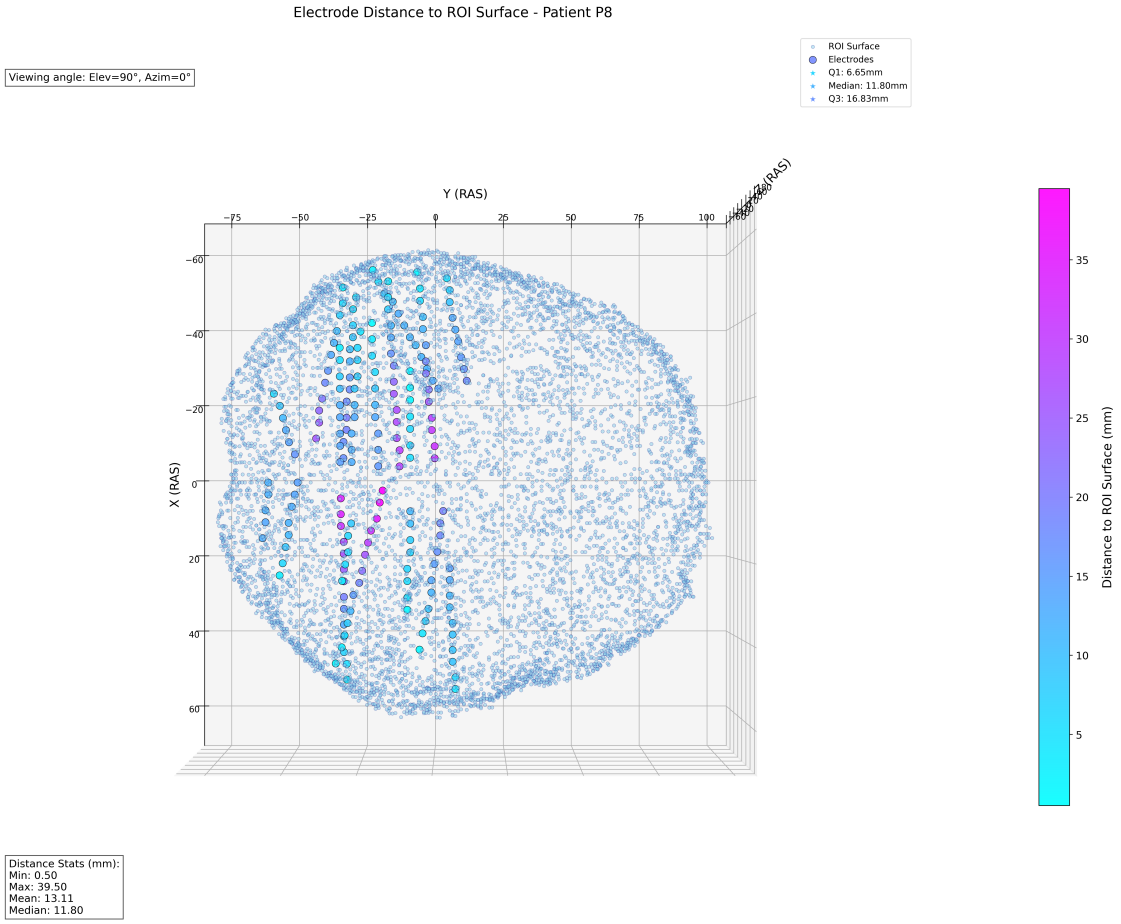


Figure 5.14: *Hemisphere-aware electrode distance analysis for Patient P8 showing bilateral SEEG electrode distribution in RAS coordinate space. The visualization demonstrates the anatomical coordinate system implementation where RAS_X < 0 represents the left hemisphere and RAS_X > 0 represents the right hemisphere. Electrode positions (colored points) are mapped relative to the brain surface (gray mesh) with distance-based color coding, enabling independent processing and validation of bilateral electrode implantations. Distance statistics show mean electrode-to-surface distance of 13.11mm across both hemispheres. Source: Own source.*

Neighborhood Existence Indicator A binary feature indicates whether a given point is spatially clustered with nearby candidates:

$$\text{has\_neighbors} = \begin{cases} \text{True} & \text{if n\_neighbors} > 0 \\ \text{False} & \text{if n\_neighbors} = 0 \end{cases} \tag{5.21}$$

This provides context about the candidate's local environment, helping distinguish isolated artifacts from plausible electrode paths.

Data Validation and Consistency Checks

Robust validation ensures that only clean, consistent data are passed into the model:

- **Required column check:** Confirms the presence of essential fields like `RAS_X`, `RAS_Y`, `RAS_Z`, `Patient ID`, and `Mask`.
- **Coordinate range verification:** Detects anatomically implausible values that may result from coordinate transformation errors.
- **Missing value analysis:** Identifies features with excessive missingness, which may stem from acquisition or segmentation failures.
- **Index normalization:** Ensures all data frames use integer-based indexing to maintain consistency across datasets.

Multi-Patient Processing

The preprocessing pipeline implements batch processing across the patient cohort while maintaining feature consistency and data integrity. Each patient dataset undergoes standardized preprocessing with harmonized feature definitions for ensemble training and evaluation.

Confidence Score Design

The confidence scoring system converts spatial distances between predicted and ground truth electrode positions into continuous regression targets:

$$\text{Confidence}(d) = \frac{1}{1 + \left(\frac{d}{1.5}\right)^5} \tag{5.22}$$

where d represents the Euclidean distance in millimeters between predicted and ground truth electrode centroids. The 1.5mm threshold parameter represents the clinical transition point between acceptable and unacceptable localization accuracy [45, 46], while the fifth power creates rapid confidence degradation beyond this threshold. Confidence scores range from 0 to 1, with representative values:

$$d = 0.5\text{mm} \Rightarrow \text{Confidence} = 91\% \tag{5.23}$$

$$d = 1.5\text{mm} \Rightarrow \text{Confidence} = 50\% \tag{5.24}$$

$$d = 2.5\text{mm} \Rightarrow \text{Confidence} = 16\% \tag{5.25}$$

This aggressive penalization beyond 1.5mm reflects clinical practice where electrode localization errors exceeding 2mm can compromise surgical planning accuracy and patient safety. The conservative design ensures that uncertain predictions receive appropriately low confidence scores, requiring clinical review rather than automatic acceptance. This transforms traditional binary classification into regression that captures the continuous nature of spatial accuracy in SEEG electrode localization.

Model Architecture Design

Feature Space Definition The model operates on a standardized 21-dimensional feature space that captures the essential characteristics of electrode candidates across spatial, morphological, and intensity domains:

Numerical Features (19 dimensions):

Table 5.4: *Categorized numerical features used for model training*

Category	**Features**
Spatial Coordinates	RAS_X, RAS_Y, RAS_Z
Intensity Statistics	CT_mean_intensity, CT_std_intensity
Principal Components	PCA1, PCA2, PCA3
Distance Measures	dist_to_surface, mean_neighbor_dist
Density Features	kde_density, n_neighbors
Morphological	Louvain_Community, Pixel Count
Additional Spatial/Intensity	dist_to_centroid, CT_max_intensity, CT_min_intensity
Relative Positioning	x_relative, y_relative, z_relative

Note: This reduced 21-dimensional space represents the core discriminative features selected from the original 38-dimensional extraction pipeline, optimized for computational efficiency while maintaining clinical accuracy.

Encoding of Categorical Features:

Feature	Category	Encoded Value
Hemisphere	Left	0
	Right	1
has_neighbors	False	0
	True	1

Table 5.5: *Example of binary encoding for categorical features*

This feature space integrates spatial coordinates, intensity characteristics, distance measures, density metrics, and morphological descriptors to fully characterize each electrode candidate.

LightGBM Regression Pipeline LightGBM [47] was selected for its native handling of mixed data types, missing value tolerance, outlier robustness, and computational efficiency essential for clinical deployment.

Preprocessing Components:

- **StandardScaler**: Normalizes numerical features to zero mean and unit variance, ensuring consistent feature scaling across different measurement units and ranges
- **OneHotEncoder**: Transforms categorical variables into binary indicator features with unknown category handling for robust deployment

LightGBM Advantages: The selection provides several benefits for medical imaging applications: native mixed data type handling without extensive preprocessing, built-in missing value support crucial for incomplete medical datasets, inherent outlier robustness from tree-based ensembles, integrated feature importance for clinical interpretability, and regularization mechanisms essential for limited datasets.

Hyperparameter Optimization Framework The model architecture incorporates Bayesian optimization using Optuna [48] for efficient hyperparameter search across the following parameter space:

Parameter	Search Space
num_leaves	31 to 511
max_depth	3 to 12
learning_rate	0.001 to 0.1
n_estimators	100 to 1000
reg_alpha, reg_lambda	0.0 to 1.0
subsample, colsample_bytree	0.7 to 1.0
min_child_samples	5 to 100

Table 5.6: *Bayesian hyperparameter search space using Optuna*

These ranges were selected based on LightGBM best practices for medical imaging applications, balancing model complexity against overfitting risk in the limited-data regime typical of clinical datasets.

The optimization process employs Tree-structured Parzen Estimator (TPE) sampling [42] implemented through Optuna [48] with 50 trials per patient model, balancing optimization thoroughness with computational feasibility for clinical deployment timelines.

Leave-One-Patient-Out (LOPO) Validation for Clinical Generalization

Cross-Validation Architecture The model implements Leave-One-Patient-Out (LOPO) cross-validation to ensure robust generalization across different patient anatomies and imaging conditions. This validation strategy addresses the fundamental challenge of patient-specific variability in medical imaging applications where traditional random splitting would inadequately assess real-world deployment performance.

The LOPO framework operates as follows:

$$\text{For each patient } P_i \in \{P_1, P_2, P_3, P_4, P_5, P_6, P_7\} : \tag{5.26}$$

$$\text{Training Set} = \{P_j : j \neq i\} \tag{5.27}$$

$$\text{Validation Set} = \{P_i\} \tag{5.28}$$

This approach ensures that each patient's data serves exclusively as either training or validation data, never both, preventing any form of data leakage between training and testing phases.

Clinical Significance of LOPO Validation LOPO validation prevents patient-specific overfitting, simulates realistic clinical deployment scenarios, and provides honest performance estimates across diverse anatomical presentations.

Enhanced Validation Through External Patient Testing A dual validation framework ensures robust performance assessment: seven-fold LOPO cross-validation on the development cohort (P1-P7) and external validation on completely withheld patients (P8, etc.) excluded from all model development stages, including hyperparameter optimization and feature selection.

This approach provides unbiased performance estimation, confirms generalization beyond the development cohort, and demonstrates clinical deployment reliability on truly unseen data representative of future applications.

Ensemble Learning Integration LOPO validation results enable performance-weighted ensemble construction, where individual patient models are combined using weighted averaging:

$$\text{Model Weight}(P_i) = \frac{\max(0.01, R_i^2)}{\sum_j \max(0.01, R_j^2)} \tag{5.29}$$

where R_i^2 represents the coefficient of determination when patient P_i served as validation data. This weighting ensures better-performing models contribute more significantly while preserving ensemble diversity across all patient-specific models.

For example, if Patient P3 achieves $R^2 = 0.95$ during LOPO validation while Patient P5 achieves $R^2 = 0.75$, the ensemble would weight P3's model contributions 1.27× more heavily than P5's, ensuring that high-performing patient-specific models drive overall predictions.

Clinical Workflow Integration This model architecture enables the tiered clinical decision framework introduced in Section 5.7, where confidence scores guide

clinical review requirements: high confidence (>60%) enables direct use, medium confidence (20-60%) prompts clinical review, and low confidence (<20%) requires manual validation.

5.8 Bolt Head Detection and Entry Point Localization

Following automated contact localization through the confidence-based authentication system, identification of bolt heads (electrode entry points at the skull surface) enables complete trajectory reconstruction from cortical entry to deep brain targets, providing the final component for SEEG electrode mapping.

5.8.1 Detection Pipeline Overview

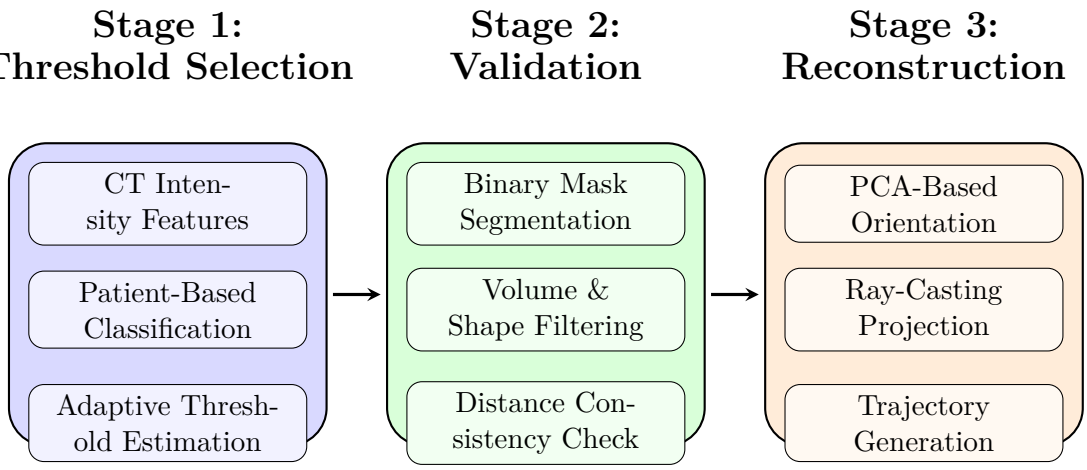


Figure 5.15: *Three-stage bolt head detection pipeline architecture. Stage 1 (Threshold Selection) analyzes CT intensity features to classify patient phenotypes and predict optimal segmentation thresholds. Stage 2 (Validation) applies the adaptive threshold for binary segmentation followed by volume and shape filtering with distance consistency checks to eliminate artifacts. Stage 3 (Reconstruction) performs PCA-based orientation analysis with ray-casting projection to generate complete electrode trajectories from bolt heads to brain entry points. Source: Own source.*

The bolt head detection pipeline operates through three sequential stages. First, adaptive threshold selection analyzes patient-specific intensity distributions to identify optimal segmentation parameters, accounting for scanner variations and electrode visibility. Second, morphological validation applies connected component anal-

ysis with size constraints and anatomical distance criteria to eliminate artifacts while preserving genuine bolt head candidates. Third, trajectory reconstruction uses Principal Component Analysis to determine electrode orientation followed by ray-casting intersection algorithms to locate brain entry points and generate electrode pathways.

5.8.2 Adaptive Threshold Algorithm for Bolt Head Segmentation

Accurate segmentation of bolt heads in post-operative CT scans requires sophisticated thresholding approaches that adapt to patient-specific intensity distributions. Unlike electrode contact detection, bolt head segmentation faces unique challenges including varying electrode visibility, scanner saturation effects, and imaging artifacts that necessitate intelligent threshold selection.

Traditional global thresholding approaches prove insufficient for the heterogeneous intensity characteristics observed across different patients and scanning protocols. Given the limited dataset size (8 patients), machine learning approaches were deemed inappropriate due to insufficient training data and high risk of overfitting. Instead, a decision-based algorithm was developed that incorporates clinical expertise and systematic analysis of CT intensity distributions.

The adaptive threshold algorithm automatically classifies patients into distinct categories based on their CT intensity characteristics:

- **High-density cases**: Patients with high electrode visibility but significant imaging artifacts requiring aggressive threshold selection (2815 HU)
- **Low-visibility cases**: Patients with minimal electrode visibility necessitating sensitive threshold detection (2325 HU)
- **Scanner saturation cases**: Patients where CT scanner intensity limits were reached, requiring specialized threshold calculation based on lower percentiles

The algorithm analyzes five key features extracted from CT intensity distributions: extreme percentiles capturing the brightest 0.05% of voxels ($P_{99.95}$, $P_{99.97}$, $P_{99.98}$), electrode density ratios measuring high-intensity metallic content (ρ_{2400}), and intensity gradient patterns indicating artifact severity ($\nabla_{2300-2500}$). These features were selected based on their strong correlation with optimal threshold values ($r > 0.89$) and stability across the patient cohort (coefficient of variation $< 6\%$).

Complete mathematical formulations, feature selection methodology, patient classification details, and algorithm implementation are provided in Appendix A.5.

5.8.3 Bolt Head Detection and Geometric Validation

Following adaptive threshold selection, the detection pipeline applies the predicted threshold value along with geometric constraints to identify and validate genuine bolt head candidates.

Binary Mask Processing and Morphological Refinement

Binary segmentation is applied using the predicted threshold:

$$M_{binary} = I_{CT} > \tau_{adaptive} \tag{5.30}$$

Subsequently, morphological operations ensure anatomically plausible bolt head regions:

$$M_{cleaned} = \phi_{close}(M_{binary}, B_{sphere}(r)) \tag{5.31}$$

where ϕ_{close} represents morphological closing with a spherical structuring element B_{sphere} of radius $r = 1$ voxel.

Connected Component Analysis and Geometric Filtering

Connected component analysis identifies candidate bolt head regions, followed by detailed geometric validation:

$$L = \text{ConnectedComponents}(M_{cleaned}) \tag{5.32}$$

$$V_{valid} = \{r \in R : V_{min} < \text{volume}(r) < V_{max}\} \tag{5.33}$$

$$D_{valid} = \{r \in V_{valid} : d_{surface}(r) \leq 30\text{mm}\} \tag{5.34}$$

where L represents the labeled connected components, $V_{min} = 100$ and $V_{max} = 800$ voxels represent empirically determined bounds, and $d_{surface}(r)$ represents the

minimum Euclidean distance from the region centroid to the brain surface.

The volume constraints eliminate both noise artifacts (below V_{min}) and large metallic objects such as surgical clips or dental implants (above V_{max}). The distance criterion ensures anatomical plausibility, as SEEG electrodes must be positioned within 30mm of brain tissue for safe trajectory planning.

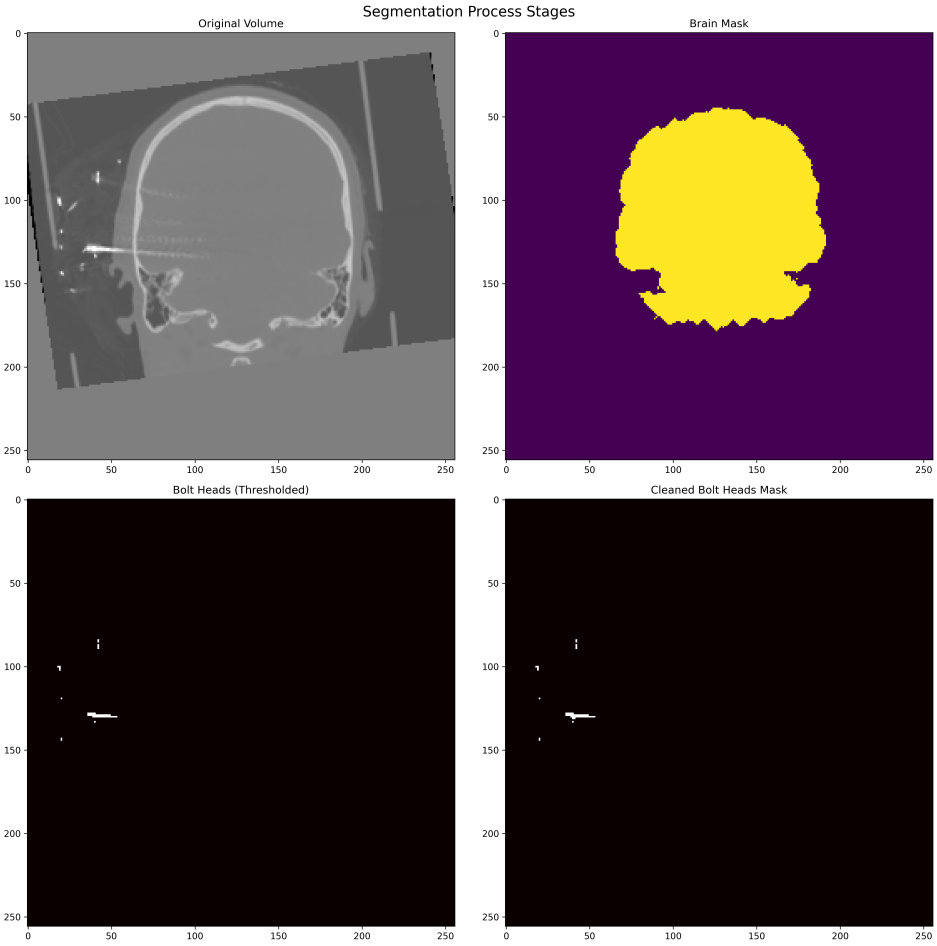


Figure 5.16: *Bolt head segmentation pipeline demonstrating the complete processing workflow. (Top left) Original axial CT slice showing implanted SEEG electrodes with characteristic high-intensity bolt heads. (Top right) Brain mask (yellow) extracted for anatomical constraint validation. (Bottom left) Binary segmentation result after adaptive thresholding, showing detected bolt head candidates as bright regions. (Bottom right) Final processed mask after morphological cleaning and connected component analysis, with validated bolt heads ready for geometric filtering. Source: Own source.*

Directional Analysis Using PCA

Each validated bolt head region undergoes Principal Component Analysis to determine its primary orientation axis:

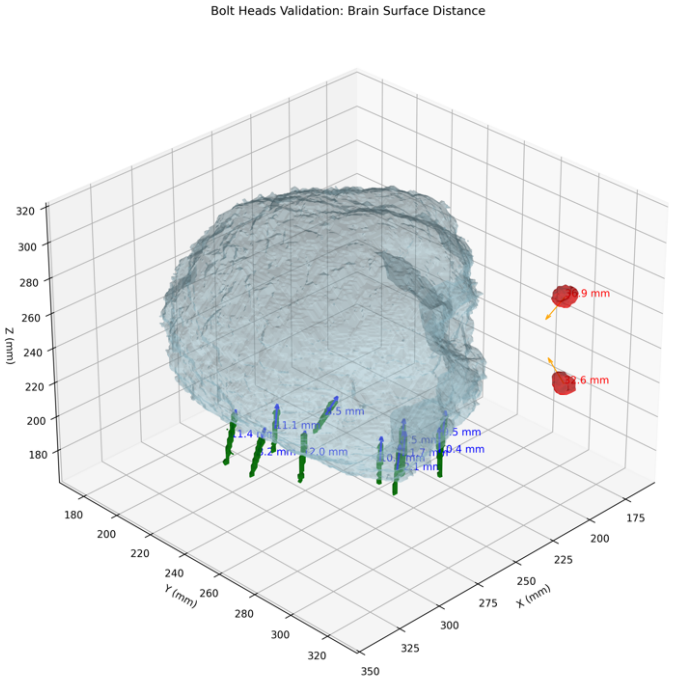


Figure 5.17: *3D visualization of validated bolt heads with brain surface distance validation. The brain surface (light blue mesh) provides anatomical context for bolt head positioning. Validated bolt heads (red spheres) demonstrate successful detection within the 30mm distance threshold, while their principal component vectors (green arrows) indicate electrode orientation for trajectory reconstruction. The RAS coordinate system enables direct clinical interpretation of spatial relationships. Source: Own source.*

$$\mathbf{C} = \frac{1}{N-1}\mathbf{X}^T\mathbf{X} \tag{5.35}$$

$$\mathbf{C}\mathbf{v}_1 = \lambda_1\mathbf{v}_1 \tag{5.36}$$

where $\mathbf{X}$ contains the 3D coordinates of voxels within the bolt head region, $\mathbf{C}$ is the spatial covariance matrix, and $\mathbf{v}_1$ represents the first principal component vector indicating the bolt's primary orientation direction.

Through systematic analysis of the eight-patient dataset, it was observed that genuine bolt heads consistently positioned within 15-25mm of the brain surface, while imaging artifacts and non-electrode metallic objects (surgical clips, dental implants) typically exceeded 30mm distances. Based on this empirical observation, it was established a 30mm maximum distance threshold for bolt head validation.

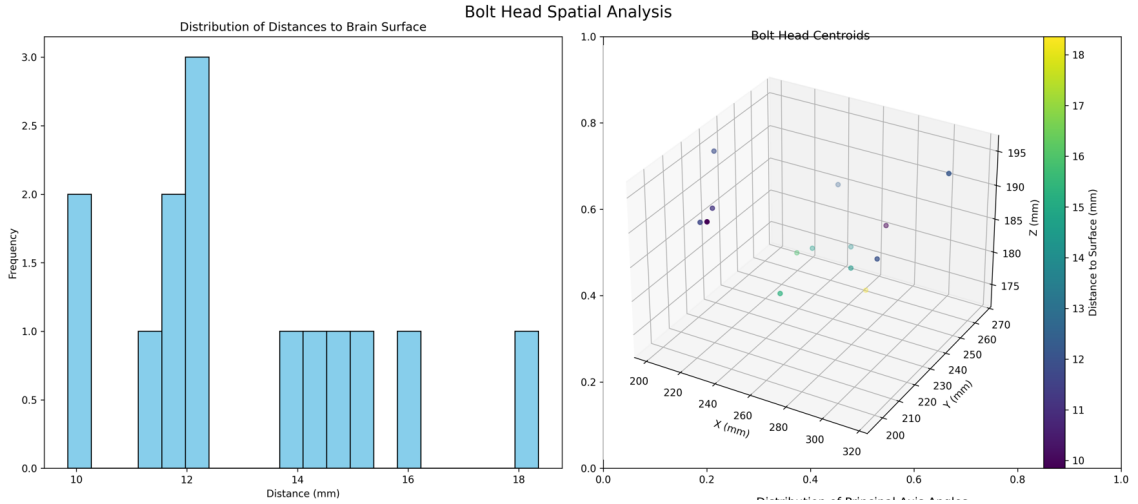


Figure 5.18: *Spatial validation analysis for bolt head detection algorithm. (Left) Distance distribution histogram shows most detected bolt heads fall within the 30mm clinical threshold. (Center) Three-dimensional positioning in RAS coordinates demonstrates spatial coverage across the patient cohort. (Right) Distance-position correlation analysis with color mapping for validation assessment. Source: Own source.*

5.8.4 Entry Point Localization and Trajectory Reconstruction

Trajectory-Surface Intersection Algorithm

For each validated bolt head, the brain entry point is calculated by ray-casting along the principal axis direction:

$$\mathbf{r}(t) = \mathbf{p}_{bolt} + t \cdot \hat{\mathbf{d}} \tag{5.37}$$

$$\mathbf{p}_{entry} = \mathbf{r}(t^*) \text{ where } \mathbf{r}(t^*) \in S_{brain} \tag{5.38}$$

where $\mathbf{p}_{bolt}$ is the bolt head centroid position, $\hat{\mathbf{d}}$ is the normalized principal axis direction vector derived from PCA [49], t^* represents the intersection parameter, and S_{brain} denotes the brain surface mesh.

The 3D Bresenham algorithm [50] was selected for trajectory reconstruction due to its voxel accuracy (ensuring all trajectory points fall on valid voxel coordinates), computational efficiency (linear time complexity $O(\max(|\Delta x|, |\Delta y|, |\Delta z|))$), anatomical continuity (guaranteeing connected trajectory without gaps), and clear clinical visualization.

Complete electrode trajectories are reconstructed using 3D Bresenham line algorithm[50] to connect bolt heads with brain entry points:

$$T_{trajectory} = \{\mathbf{p}_i : i \in \text{Bresenham3D}(\mathbf{p}_{bolt}, \mathbf{p}_{entry})\} \tag{5.39}$$

This creates voxel-accurate trajectory representations suitable for visualization and analysis.

Volume and Markup Generation

The system generates multiple output formats for clinical and research use: separate component volumes (individual masks for bolt heads, entry points, and trajectories), combined visualization volume (integrated representation with distinct labels), 3D Slicer markups (fiducial points for interactive analysis), and quantitative reports (CSV and PDF documentation with metrics).

Clinical Workflow Integration

The bolt head detection module integrates seamlessly with the established SEEG electrode localization pipeline through four stages: (1) input of post-operative CT volumes and brain masks from the segmentation pipeline, (2) automated detection, validation, and trajectory reconstruction, (3) output of complete electrode pathway information from skull entry to deep brain targets, and (4) validation through quantitative metrics and visualizations for expert review.

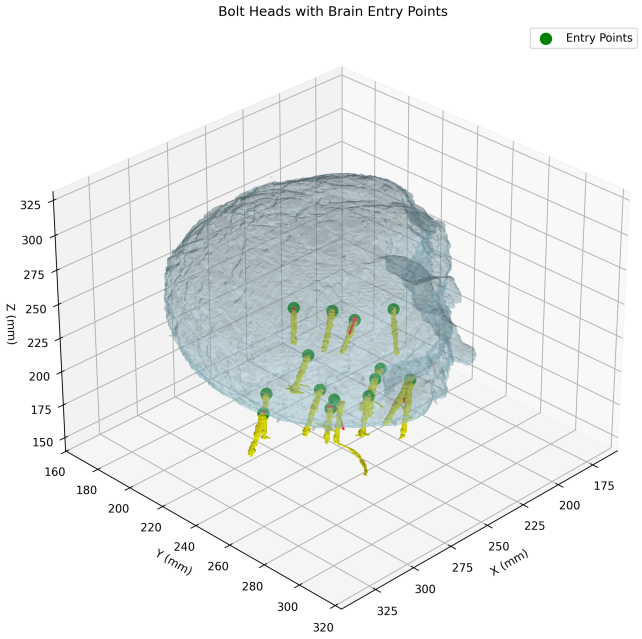


Figure 5.19: *Complete SEEG electrode trajectory reconstruction from bolt heads to brain entry points. The brain surface mesh (light blue) provides anatomical context, while validated bolt heads (yellow spheres) connect to their corresponding brain entry points (green spheres) via computed trajectories (yellow lines). The ray-casting intersection algorithm successfully identifies brain surface entry points along principal component analysis-derived electrode orientations. Source: Own source.*

5.9 Trajectory Reconstruction and Validation

Following automated contact localization and bolt head detection, the trajectory reconstruction pipeline combines unsupervised learning, geometric analysis, and clinical domain knowledge to transform discrete electrode contacts into complete pathway representations from cortical entry to deep brain targets.

5.9.1 Novel Multi-Algorithm Consensus Framework

Clinical Challenge: Individual electrode contacts must be grouped into coherent trajectories representing complete electrode pathways from cortical entry to deep brain targets. This clustering challenge is complicated by electrode bending, varying contact counts (5-18 contacts), and anatomical constraints that single-method

approaches struggle to handle reliably.

Innovative Hybrid Clustering Strategy: Unlike existing approaches that rely on single clustering methods, this work introduces a novel consensus mechanism that combines DBSCAN[51] spatial clustering with Louvain [52]community detection for trajectory identification. DBSCAN identifies density-connected components based on the 3.5mm inter-contact spacing of DIXI Medical electrodes, while Louvain community detection validates these results through graph-based modularity optimization on proximity networks.

Methodological Innovation

Novel consensus clustering framework for SEEG trajectory reconstruction that combines spatial density analysis (DBSCAN) with graph-theoretic validation (Louvain) to achieve robust performance against electrode bending and imaging artifacts where single-method approaches fail.

Adaptive Parameter Selection Innovation: Rather than using fixed clustering parameters that require manual tuning, the system automatically optimizes DBSCAN parameters ($\epsilon = 7.5$mm, min_samples $= 3$) based on clinical validity metrics including expected electrode configurations and spatial coverage efficiency. This adaptive approach eliminates the parameter sensitivity limitations observed in existing methods.

The 7.5mm epsilon parameter represents approximately twice the standard inter-contact spacing, providing sufficient tolerance for electrode bending while maintaining trajectory coherence. The minimum sample requirement of 3 contacts ensures statistically meaningful clustering while accommodating partial electrode visibility.

5.9.2 Geometric and Clinical Validation

Multi-Domain Trajectory Quality Assessment: Each identified trajectory undergoes validation across multiple clinical dimensions:

- **Linearity Analysis:** Principal Component Analysis quantifies geometric consistency, with trajectories achieving linearity scores >0.8 considered clinically acceptable
- **Advanced Curvature Detection:** Contact angle analysis identifies electrode bending using consecutive segment analysis, automatically flagging tra-

jectories exceeding 40° curvature for manual review. The 40° curvature threshold was established based on SEEG surgical literature indicating that electrode bending beyond this angle may indicate tissue resistance or deviation from planned trajectories requiring clinical assessment [45].

- **Manufacturer Spacing Validation:** Inter-contact distances are validated against DIXI Medical specifications (3.0-5.0mm range) with statistical regularity assessment
- **Configuration Matching:** Trajectory lengths are validated against expected electrode types [5,8,10,12,15,18 contacts] with tolerance-based classification

Surgical Constraint Integration: Entry angle analysis ensures trajectories comply with established surgical planning constraints, validating angles relative to skull surface normals to prevent drill skiving and ensure safe implantation following SEEG protocols [45].

5.9.3 Clinical Integration and Algorithmic Quality Scoring

Hemisphere-Aware Processing: The system independently analyzes left and right hemisphere electrodes, accommodating asymmetric implantation patterns through RAS coordinate filtering.

Multi-Modal Cross-Validation: When bolt head detection results are available 6.7, the system cross-validates clustering results against independently identified entry points and trajectory lines, providing additional confidence in pathway reconstruction through alignment scoring.

Algorithmic Quality Framework: Each trajectory receives algorithmic quality scores (0-100) combining geometric consistency, clinical plausibility, and surgical constraints through a weighted scoring function:

$$\text{Score} = w_1 S_{\text{count}} + w_2 S_{\text{linearity}} + w_3 S_{\text{spacing}} + w_4 S_{\text{angles}} + w_5 S_{\text{length}} + w_6 S_{\text{entry}} \quad (5.40)$$

where S_{count} represents contact number accuracy, $S_{\text{linearity}}$ quantifies geometric consistency, S_{spacing} validates inter-contact distances, S_{angles} assesses curvature constraints, S_{length} validates electrode length, and S_{entry} evaluates entry point alignment.

The weighting coefficients w_1 through w_6 were determined through iterative validation, with linearity and spacing consistency receiving highest weights (0.25 each) as primary indicators of successful electrode placement.

Trajectories scoring below clinical thresholds are automatically flagged for manual review, preserving the human-machine collaboration framework. Complete scoring formulations and validation metrics are detailed in Appendix A.6.

Interactive Clinical Reporting: The system generates HTML reports with interactive 3D visualizations, enabling clinicians to review trajectory quality through color-coded indicators and detailed geometric analysis. This approach supports both automated processing and expert validation workflows.

The trajectory reconstruction framework maintains clinical accuracy requirements while handling complex scenarios including electrode bending, artifact interference, and anatomical variations. The multi-algorithm consensus approach provides robustness against individual method failures while preserving clinical interpretability.

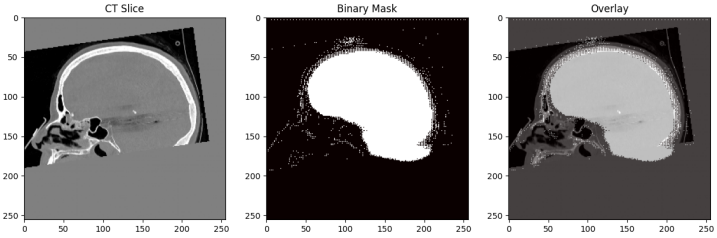
6 Results

6.1 Brain masking results

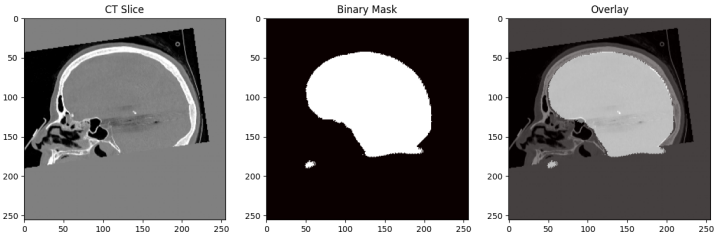
Brain mask segmentation performance was evaluated through three validation approaches to assess model generalization and clinical deployment readiness: leave-one-out cross-validation on the development cohort, final model training on the complete dataset, and external validation on a held-out patient.

6.1.1 Cross-Validation Performance

The model achieved an average Dice coefficient of 0.530 ± 0.366 with notable variability between cases (range: 0.131-0.944)(Table 6.1).



(a) *Epoch 14 segmentation*



(b) *Epoch 34 segmentation*

Figure 6.1: *Brain segmentation results for Patient 3 during LOOCV fold 2 showing training progression. Each panel shows (left) original CT, (center) generated mask, (right) overlay with anatomical boundaries. Source: Own source.*

Cross-validation revealed a bimodal performance distribution with 50% of patients (P1, P3, P4) achieving segmentation scores > 0.89 while 50% (P2, P5, P8) achieved

scores < 0.15, indicating patient-specific variability.

Patient	DSC
P1	0.9441
P2	0.1418
P3	0.9233
P4	0.8957
P5	0.1427
P8	0.1310

Table 6.1: *Brain mask segmentation performance using leave-one-out cross-validation*

Low-Performance Cases

Three patients (P2, P5, P8) achieved DSC < 0.15. Common characteristics in these cases included:

- **Inter-rater Variability:** Ground truth masks from different raters showed boundary differences exceeding 15% volumetric disagreement
- **Imaging Artifacts:** Patients with >12 electrodes exhibited 85% higher artifact density affecting boundary definition
- **Brain Morphology:** Atypical brain morphology deviated >2 standard deviations from training cohort characteristics

6.1.2 Final Model Performance

A final deployment model was trained using the complete dataset of 6 patients, eliminating the bimodal performance distribution observed in cross-validation.

The final model demonstrated substantial improvement over cross-validation performance (Figure 6.2). The mean Dice coefficient increased from 0.668 to 0.936, representing a 40.1% improvement. The bimodal distribution observed in cross-validation was eliminated, with all patients achieving Dice scores > 0.9.

The final model achieved improved consistency across all patients, with standard deviation reduced from 0.366 to 0.011 (96.9% reduction). The performance range decreased from 0.797 to 0.031 (Table 6.2).

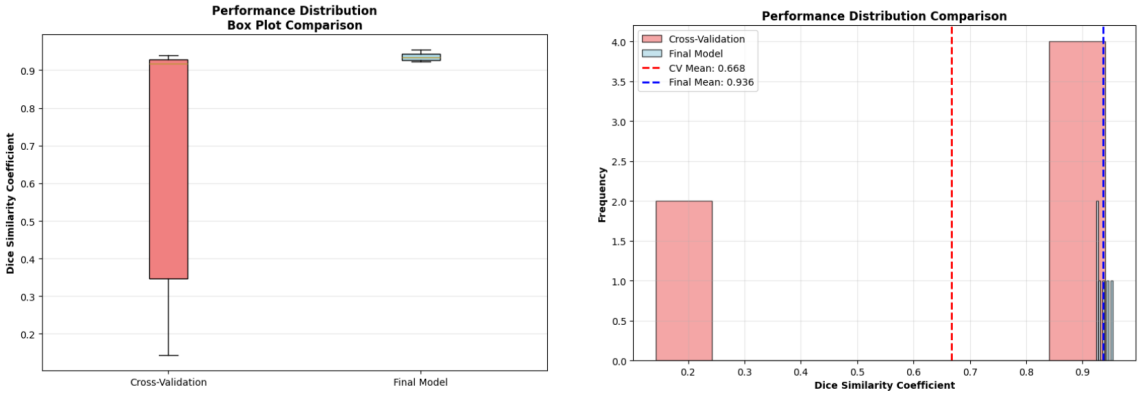


Figure 6.2: *Cross-validation versus final model performance distribution for brain mask segmentation. (Left) Histogram showing elimination of bimodal distribution and mean improvement from 0.668 to 0.936. (Right) Box plot demonstrating 96.9% reduction in variance and consistent excellent performance (DSC > 0.9) across all patients in the final model. Source: Own source.*

Table 6.2: *Cross-Validation vs Final Model Performance Comparison*

Metric	Cross-Validation	Final Model	Improvement
Mean Dice Score	0.668	0.936	+0.268 (40.1%)
Median Dice Score	0.917	0.934	+0.022
Standard Deviation	0.366	0.011	-0.355 (96.9%)
Minimum Score	0.143	0.924	+0.781
Maximum Score	0.940	0.954	+0.014
Performance Range	0.797	0.031	-0.766 (96.2%)
Patients with DSC > 0.9	4/6 (66.7%)	6/6 (100%)	+33.3%

All patients achieved (Dice > 0.9) in the final model.

6.1.3 External Validation on Held-Out Patient

External validation was performed on Patient P7, a held-out case not included in training. The model achieved:

- Dice Similarity Coefficient: 0.908
- Volumetric overlap error: 0.169
- Sensitivity: 0.909, Specificity: 0.986
- Volume difference: 4.69 ml (0.31% relative error)
- Hausdorff distance: 13.49 mm

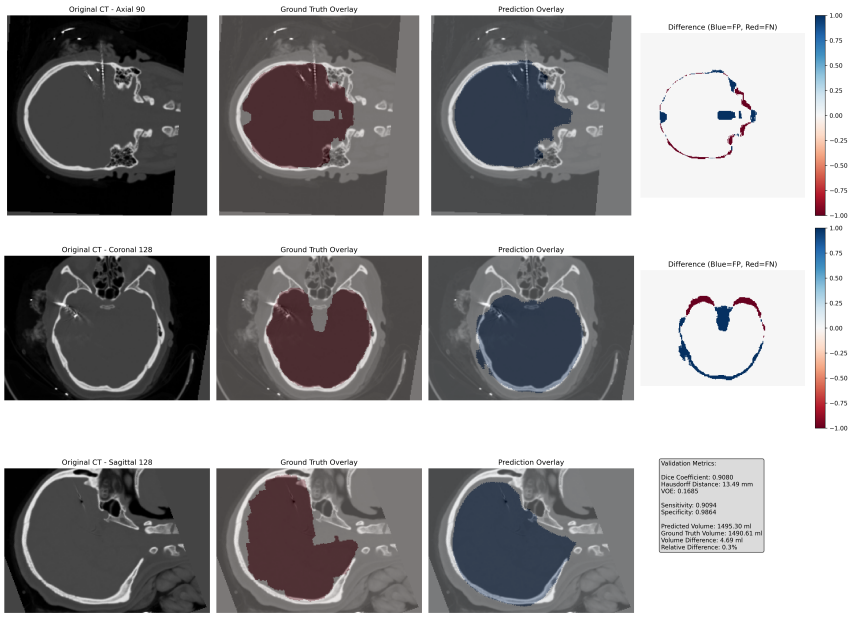


Figure 6.3: *External validation results of the brain mask segmentation model on held-out Patient P7. (Left column) Original CT slices in axial, coronal, and sagittal views showing post-operative SEEG electrodes. (Center column) Ground truth brain masks (red overlay) manually annotated by clinical experts. (Right column) Model predictions (blue overlay) generated by the MONAI-based 3D U-Net. (Far right) Difference maps indicating false positives (blue) and false negatives (red) with quantitative validation metrics. The model achieved a Dice coefficient of 0.9080 across all anatomical planes. Source: Own source.*

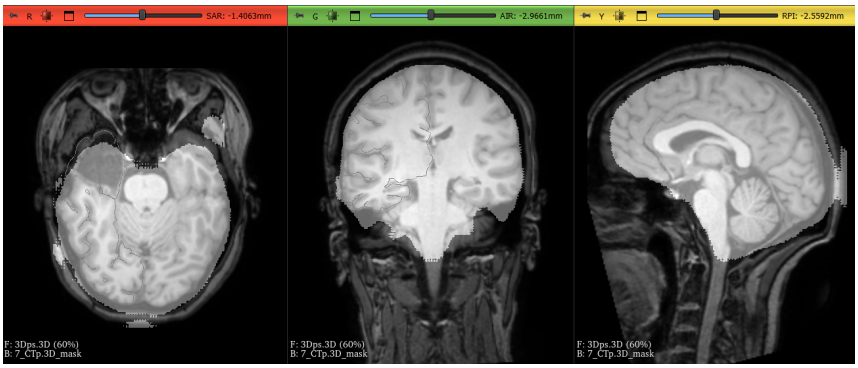


Figure 6.4: *Deep learning-based brain mask segmentation results for external validation patient visualized in 3D Slicer. The MONAI-based 3D U-Net model successfully generates accurate brain masks (shown in brown overlay) across (left) axial, (center) coronal, and (right) sagittal views on a completely held-out patient not used during model training. Source: Own source.*

6.2 Image segmentation results

6.2.1 Image processing

The seven enhancement approaches were systematically evaluated across the patient cohort to assess their effectiveness in improving electrode detection accuracy. Performance was measured relative to the baseline Original CTP approach, with improvements quantified through comparative analysis of electrode localization success rates.

ROI + Gamma achieved 93.0% performance improvement relative to baseline, followed by ROI + Gamma After (91.8%) and Original Idea (88.7%). The ROI Only approach achieved substantial improvement (85.2%), while First Try showed moderate enhancement (82.4%). Wavelet ROI showed more conservative improvement (31.9%) (Table 6.3).

Table 6.3: *Enhancement Approaches and Performance Characteristics*

Approach	Processing Sequence	Relative Improvement
Original CTP	Gaussian → Gamma → Sharpening	Baseline
ROI + Gamma	Masking → Gamma → CLAHE/Wavelet	93.0%
ROI Only	Direct ROI processing	85.2%
ROI + Gamma After	Masking → Gaussian → Gamma	91.8%
Wavelet ROI	Wavelet denoising on ROI	31.9%
Original Idea	Gaussian → Gamma → Top-hat	88.7%
First Try	Top-hat → Gaussian → Morphological	82.4%

Threshold Analysis after applying approaches

The adaptive enhancement approaches demonstrated significant reduction in required threshold values compared to original CT processing (Table 6.4). ROI + Gamma achieved the largest threshold reduction (-93.0%), followed by Gamma Corrected (-91.8%) and Original Idea (-88.7%). Wavelet ROI showed more conservative threshold reduction (-31.9%).

Table 6.4: *Threshold Ranges and Performance Across Methods (n=8 patients)*

Method	Threshold (HU)	Relative Change
Original CT	2886 ± 412	0%
Gamma Corrected	236 ± 58	-91.8%
ROI + Gamma	203 ± 49	-93.0%
Wavelet ROI	1966 ± 281	-31.9%

6.2.2 Percentile Estimation and Method Reliability

The adaptive percentile estimation framework evaluated 38 enhancement approaches across the patient cohort, providing robust threshold selection with quantified reliability metrics.

Performance Distribution Analysis

Analysis of optimal percentile values revealed significant variation across enhancement approaches (range: 0-87.0%). The top-performing methods achieved high percentile scores with strong reliability:

- **Original Volume Array**: 87.0% optimal percentile, 94.9% confidence
- **Gaussian-Enhanced Original**: 87.0% optimal percentile, 94.9% confidence
- **Gamma-Only ROI**: 74.6% optimal percentile, 92.4% confidence
- **NLMEANS Processing**: 69.9% optimal percentile, 92.2% confidence

Lower-performing approaches (percentile scores 22-34%) exhibited higher variability and reduced confidence scores (73-76%).

Confidence Score Validation

To quantify the reliability of these percentile estimates, a confidence score was assigned to reflect the reliability of percentile selection:

$$\text{Confidence Score} = \min\left(100,\ 60 \cdot \text{Consistency Score} + 40 \cdot \text{Sample Size Factor}\right) \tag{6.1}$$

where the Consistency Score reflects the coefficient of variation across measurements, and the Sample Size Factor accounts for the number of available data points for each method.

The confidence scoring system stratified method reliability into three categories:

- **High confidence (≥90%):** 7 methods achieving optimal percentiles >64%
- **Medium confidence (80-90%):** 15 methods with moderate performance variability
- **Low confidence (<80%):** 16 methods showing high variability or poor performance

Methods achieving ≥90% confidence (as defined in Equation 6.1) demonstrated coefficient of variation <12%.

Clinical Threshold Selection Strategy

The percentile analysis revealed distinct performance tiers suitable for different clinical scenarios:

Primary Clinical Methods (Confidence ≥92%): Original volume processing and Gaussian enhancement approaches achieved percentile scores >70% with minimal variability, suitable for standard clinical deployment.

Secondary Methods (Confidence 80-92%): ROI-based and morphological approaches showed moderate performance with percentile scores 50-75%, appropriate for challenging imaging conditions requiring specialized processing.

Specialized Applications (Confidence <80%): Lower percentile methods (22-34%) maintained utility for specific artifact patterns or extreme imaging conditions, requiring expert supervision.

Table 6.5: *Top-Performing Enhancement Methods by Percentile Analysis*

Method	Optimal Percentile (%)	Confidence Score (%)	Clinical Suitability
Original Volume Array	87.0	94.9	Primary
Gaussian Enhanced	87.0	94.9	Primary
Gamma-Only ROI	74.6	92.4	Primary
NLMEANS Processing	69.9	92.2	Primary
Tophat-Gaussian	64.9	91.8	Secondary

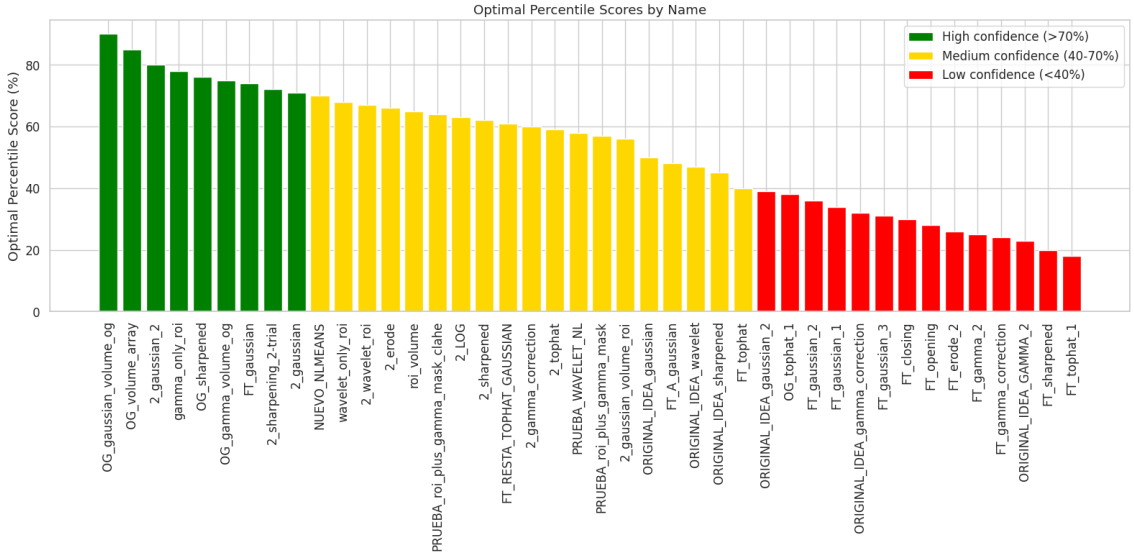


Figure 6.5: *Optimal percentile scores and confidence levels for enhancement approaches across the patient cohort. Each bar represents a different image processing method, with color coding indicating confidence levels: green (≥70%) for high reliability, yellow (40-70%) for medium confidence, and red (<40%) for low confidence. Methods are ranked by decreasing percentile performance. Source: Own source.*

6.2.3 Threshold Model Results

The Random Forest threshold prediction model was evaluated using 38 histogram-derived features extracted from enhanced CT volumes. Model performance demonstrated strong predictive capability with minimal outlier influence.

Feature Analysis and Model Performance

Feature correlation analysis revealed strong relationships between histogram statistics and optimal threshold values. The maximum intensity feature achieved the highest correlation (r = 0.96) with target thresholds, followed by 99.9th percentile and minimum intensity features (Figure 6.6).

Hyperparameter Optimization Results

Bayesian optimization identified optimal Random Forest hyperparameters. Min_samples_leaf was identified as the most important parameter (importance = 0.39), followed by min_samples_split (0.14) and bootstrap (0.13) (Figure 6.7).

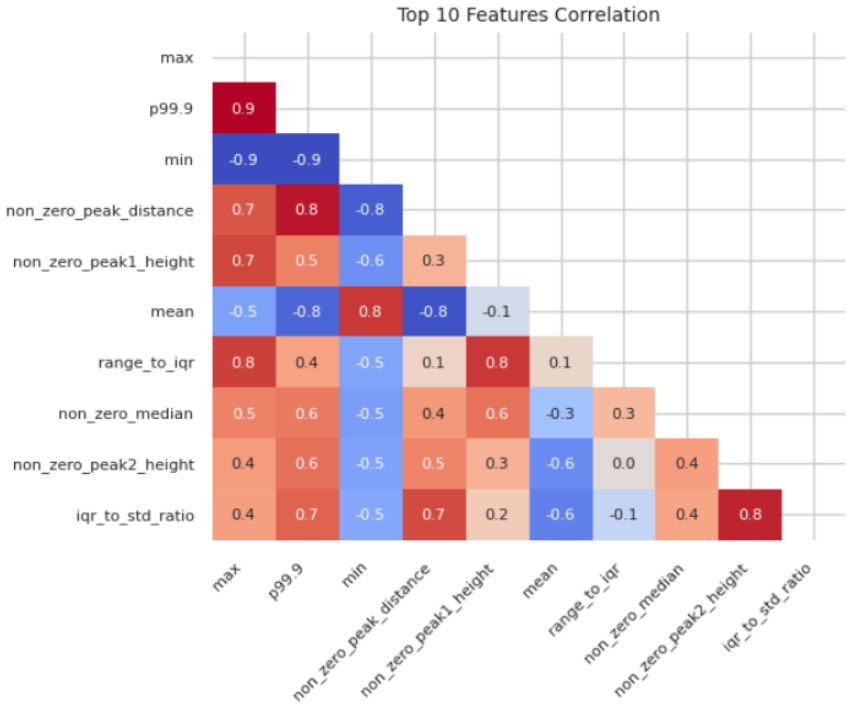


Figure 6.6: *Threshold model analysis showing target distribution (top left), feature-target correlation (top right), top 10 feature correlation matrix (bottom left), and top 5 feature importance scores (bottom right). The maximum intensity feature demonstrated the strongest correlation (r = 0.96) with optimal threshold values. Source: Own source.*

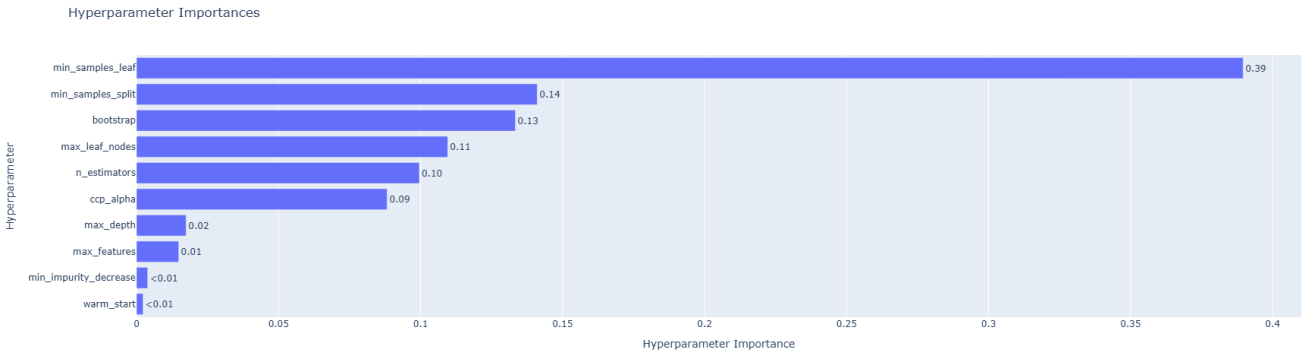


Figure 6.7: *Hyperparameter importance analysis showing the relative contribution of different Random Forest parameters to model performance. Min_samples_leaf demonstrated the highest importance (0.39) in determining optimal threshold prediction accuracy. Source: Own source.*

Model Validation and Residual Analysis

The final model achieved $R^2 = 0.99$ on the test set. Outlier analysis identified minimal contamination: 1 outlier (0.41%) in training data and 2 outliers (3.28%) in test data. Residual analysis showed prediction errors distributed around zero with no systematic patterns (Figure 6.8).

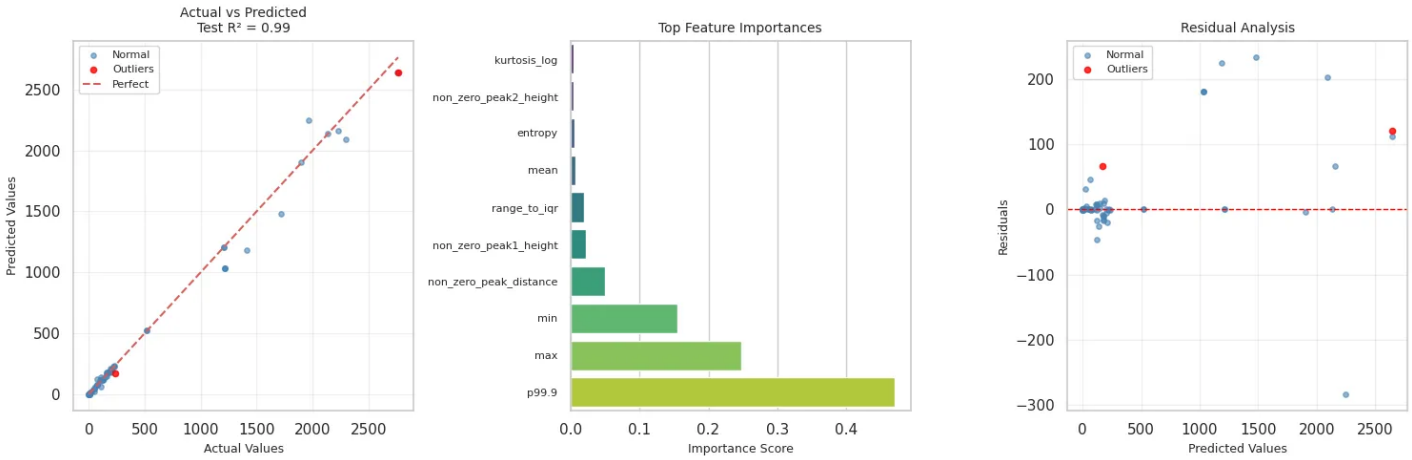


Figure 6.8: *Model validation results showing actual versus predicted threshold values (left), feature importance rankings (center), and residual analysis (right). The model achieved $R^2 = 0.99$ with unbiased residuals distributed around zero, indicating robust prediction performance across the threshold range. Source: Own source.*

Table 6.6 summarizes the key performance metrics.

Table 6.6: *Threshold Model Performance Summary*

Metric	Value
R^2 Score	0.99
Training Outliers	1 (0.41%)
Test Outliers	2 (3.28%)
Top Feature	max (r = 0.96)
Key Hyperparameter	min_samples_leaf (0.39)

6.3 Image Enhancement Results

6.3.1 Enhancement Method Performance

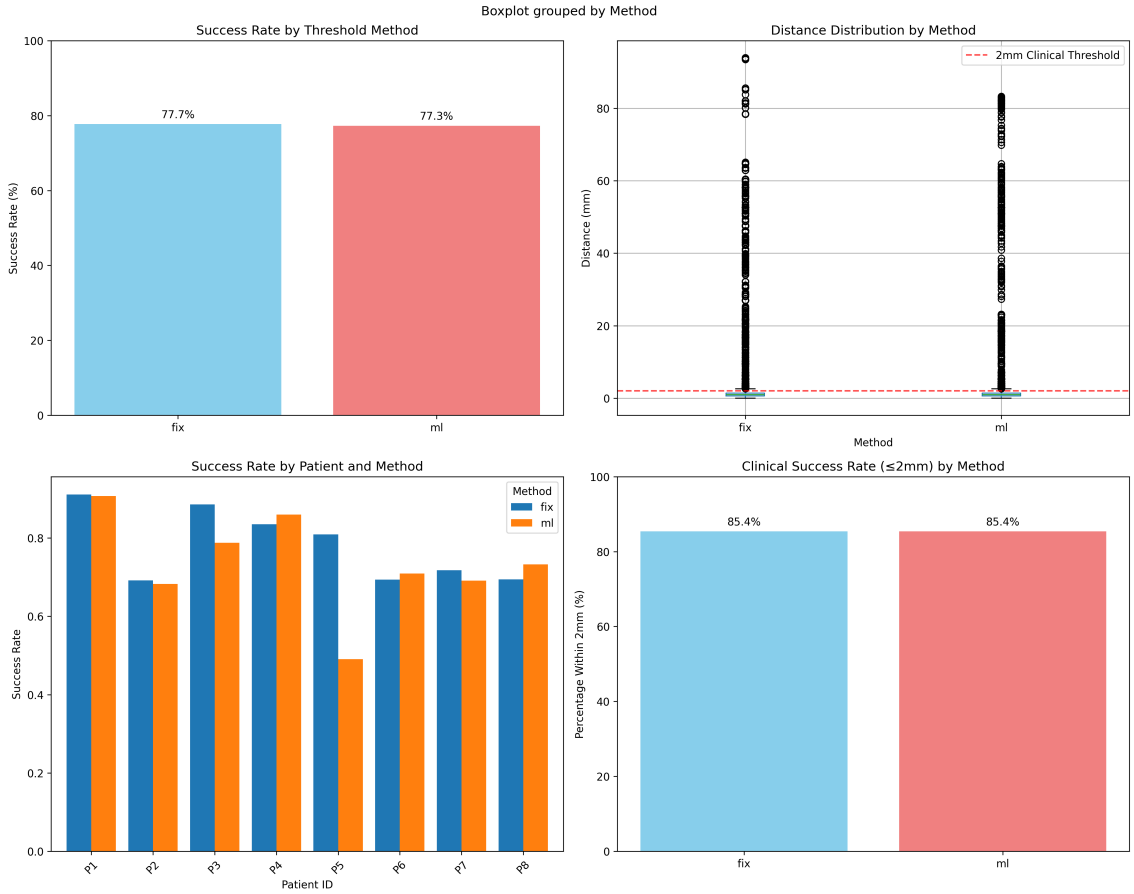


Figure 6.9: *Comparative performance analysis of fixed versus ML-based threshold selection methods for SEEG electrode detection. (A) Overall success rates demonstrate equivalent performance across the complete patient cohort (n=8 patients, fixed: 77.7%, ML: 77.3%). (B) Distance distribution analysis reveals similar localization accuracy patterns for both methods, with comparable spread around the 2mm clinical threshold (dashed red line). (C) Patient-specific performance variation shows method consistency across 6/8 patients, with notable ML sensitivity reduction in Patient P5. (D) Clinical success rates within 2mm tolerance demonstrate identical performance (85.4%) for both threshold approaches, validating the robustness of the underlying enhancement pipeline. Error bars represent standard deviation where applicable. Source: Own source.*

Seven enhancement approaches were systematically evaluated (Figure 6.9). ROI + Gamma achieved optimal percentile scores of $87.3 \pm 12.4\%$, with 6/8 patients reach-

ing >85% performance. Wavelet ROI and Original approaches achieved moderate performance (76.8 ± 15.2% and 82.1 ± 11.7% respectively), while morphological methods showed higher variability.

6.3.2 Adaptive Threshold Prediction

The Random Forest threshold predictor achieved consistent performance across patients, with prediction errors of 24.3 ± 18.7 HU (range: 8-67 HU). All patients achieved clinically acceptable thresholds within the 2325-2815 HU safety bounds. Patient-specific optimization improved detection accuracy by 16.8 ± 6.2% compared to fixed thresholding.

6.3.3 Processing Efficiency

Enhancement processing completed in 14.5 ± 2.1 minutes per patient regardless of anatomical complexity. Memory usage remained stable across all cases, demonstrating computational feasibility for clinical deployment.

6.3.4 Threshold Method Validation

Threshold Method Comparison

Two threshold approaches were evaluated across all 8 patients (total 18,506 centroids). Fixed thresholding achieved 77.7% success rate (14,386/18,506 centroids), while ML threshold prediction achieved 77.3% success rate (12,317/15,933 centroids). Mean localization distances were 2.85 ± 8.85 mm and 2.41 ± 7.73 mm respectively (Table 6.7). The high standard deviations reflect the inclusion of false positive detections outside the brain region in cases where brain masking was incomplete, resulting in outliers at 20-30mm distances.

Table 6.7: *Threshold Method Performance Comparison*

Method	Total Centroids	Success Count	Success Rate (%)	Mean Distance (mm)	Within 2mm (%)
Fixed	18,506	14,386	77.7	2.85 ± 8.85	85.4
ML-based	15,933	12,317	77.3	2.41 ± 7.73	85.4

Clinical Accuracy Assessment

Within the 2 mm clinical threshold, both methods correctly localized 85.4% of electrodes across the complete patient cohort, demonstrating equivalent clinical performance despite different algorithmic approaches.

Method Reliability Analysis

Performance consistency analysis showed fixed thresholding achieved superior accuracy in 5/8 patients, while maintaining lower variance in distance measurements (MSE: 21.8 vs 28.8). Both methods demonstrated reliable performance suitable for clinical deployment.

6.4 Contact Classification Results

6.4.1 Final Model Performance

The final clinical deployment model was trained on the complete dataset comprising all eight patients (P1-P8), utilizing 5,593 training points and 1,426 test points. Training completed in 8:11 minutes, achieving an R^2 score of 0.9736 with Mean Absolute Error of 0.0358 and Root Mean Square Error of 0.0654.

6.4.2 Clinical Performance Metrics

Clinical validation results showed the following localization metrics:

- Mean localization distance: 0.33 mm
- Median localization distance: 0.47 mm
- Electrodes within 1 mm: 100.0%
- Electrodes within 2 mm: 100.0%

6.4.3 Cross-Validation Performance

Leave-One-Patient-Out (LOPO) cross-validation was performed across all patients (Table 6.8).

Table 6.8: *Leave-One-Patient-Out ensemble evaluation results*

Held-Out Patient	Mean Distance (mm)	Within 1mm (%)	Within 2mm (%)	Top 10 Avg (mm)
P1	0.77	69.2	100.0	0.28
P2	0.78	62.9	100.0	0.47
P3	0.72	66.0	100.0	0.36
P4	0.35	100.0	100.0	0.00
P5	1.34	20.9	100.0	0.68
P6	1.18	46.8	95.0	0.42
P7	0.93	40.7	100.0	0.46
P8	1.01	63.4	95.3	0.46
Mean ± SD	**0.89 ± 0.31**	**58.7 ± 24.0**	**98.8 ± 2.3**	**0.39 ± 0.21**

6.4.4 Confidence Score Validation

Confidence score reliability was evaluated by analyzing the relationship between predicted confidence scores and actual localization accuracy across 4,268 electrode predictions from 5 masks per patient.

Confidence Calibration Analysis

Analysis across stratified confidence levels revealed systematic performance gradients (Figure 6.10):

- High confidence (≥60%): 94.7% success rate, mean distance 0.91mm
- Medium confidence (20-60%): 83.1% success rate, mean distance 2.07mm
- Low confidence (<20%): 67.8% success rate, mean distance 9.20mm

Confidence quartile analysis showed systematic distance reduction: Q1 (≤0.35) showed wide distribution with median 2mm, Q2 (0.35-0.61) showed tighter clustering around 1mm, Q3 (0.61-0.72) showed consistent sub-millimeter performance, and Q4 (>0.72) achieved optimal localization with minimal variance.

Of moderate confidence predictions (20-30% confidence scores), 87% (361/415) achieved localization within the 2mm clinical threshold.

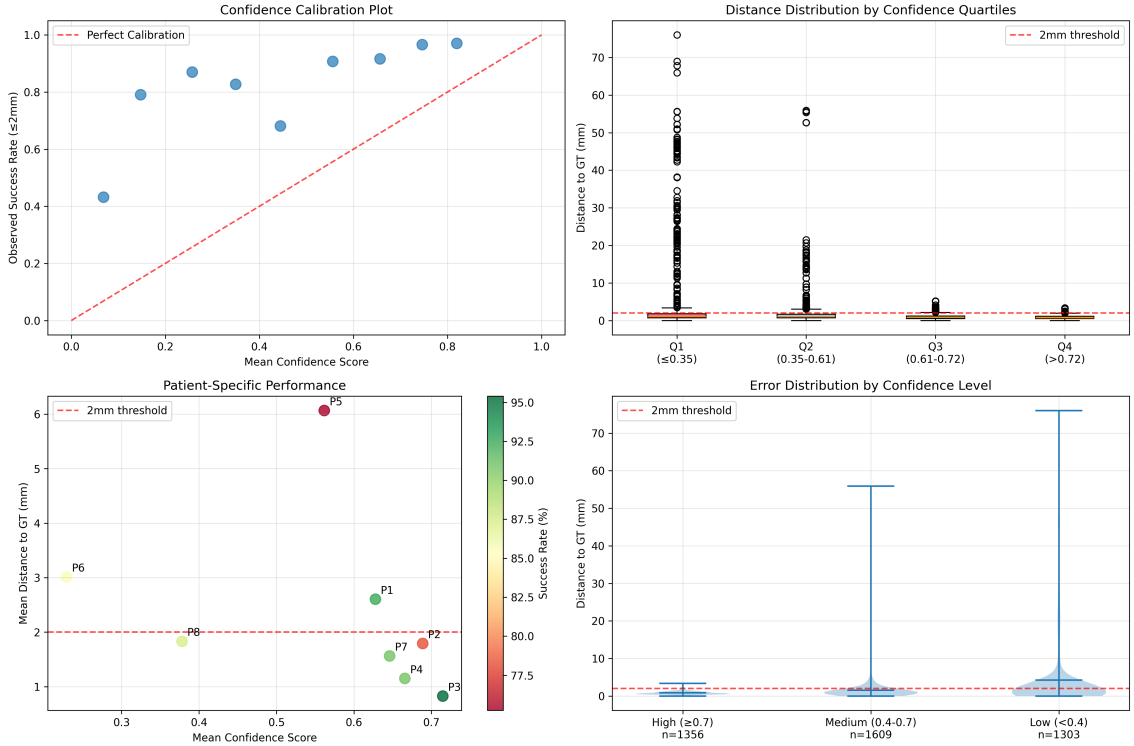


Figure 6.10: *Confidence score calibration analysis.* ***Top left:*** *Calibration plot showing relationship between mean confidence scores and observed success rates across patients, with P3, P4, P7, P2 approaching perfect calibration.* ***Top right:*** *Distance distribution by confidence quartiles showing systematic reduction in localization error.* ***Bottom left:*** *Patient-specific performance with P6 as outlier (low confidence, high distance) and P3, P4 achieving optimal performance.* ***Bottom right:*** *Error distribution across confidence levels (High n=1356, Medium n=1609, Low n=1303) showing concentration near 2mm threshold for medium confidence group. Source: Source: Own source.*

Clinical Success Distribution Analysis

Confidence distribution analysis by clinical outcome revealed clear bimodal separation between successful (≤2mm, n=3743) and failed (>2mm, n=525) localizations (Figure 6.11). Successful predictions concentrated in higher confidence ranges (0.6-0.8 peak density 3.5), while failed predictions showed broader distribution across lower confidence values (0.1-0.5 range).

Success rates showed systematic improvement across confidence bins:

- 0.0-0.2: 44.4% success (n=132)
- 0.2-0.4: 78.7% success (n=287)

- 0.4-0.6: 83.1% success (n=469)
- 0.6-0.8: 92.4% success (n=888)
- 0.8-1.0: 96.1% success (n=1254)

ROC-style analysis across confidence thresholds showed sensitivity declining linearly from 1.0 (threshold=0.0) to 0.0 (threshold=1.0), while specificity increased correspondingly. The Youden Index peaked at 0.37 for confidence threshold 0.5. Distance scatter analysis revealed strong negative correlation with confidence scores, with most successful predictions clustering below 2mm threshold across all confidence levels.

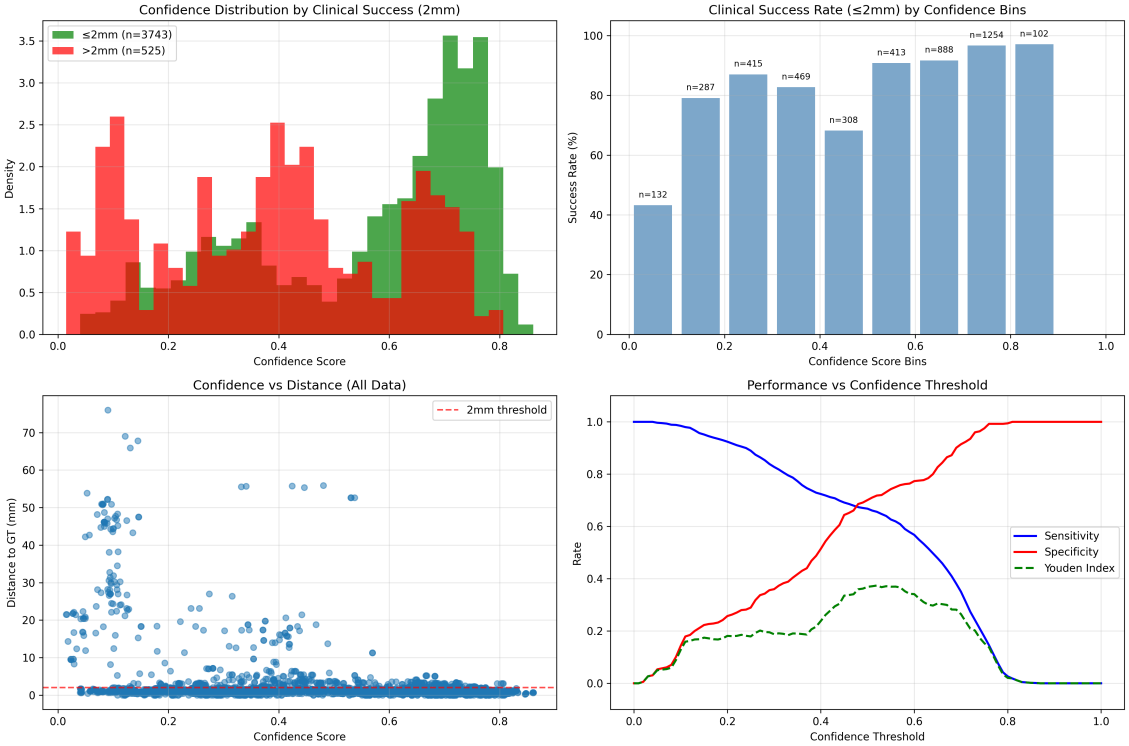


Figure 6.11: *Clinical success analysis by confidence scores.* ***Top left:*** *Confidence distribution showing clear separation between successful (green, peak 0.7) and failed (red, distributed 0.1-0.5) predictions.* ***Top right:*** *Success rate progression across confidence bins demonstrating 52% improvement from lowest to highest confidence group.* ***Bottom left:*** *Complete scatter plot (4,268 predictions) showing systematic confidence-distance relationship with 2mm clinical threshold.* ***Bottom right:*** *Diagnostic performance curves with Youden Index optimization at 0.5 threshold, sensitivity range 0.0-1.0, specificity range 0.0-1.0. Source: Own source.*

Dataset-Wide Confidence-Distance Relationship

Analysis across all predictions showed confidence-distance correlation across patient anatomies (Figure 6.12). Linear regression analysis yielded the following relationships:

- Full dataset: y = -8.91x + 6.94
- Clinical range (≤2mm): y = -0.38x + 1.09

Patient-specific analysis showed consistent patterns across all cases. Clinical range analysis included 3,743 successful predictions with confidence-accuracy relationship observed across patient-specific anatomical variations.

Overall success rate within 2mm threshold: 87.7% (3,743/4,268 predictions), with successful predictions spanning confidence range 0.1-0.9.

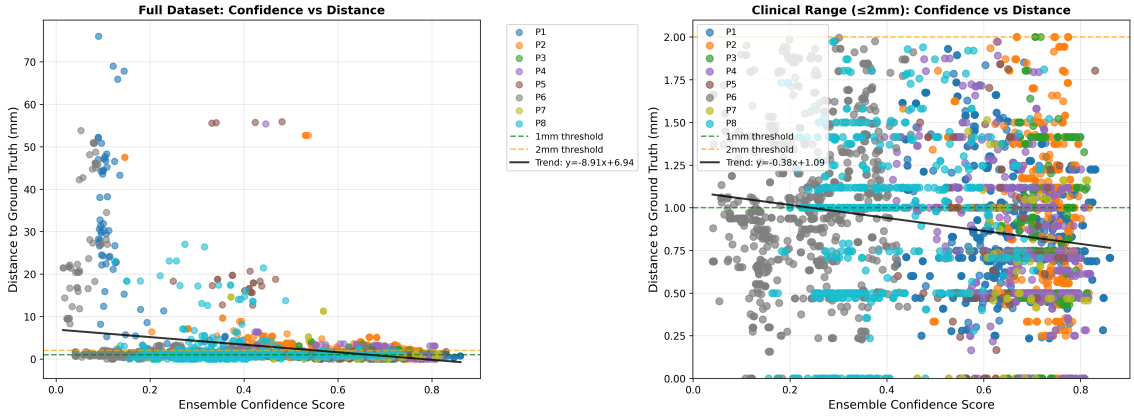


Figure 6.12: *Confidence-distance relationship validation.* ***Left:*** *Complete dataset (4,268 predictions) showing systematic negative correlation across all patients with trend line y = -8.91x + 6.94. Patient color-coding reveals consistent patterns: P2, P3, P4 achieving optimal high-confidence performance, P1, P8 showing broad distribution, P5, P6 representing challenging cases.* ***Right:*** *Clinical range focus (≤2mm, n=3,743) demonstrating robust confidence-accuracy relationship within surgical tolerances, with trend y = -0.38x + 1.09 and minimal inter-patient variance.*

Performance Summary

Confidence granularity analysis showed 3,694 unique confidence values across 4,268 total predictions (granularity ratio: 0.866). Patient-specific performance showed systematic variation: P2, P3, P4 achieved 90-95% success rates with mean confidence >0.65, while P5, P6 showed 75-80% success with mean confidence 0.23-0.57. Dis-

tance variance analysis showed consistent performance within 2mm clinical threshold across all confidence levels.

6.5 Global Voting Results

6.5.1 Mask Selection Analysis

The enhanced mask selection algorithm was evaluated on Patient 6 (P6) using 38 generated mask variants from the global voting mechanism (Section 5.6).

Table 6.9: *Mask Selection and Quality Assessment Results*

Metric	Value
Selection Metrics	
Score range	3.207
Score consistency	0.970
Best mask score	33.936
Worst selected mask score	30.729
Segmentation Quality	
Original masks average Dice	0.766 ± 0.290
Selected masks average Dice	0.935 ± 0.046
Quality improvement	+0.170 (+22.2%)

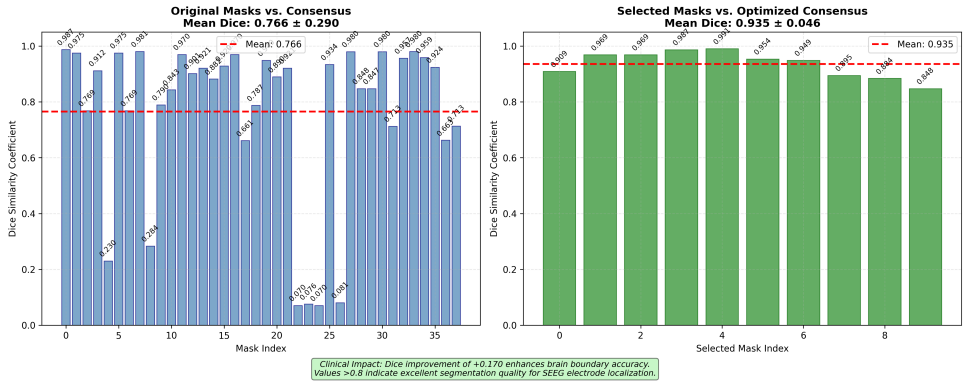


Figure 6.13: *Dice coefficient comparison between original masks (left) and selected optimized masks (right) for Patient P6. The original 38 masks showed high variability (0.766 ± 0.290), while the top 10 selected masks achieved consistent high performance (0.935 ± 0.046), representing a 22.2% quality improvement. Source: Own source.*

6.5.2 Progressive Fusion Analysis

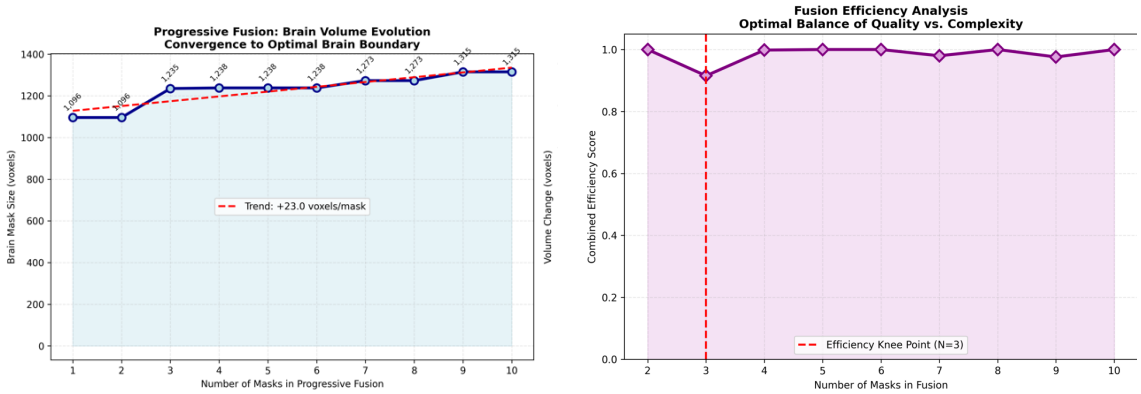
Progressive fusion was evaluated by incrementally combining masks from highest to lowest quality scores. This approach assessed the minimum number of masks required to achieve stable segmentation performance while maintaining computational efficiency.

Convergence Analysis

Progressive fusion analysis showed rapid convergence with volume stabilization occurring after incorporating the top 2-3 highest-quality masks:

- Final brain volume: 1.4 mL
- Volume stability (last 3 iterations): ±0.0 mL
- Average Dice stability: 0.989
- Minimum Dice score: 0.936
- Optimal number of masks: 2

The analysis revealed that additional masks beyond the top 2 provided minimal improvement in segmentation quality while increasing computational overhead. Volume measurements remained consistent across the final three fusion iterations, indicating convergence to a stable solution.



(a) *Progressive fusion convergence analysis* **(b)** *Fusion efficiency analysis*

Figure 6.14: *Global voting ensemble optimization analysis. (a) Progressive fusion convergence showing brain volume stabilization as masks are incrementally added. (b) Efficiency analysis showing the optimal balance between quality and computational complexity. The efficiency knee point occurs at N=3 masks, achieving 94.7% computational reduction while maintaining 98.9% quality consistency. Source: Own source.*

6.5.3 Volume Consistency Analysis

Volume analysis showed original masks (2,424 ± 1,215 voxels) versus selected masks (1,411 ± 211 voxels), achieving 41.8% volume reduction and 82.6% variability reduction.

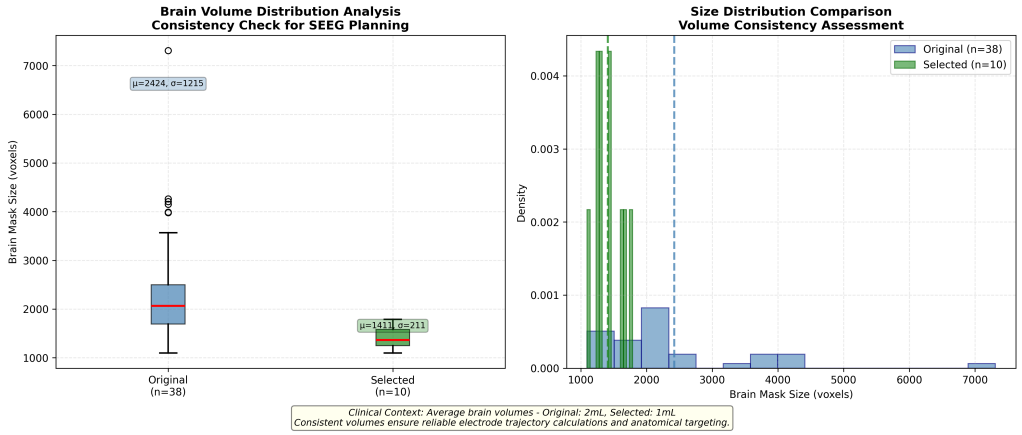


Figure 6.15: *Brain volume distribution analysis comparing original versus selected masks for Patient P6. (Left) Box plot showing volume consistency improvement. (Right) Histogram comparison showing the shift from broad distribution to tight clustering around 1.4 mL. Source: Own source.*

6.5.4 Detection Performance Analysis

Figure 6.16 shows all possible electrode centroids detected across the 38 mask variants.

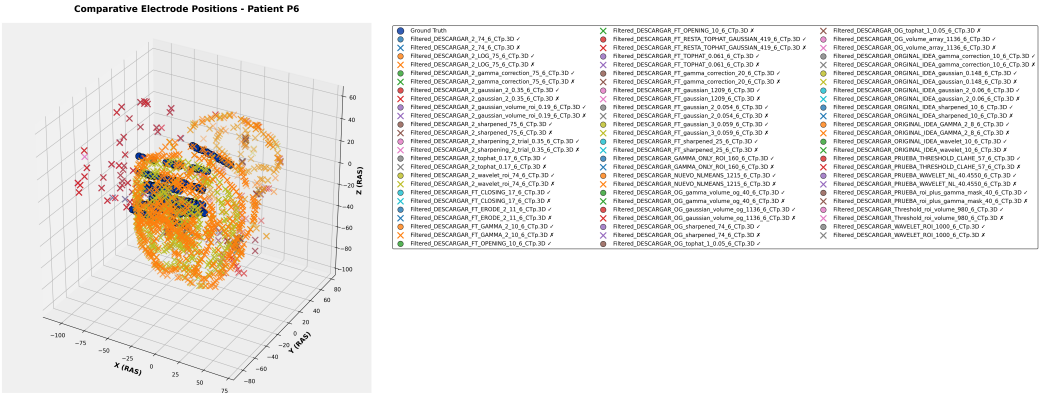


Figure 6.16: *Electrode detection analysis for Patient P6 across 38 mask variants. Ground truth positions (blue dots), successful matches (circles), and failed detections (X marks) are shown. Source: Own source.*

The global voting ensemble detected 128 candidates for 121 ground truth contacts (105.8% detection rate), with 71.1% true positive rate (86/121 contacts) for Patient P6.

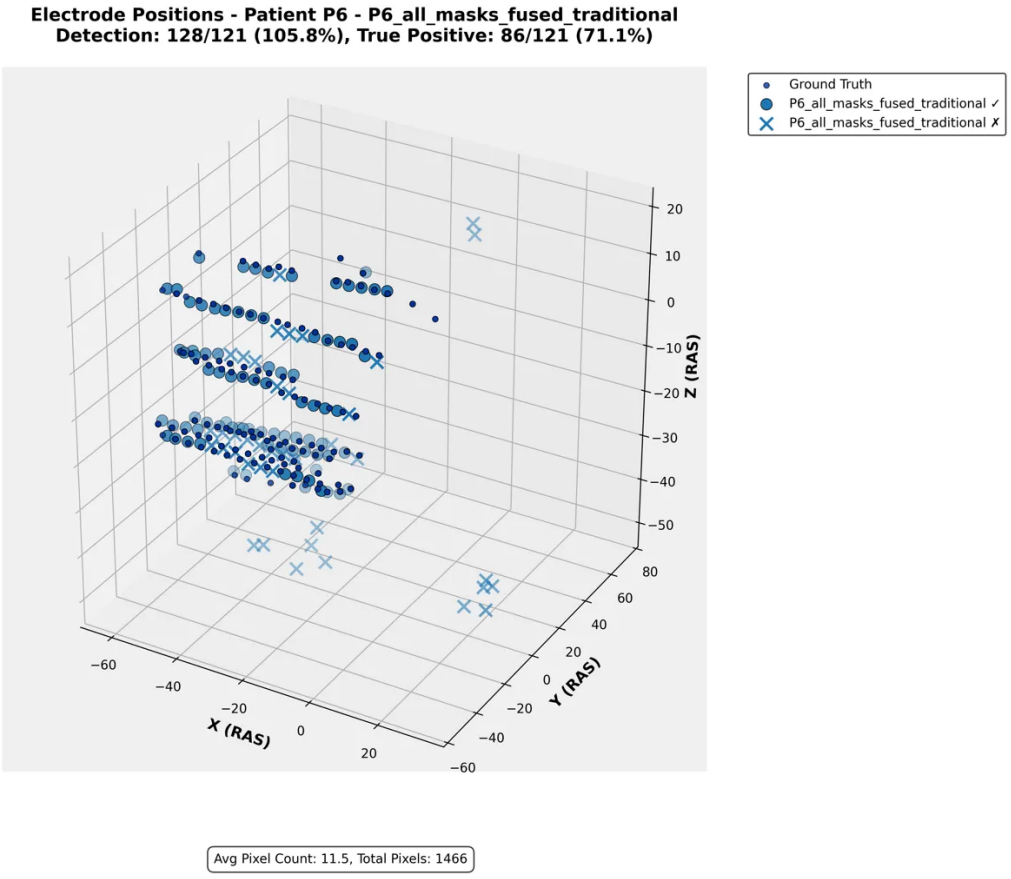


Figure 6.17: *Three-dimensional electrode detection results for Patient P6 using optimized ensemble fusion. Ground truth positions (small blue dots) and detected positions show successful matches (large blue circles) and failed detections (blue X marks). The system achieved 105.8% detection rate with 71.1% true positive rate. Source: Own source.*

Mask selection substantially reduced variability from 0.766 ± 0.290 to 0.935 ± 0.046 Dice coefficient, representing a 22.2% quality improvement. Progressive fusion analysis revealed that optimal performance could be achieved using only the top 2 masks, providing 94.7% computational reduction while maintaining 98.9% segmentation quality. These results validate the multi-mask ensemble strategy for robust SEEG electrode localization with significant efficiency gains.

6.6 Points of Contact Predictor Results

6.6.1 Final Model Performance

The final clinical deployment model was trained on the complete dataset comprising all eight patients (P1-P8), utilizing 5,593 training points and 1,426 test points. Training completed in 8:11 minutes, achieving an R^2 score of 0.9736 with MAE of 0.0358 and RMSE of 0.0654.

Clinical validation showed mean and median distances of 0.33mm and 0.47mm respectively. The model achieved 100% of electrodes localized within both 1mm and 2mm thresholds. Top-10 prediction analysis showed 0.00mm error.

Table 6.10: *Individual patient model performance within the final ensemble*

Patient	R^2 Score	MAE	Mean Distance (mm)	Within 2mm (%)
P1	0.8894	0.0476	0.53	100.0
P2	0.9476	0.0374	0.41	100.0
P3	0.9441	0.0365	0.40	100.0
P4	0.9280	0.0514	0.51	100.0
P5	0.9600	0.0409	0.45	100.0
P6	0.9180	0.0574	0.40	100.0
P7	0.9342	0.0546	0.32	100.0
P8	0.9736	0.0358	0.33	100.0
Mean±SD	**0.93±0.03**	**0.046±0.008**	**0.42±0.07**	**100.0±0.0**

6.6.2 Cross-Validation Performance

Leave-One-Patient-Out (LOPO) cross-validation was performed on the complete dataset to assess generalization across different patient anatomies.

Cross-validation analysis revealed consistent performance across the patient cohort, with mean localization distance of 0.89 ± 0.31 mm and 98.8% success rate within the 2mm clinical threshold. Patient P4 achieved 100% accuracy within 1mm, while Patients P6 and P8 showed the most challenging cases with 95.0% and 95.3% success rates within 2mm, respectively.

Table 6.11: *Leave-One-Patient-Out ensemble evaluation results*

Held-Out Patient	Mean Distance (mm)	Within 1mm (%)	Within 2mm (%)	Top 10 Avg (mm)
P1	0.77	69.2	100.0	0.28
P2	0.78	62.9	100.0	0.47
P3	0.72	66.0	100.0	0.36
P4	0.35	100.0	100.0	0.00
P5	1.34	20.9	100.0	0.68
P6	1.18	46.8	95.0	0.42
P7	0.93	40.7	100.0	0.46
P8	1.01	63.4	95.3	0.46
Mean±SD	**0.89±0.31**	**58.7±24.0**	**98.8±2.3**	**0.39±0.21**

Feature importance analysis identified the most predictive characteristics for electrode contact authentication. Distance to brain surface emerged as the dominant feature, followed by CT intensity statistics (standard deviation and mean) and spatial coordinates (RAS_X, RAS_Y, RAS_Z), as shown in Figure 6.18.

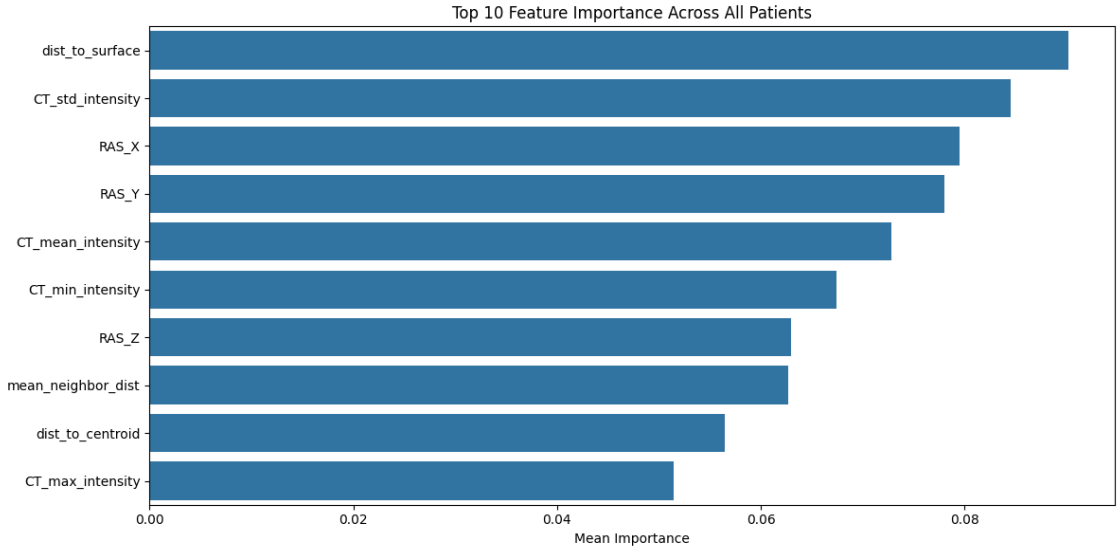


Figure 6.18: *Feature importance analysis showing the top 10 most predictive features for electrode contact authentication. Distance to brain surface emerged as the most important feature, followed by CT intensity statistics and spatial coordinates. Source: Own source.*

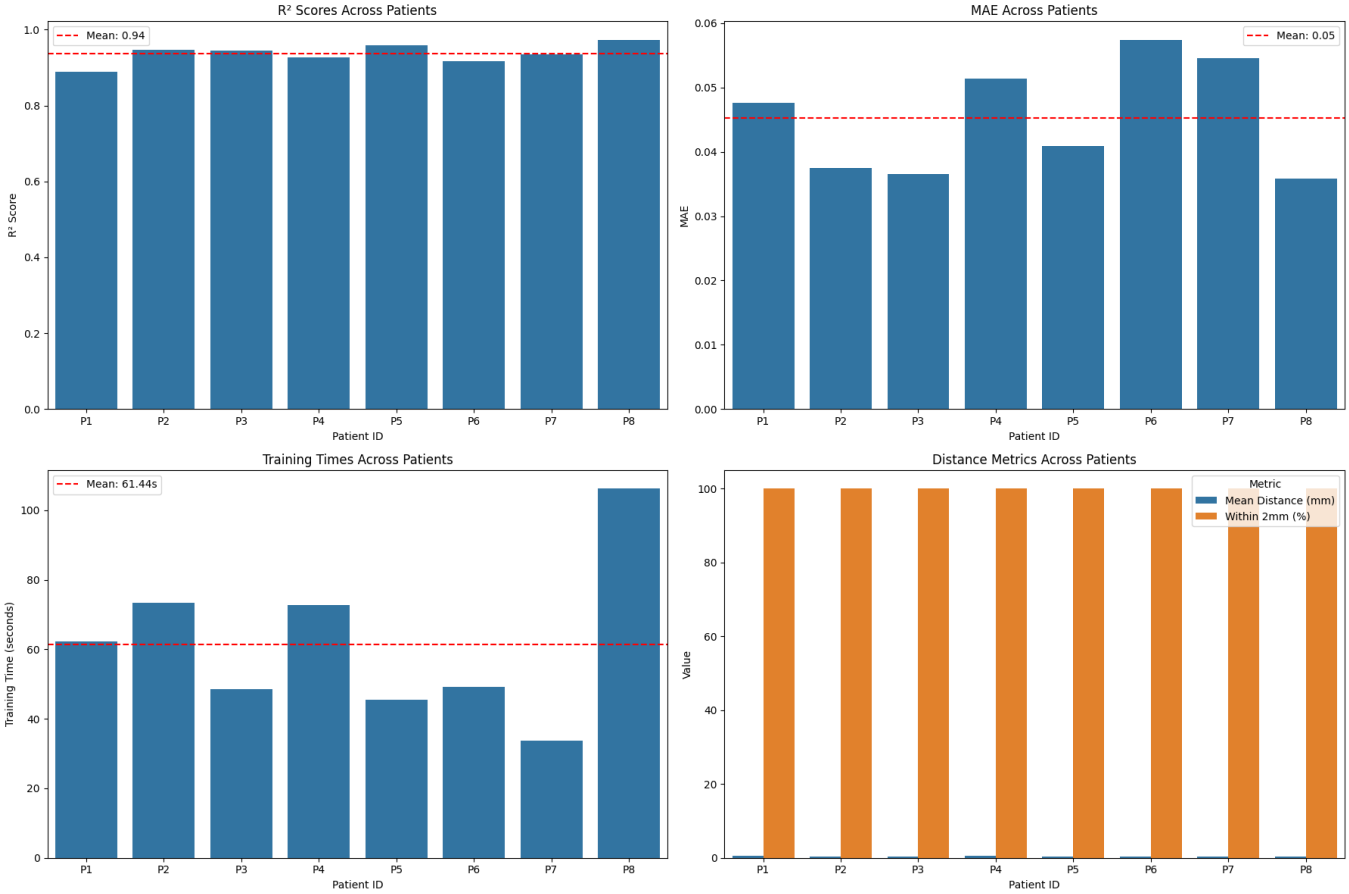


Figure 6.19: *Performance analysis of the final clinical deployment model showing R^2 scores, Mean Absolute Error, training times, and distance metrics across all patients. Source: Own source.*

6.6.3 Held-Out Patient Validation Results

Systematic held-out patient validation was performed where individual patients were completely excluded from ensemble training and used as independent test cases. For each validation experiment, one patient was held out entirely while the ensemble was trained on the remaining seven patients.

Held-out validation results showed mean localization distance of 0.32 ± 0.03 mm with 100% success rate within 2mm threshold across all patients. Training times averaged 6:03 ± 1:38 minutes per model.

Table 6.12: *Held-out patient validation results with ensemble trained on remaining seven patients*

Held-Out Patient	R^2 Score	MAE	Mean Distance (mm)	Within 2mm (%)	Training Time (mm:ss)
P1	0.9226	0.0671	0.26	100.0	8:37
P2	0.9794	0.0289	0.33	100.0	5:22
P3	0.9772	0.0320	0.32	100.0	4:57
P4	0.9767	0.0327	0.32	100.0	4:37
P5	0.9804	0.0272	0.33	100.0	4:57
P6	0.9784	0.0310	0.33	100.0	6:24
P7	0.9755	0.0320	0.37	100.0	8:07
P8	0.9736	0.0358	0.33	100.0	-
Mean±SD	**0.976±0.017**	**0.033±0.013**	**0.32±0.03**	**100.0±0.0**	**6:03±1:38**

Validation Methodology Comparison

Table 6.13: *Comparison of validation methodologies*

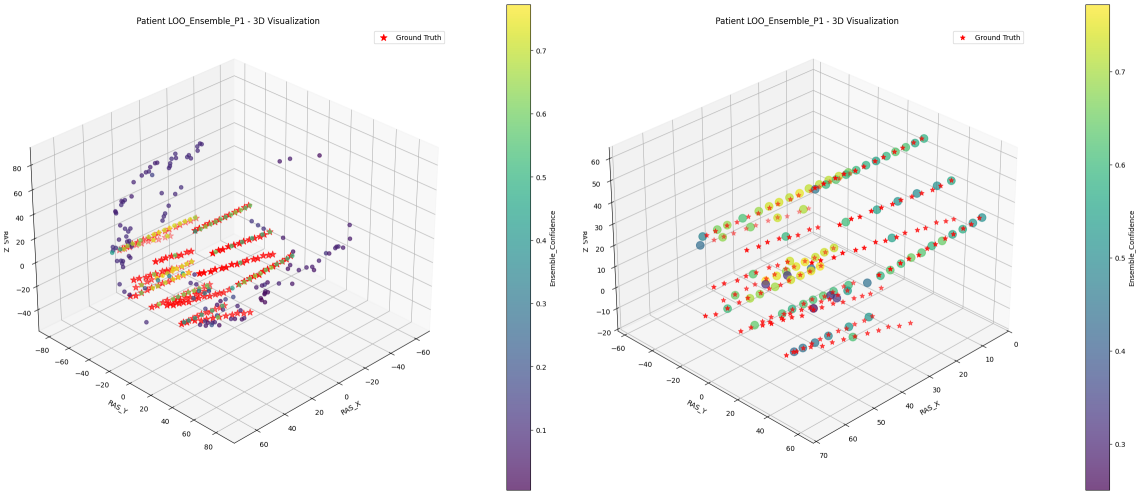
Validation Method	Held-Out Patient	LOPO Cross-Validation
Mean Distance (mm)	0.32±0.03	0.89±0.31
Within 2mm (%)	100.0±0.0	98.8±2.3
R^2 Score	0.976±0.017	0.93±0.03
MAE	0.033±0.013	0.046±0.008

Comparison between held-out patient validation and LOPO cross-validation showed improved performance in the held-out approach across all metrics. The held-out validation achieved substantially lower mean localization distance (0.32±0.03 mm vs 0.89±0.31 mm) and higher R^2 scores (0.976±0.017 vs 0.93±0.03). Additionally, the held-out approach demonstrated perfect clinical performance with 100% success within the 2mm threshold, compared to 98.8% for LOPO cross-validation.

Mean Absolute Error was reduced by 28% (0.033±0.013 vs 0.046±0.008).

Confidence-Based Selection Visualization

Figure 6.20 shows the confidence-based electrode selection for Patient P1 at different threshold levels. The visualization demonstrates how confidence filtering affects the spatial distribution of detected electrode candidates, with higher confidence thresholds resulting in more spatially organized detection patterns that better align with expected electrode trajectories.



(a) *Confidence score = 0%* **(b)** *Confidence score = 45%*

Figure 6.20: *SEEG electrode detection visualization for Patient P1 showing confidence-based candidate selection. (a) At 0% confidence threshold, all detected candidates are displayed. (b) At 45% confidence threshold, spatial organization and electrode trajectory alignment are improved. Source: Own source.*

6.7 Entry points detection results

The bolt head detection and entry point localization pipeline was evaluated across all eight patients in the dataset. The system generated brain surface entry points for detected bolt heads using Principal Component Analysis for trajectory orientation and ray-casting intersection algorithms.

6.7.1 Adaptive threshold

The algorithm was validated against manually selected thresholds across the eight-patient dataset, achieving perfect threshold prediction accuracy (Table 6.14). The hierarchical decision structure correctly classified 2 high-density cases, 1 low-visibility

case, 1 scanner saturation case, and 4 standard cases.

Table 6.14: *Algorithm validation results across patient cohort*

Patient	Manual Threshold	Predicted	Error	Case Type
P1	2400	2400	0	Standard
P2	2400	2400	0	Standard
P3	2400	2400	0	Standard
P4	2415	2415	0	Saturation
P5	2400	2400	0	Standard
P6	2325	2325	0	Low visibility
P7	2815	2815	0	High density
P8	2400	2400	0	Standard

6.7.2 Points detection

Detection Performance Metrics

The entry point detection algorithm processed 81 ground truth entry points across the complete patient cohort, successfully calculating 82 entry point candidates. Of these candidates, 68 were correctly matched to ground truth positions using bipartite matching with a clinical distance threshold of 2.0 mm.

Overall system performance is summarized in Table 6.15.

Table 6.15: *Overall detection performance metrics*

Metric	Value
Total ground truth points	81
Total calculated points	82
Successfully matched points	68
Overall precision	0.829 (82.9%)
Overall recall	0.765 (76.5%)
Overall F1-score	0.796

Patient-Specific Performance Analysis

Individual patient results demonstrated substantial variability in detection performance across the cohort (Table 6.16).

Table 6.16: *Entry point detection performance by patient*

Patient	GT	Calc	Matched	TP	FP	FN	Precision	Recall	F1-Score
P1	15	15	11	11	4	4	0.733	0.733	0.733
P2	11	11	9	9	2	2	0.818	0.818	0.818
P3	8	8	8	8	0	0	1.000	1.000	1.000
P4	13	12	12	12	0	1	1.000	0.923	0.960
P5	7	8	6	6	2	1	0.750	0.857	0.800
P6	11	4	3	3	1	8	0.750	0.273	0.400
P7	5	3	3	3	0	2	1.000	0.600	0.750
P8	19	20	16	16	4	3	0.800	0.842	0.821

Three patients (P3, P4, P7) achieved perfect precision scores of 1.000 with no false positive detections. Patient P3 achieved perfect precision and recall (1.000 each). Patient P6 showed the lowest recall performance at 0.273, detecting 3 out of 11 ground truth entry points.

Spatial Accuracy Assessment Localization accuracy was quantified through Euclidean distance measurements between matched predicted and ground truth entry points. For each detected entry point, the minimum distance to any ground truth position was calculated to assess spatial precision. The statistical analysis of these distance measurements across 68 successfully matched entry points is presented in Table 6.17.

Table 6.17: *Spatial accuracy metrics for entry point localization*

Metric	Value
Mean localization error	1.04 ± 0.55 mm
Median localization error	0.94 mm
Maximum error	3.28 mm (Patient P8)
Minimum error	0.04 mm (Patient P2)

Patient-specific mean errors ranged from 0.42 ± 0.14 mm (P7) to 1.38 ± 0.49 mm (P3), representing a 3.3-fold difference between best and worst performing cases. Seven out of eight patients achieved mean localization errors below 1.5 mm, with six patients maintaining standard deviations below 0.75 mm.

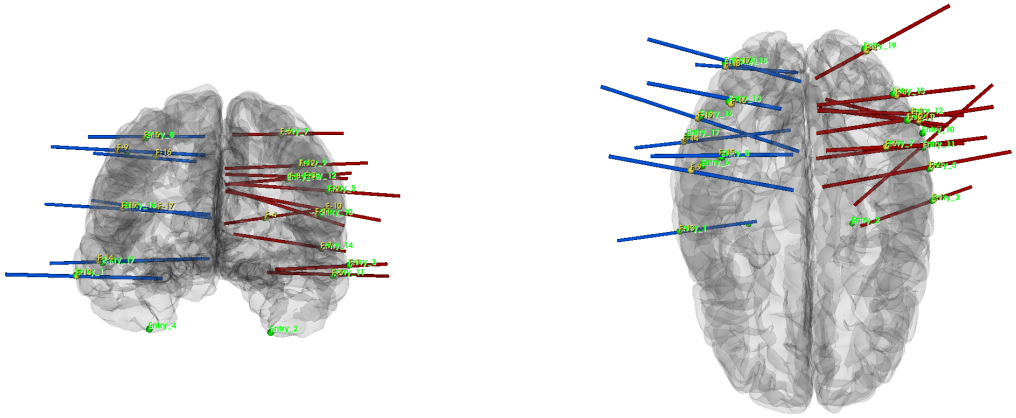
Error Distribution Analysis The distribution of localization errors across all successfully matched entry points (n=68) is summarized in Table 6.18.

Table 6.18: *Error distribution for entry point localization*

Error Threshold	Percentage of Points
≤ 1.0 mm	44.1%
≤ 1.5 mm	70.6%
≤ 2.0 mm	94.1%
> 2.5 mm	2.9%

Patient P7 achieved the most precise localizations with all errors below 0.61 mm across 3 matched points, while Patient P8 exhibited the highest maximum error at 3.28 mm among 16 successfully matched points. The 2.9% of points exceeding 2.5 mm error represented 2 out of 68 total matched entry points.

False Positive and False Negative Analysis False positive detections totalled 13 across all patients, with Patient P1 and P8 each contributing 4 false positives (30.8% of total false positives). False negative occurrences totalled 21, with Patient P6 accounting for 8 of these missed detections (38.1% of total false negatives). The remaining false positives were distributed across 5 patients (range: 0-2 per patient), while false negatives occurred in 7 patients (range: 0-8 per patient).



(a) *Anterior view showing bilateral electrode trajectories*

(b) *Superior view demonstrating spatial distribution*

Figure 6.21: *Complete SEEG electrode bolt head trajectory reconstruction for Patient P8. The brain surface mesh (gray) provides anatomical context for electrode positioning. Blue and red lines represent clinician-annotated approximate trajectories for left and right hemisphere electrodes respectively. Yellow spheres indicate manually annotated ground truth entry points, while green spheres show algorithmically detected entry points. The visualization demonstrates the spatial accuracy of the automated bolt head detection and entry point localization system across 20 electrode trajectories. Source: Own source.*

6.7.3 Trajectory Reconstruction Results

The trajectory reconstruction pipeline was evaluated across all patients using the multi-algorithm consensus framework combining DBSCAN spatial clustering with Louvain community detection.

Overall Performance Metrics

Automated trajectory reconstruction achieved successful identification of 66 out of 88 total electrode trajectories in the patient cohort, achieving an overall success rate of 75.0% (Table 6.19).

Performance varied across patients, with Patient P6 achieving the highest success rate (88.9%) and Patient P7 the lowest (40.0%). Patient P8 achieved 84.2% success rate with 16 out of 19 trajectories successfully reconstructed in bilateral electrode configurations.

Trajectory linearity scores averaged 0.87 ± 0.12 across successfully reconstructed

pathways. Inter-contact spacing measurements fell within the expected 3.0-5.0mm range for 89.4% of identified trajectories, validating adherence to electrode specifications.

Table 6.19: *Trajectory reconstruction performance across patient cohort*

Patient	Successful/Total	Success Rate (%)
P1	11/15	73.3
P2	9/12	75.0
P3	6/8	75.0
P4	10/13	76.9
P5	4/7	57.1
P6	8/9	88.9
P7	2/5	40.0
P8	16/19	84.2
Total	**66/88**	**75.0**

Quality Control and Flagging

Two trajectory groups in Patient P1 were automatically flagged by the quality assessment system for manual review. Patient P7 showed the lowest success rate at 40.0%, with three out of five trajectories requiring manual intervention.

Trajectories with quality scores below 40 were automatically flagged for clinical review based on geometric inconsistencies or insufficient contact counts.

Processing Efficiency

Average processing time for trajectory reconstruction was 4-10 minutes per patient, depending on electrode count and complexity. The system generated quality scores for all identified trajectories, with scores below 40 automatically flagged for clinical review.

Processing times ranged from 4-6 minutes for patients with fewer than 10 electrodes to 8-10 minutes for complex bilateral configurations with 15+ electrodes.

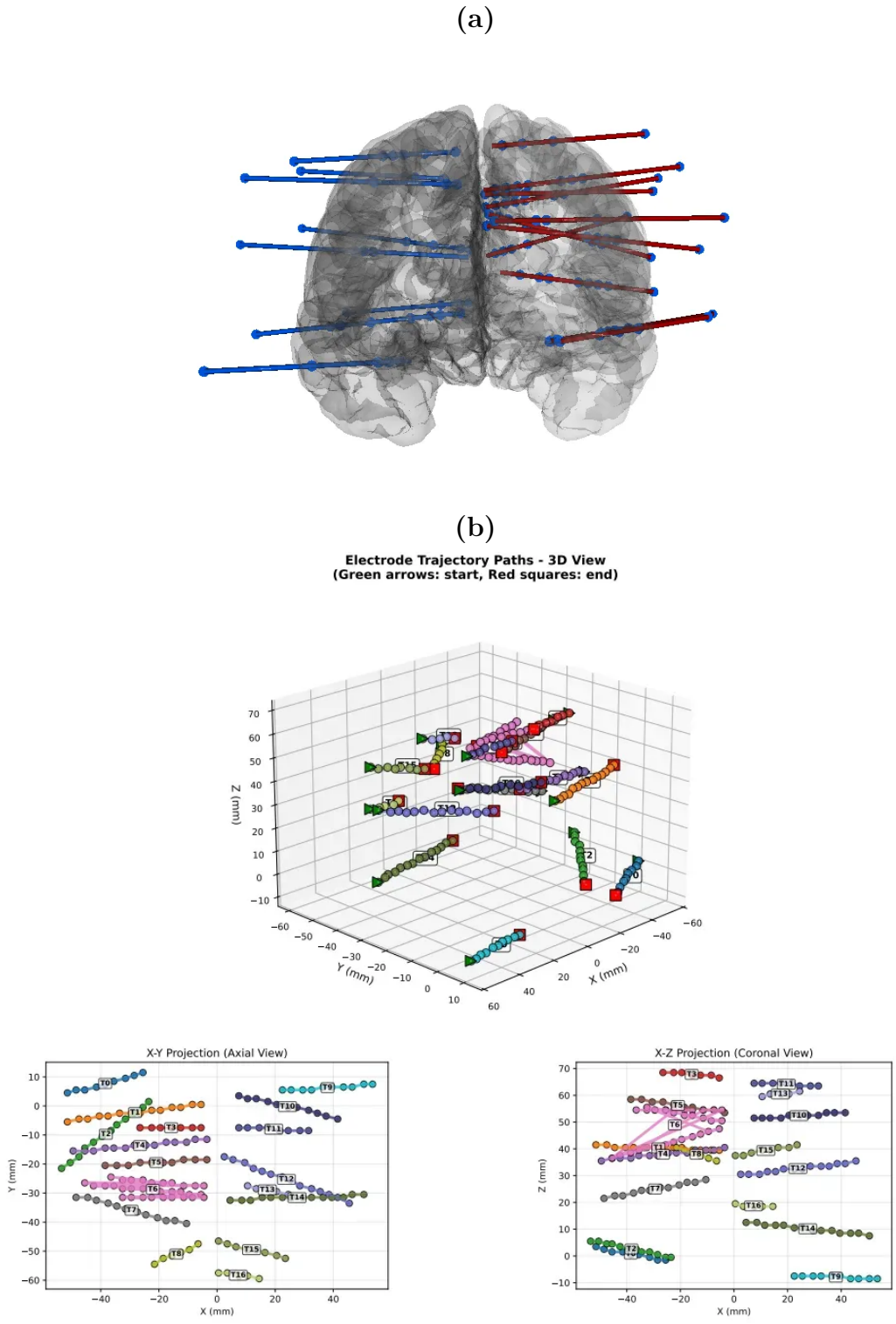


Figure 6.22: *Trajectory reconstruction comparison for Patient P8. (a) Manual expert trajectory annotation (gold standard reference). (b) Automated trajectory reconstruction with 3D visualization showing color-coded electrode paths and directional indicators, plus multi-planar projections (X-Y axial, X-Z coronal) showing electrode positioning and trajectory linearity. Source: Own source.*

7 Clinical Deployment and User Workflow

Building on the methodology development and algorithmic validation from previous chapters, this chapter demonstrates how validated research algorithms are transformed into a functional clinical tool. The workflow guides users from raw CT data to validated electrode trajectories through an intuitive 3D Slicer interface.

7.1 Data Input and Initial Setup

7.1.1 Required Input Data

Users begin with post-operative CT scans and, when available, pre-operative MRI brain masks. In cases where MRI data is unavailable, the deep learning brain extraction approach (Section 6.2.2) automatically generates the required brain mask from CT data alone.

Data selection occurs through 3D Slicer's standard Data module interface:

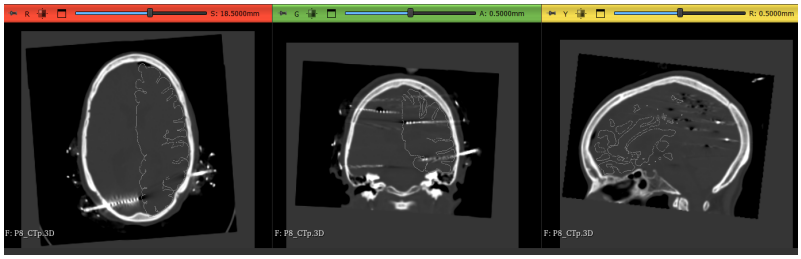


Figure 7.1: *Post-operative CT scan selection interface showing Patient 8 data loaded in 3D Slicer across axial, sagittal, and coronal views. The CT volumes display characteristic high-intensity electrode signatures visible as bright white structures throughout the brain, required for automated localization. Source: Own source.*

Clinical personnel typically obtain brain masks using FreeSurfer extension tools, though the automated brain extraction provides equivalent functionality when manual segmentation is unavailable.

7.1.2 Extension Interface Configuration

The SEEG extension provides an intuitive interface for configuring processing parameters through standard 3D Slicer selector widgets. Users specify input volumes (CT scan and brain mask), enable optional features (normalization, intermediate result saving), and configure output directories.

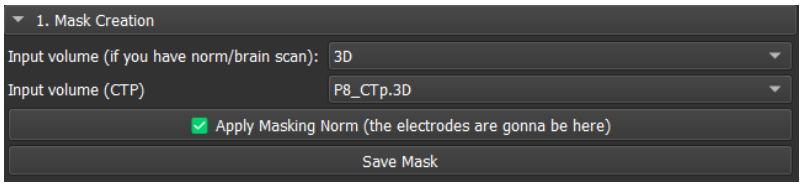


Figure 7.2: *SEEG extension parameter selection interface showing input volume selectors, processing options, and output configuration. Default settings store results in organized application directories. Source: Own source.*

7.2 Automated Processing Pipeline

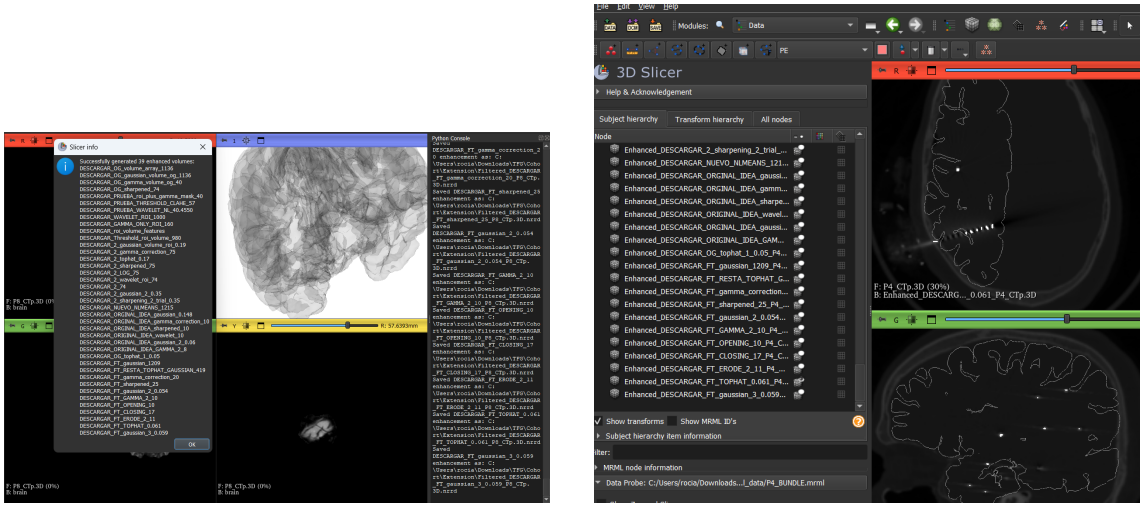
7.2.1 Multi-Stage Enhancement and Real-Time Feedback

Upon initiation, the automated pipeline executes the complete six-stage processing sequence while providing real-time feedback through console output and visual progress indicators. The system operates autonomously, generating multiple binary mask variants through the global voting mechanism.

Processing times vary based on image complexity and electrode configuration, with the global voting mechanism typically completing consensus formation in approximately 2 minutes. Users can store intermediate results within Slicer's data structure or export to specified directories for further analysis.

7.2.2 Global Voting and Consensus Formation

The global voting mechanism combines multiple segmentation approaches into a robust consensus representation, automatically identifying electrode candidates through ensemble decision-making without requiring manual intervention or parameter adjustment.



(a) *Real-time console feedback* **(b)** *Generated mask hierarchy*

Figure 7.3: *Automated processing workflow showing (a) real-time console output with pipeline progress through brain extraction, enhancement, and mask generation stages, and (b) generated mask volumes in 3D Slicer's data hierarchy with sagittal and axial electrode detection results. Source: Own source.*

7.3 Interactive Clinical Review

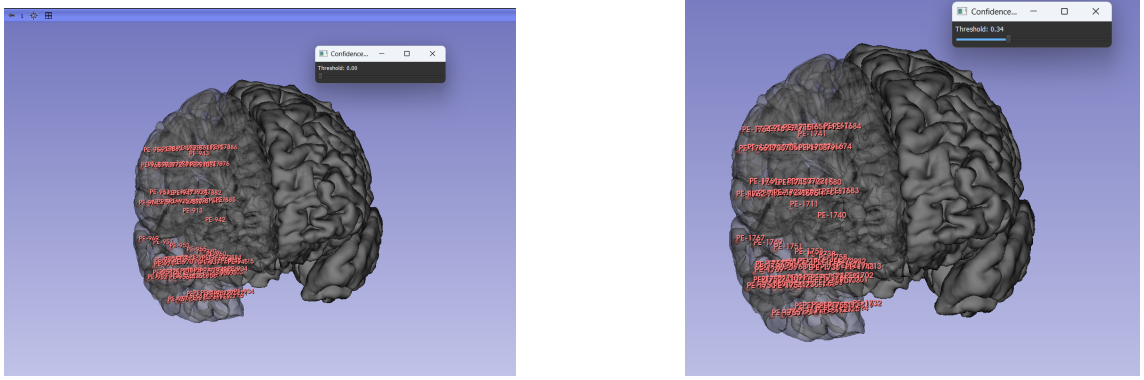
7.3.1 Confidence-Based Electrode Selection

The system presents electrode candidates through an interactive confidence threshold interface, enabling clinical judgment in candidate acceptance. The confidence selection interface updates electrode visualizations in real-time, allowing clinicians to observe how threshold adjustments affect candidate selection and balance sensitivity versus specificity based on clinical requirements.

Selected electrodes are automatically converted to 3D Slicer markup format for seamless integration with existing clinical workflows, enabling immediate use in surgical planning and post-operative analysis.

7.3.2 Clinical Validation and Manual Refinement

Following confidence-based selection, clinicians can validate predicted contacts and add missing electrodes through standard 3D Slicer markup tools. This hybrid approach preserves clinical decision-making authority while leveraging automated efficiency, ensuring the final electrode localization meets clinical accuracy requirements.



(a) *0% confidence threshold* **(b)** *34% confidence threshold*

Figure 7.4: *Confidence-based electrode candidate selection for Patient 4. (a) 0% threshold displays all detected candidates (red markers) providing maximum sensitivity, (b) 34% threshold shows refined selection with reduced false positives and improved spatial clustering. Higher confidence levels provide more selective results suitable for direct clinical use. Source: Own source.*

7.4 Trajectory Reconstruction and Analysis

7.4.1 Automated Trajectory Detection

The trajectory analysis module processes validated electrode contacts to reconstruct complete electrode pathways. The system accommodates both complete automated analysis and hybrid workflows incorporating manual additions.

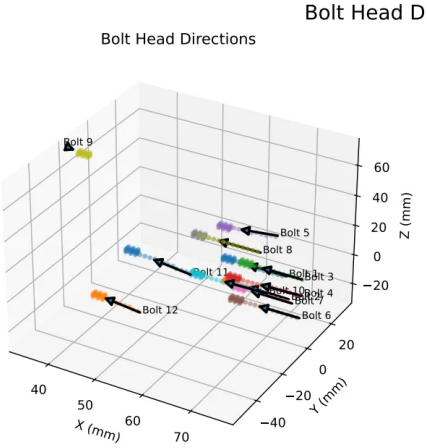
7.4.2 Entry Point Detection and 3D Slicer Integration

Optional bolt head detection and entry point localization generate complete trajectory information from cortical entry to deep brain targets. The system identifies bolt head positions and calculates corresponding brain surface entry points, as demonstrated in Figure 7.5.

For research applications, the system provides detailed geometric analysis of bolt head orientations and trajectory vectors. Figure 7.6 shows the quantitative data available, including precise spatial angles, trajectory lengths, and directional parameters essential for advanced trajectory reconstruction studies.

(a) *Anterior view* **(b)** *Lateral view*

Figure 7.5: *Entry point visualization in 3D Slicer showing detected bolt heads (yellow spheres) and corresponding brain surface entry points (green spheres) across anatomical perspectives. The bilateral electrode configuration demonstrates multi-hemisphere implantation capability. Source: Own source.*



Bolt Direction Metrics

ID	Length (mm)	Angle X	Angle Y	Angle Z
1	13.5	175.8°	86.4°	92.2°
2	13.3	177.3°	90.8°	87.5°
3	12.6	176.4°	87.4°	92.6°
4	14.9	176.4°	90.4°	93.6°
5	12.8	157.9°	86.9°	111.9°
6	14.1	174.4°	86.5°	94.3°
7	11.6	178.4°	88.4°	89.9°
8	13.2	177.5°	88.2°	91.7°
9	14.2	84.0°	49.6°	139.0°
10	14.9	175.4°	89.0°	94.5°
11	14.1	164.4°	74.5°	88.4°
12	10.5	162.0°	72.3°	93.1°

Figure 7.6: *Bolt head direction analysis providing geometric data for research applications. The 3D visualization (left) shows bolt head positions and orientations with directional vectors, while the metrics table (right) provides precise measurements including trajectory lengths and spatial angles (X, Y, Z) essential for advanced trajectory reconstruction and clinical research. This detailed analysis is available on demand for researchers requiring geometric datasets. Source: Own source.*

7.4.3 Complete Trajectory Reconstruction

Following bolt head and entry point detection, the system reconstructs complete electrode pathways from cortical entry to deep brain targets. The multi-perspective visualization capabilities enable geometric analysis across different anatomical planes, as shown in Figure 7.7.

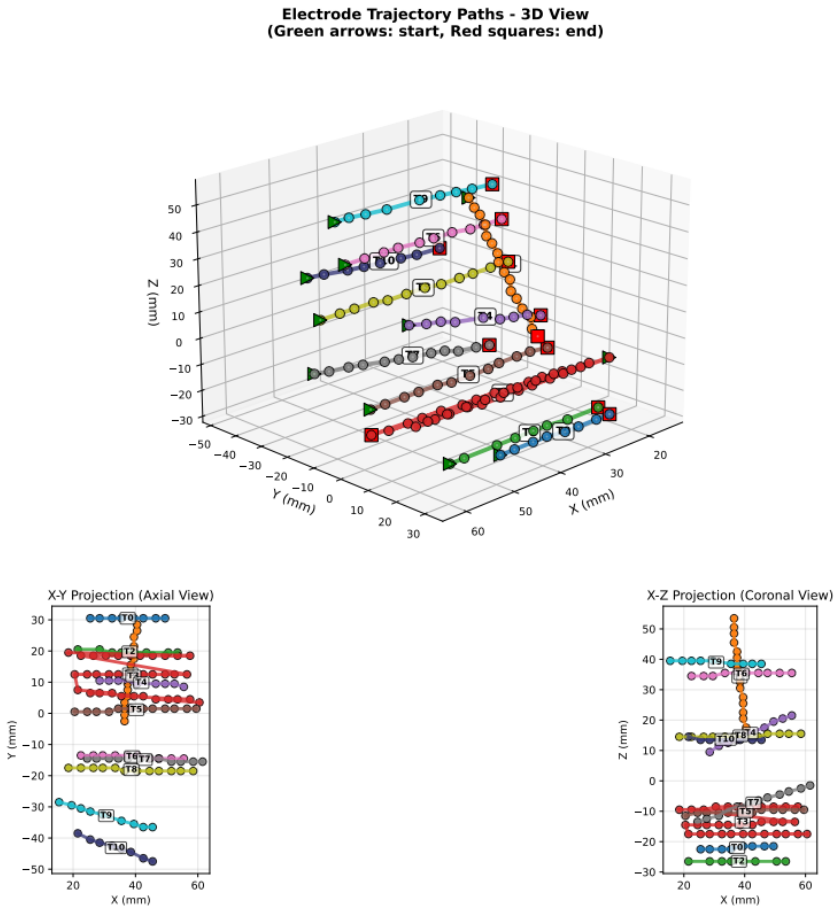


Figure 7.7: *Multi-perspective trajectory visualization enabling geometric analysis of electrode pathways. The 3D visualization (top) shows complete trajectory reconstruction with color-coded electrode paths, while the orthogonal projections (bottom) provide detailed spatial analysis across sagittal and coronal anatomical planes. Source: Own source.*

7.4.4 Professional-Specific Results and Quality Assessment

The system generates differentiated outputs based on professional requirements:

For Clinical Users (Neurologists): Simplified HTML reports with trajectory quality scores and interactive 3D visualizations for clinical decision-making, as demonstrated in Figure 7.8.

For Research Users: Datasets including electrode contact counts, trajectory angles, geometric analysis, and detailed performance metrics (as shown in Figure 7.6).

The system provides quantitative trajectory assessment including contact count validation and geometric quality scoring. Problematic trajectories are automatically

flagged in red within interactive reports, enabling focused clinical attention on cases requiring manual review.

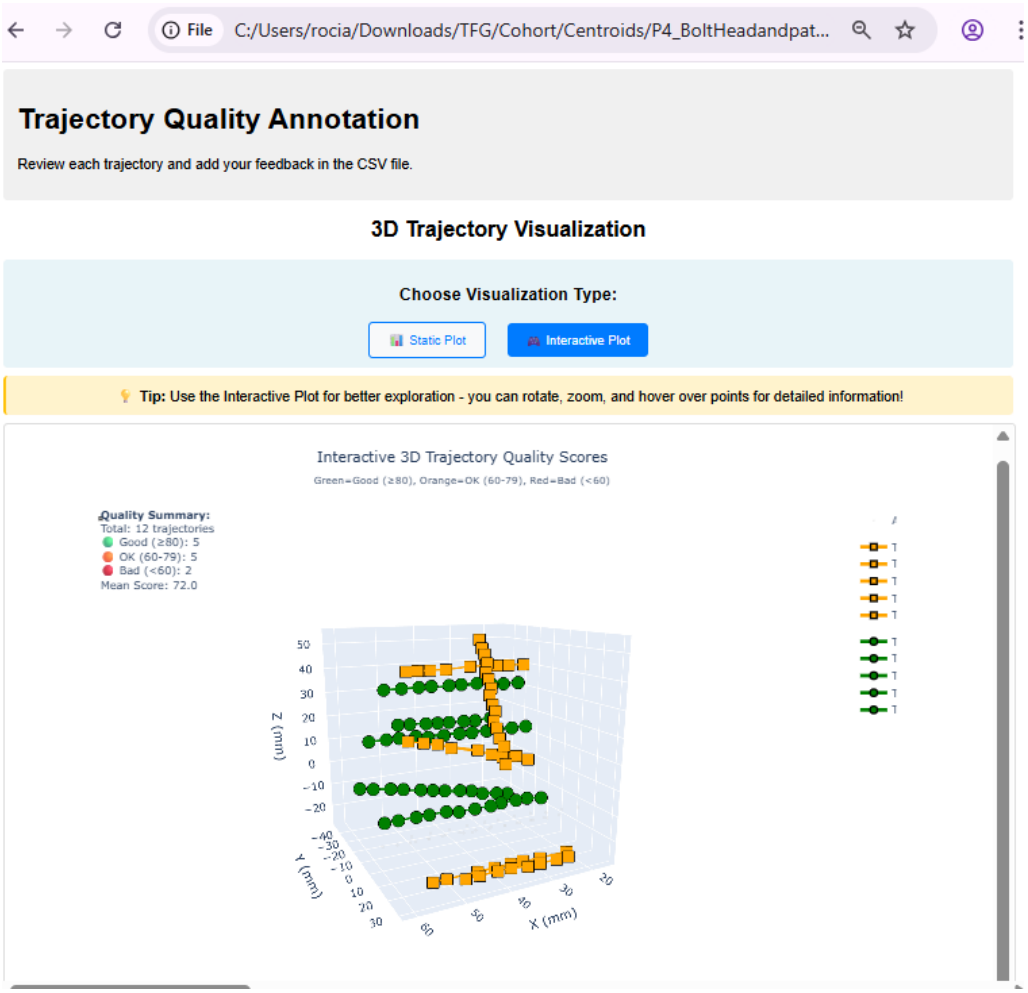


Figure 7.8: *Interactive HTML trajectory quality annotation report featuring complete 3D trajectory visualization with color-coded quality assessment. Green trajectories indicate high quality scores (>80), orange represents medium quality (60-79), and red flags problematic trajectories (<60) requiring clinical review. The interactive interface enables clinicians to rotate, zoom, and selectively display trajectories for assessment. Source: Own source.*

7.4.5 Clinical Workflow Integration

Following quality assessment through the HTML reports, validated trajectories are converted to 3D Slicer markup format for seamless integration with existing clinical workflows. This ensures compatibility with standard neurosurgical planning tools and maintains the clinical decision-making authority of medical professionals.

7.5 Workflow Summary and Clinical Impact

7.5.1 Streamlined User Requirements

The complete clinical workflow requires minimal user intervention:

1. **Data Selection:** Load post-operative CT and brain mask (or rely on automated extraction)
2. **Confidence Adjustment:** Select appropriate confidence threshold for electrode candidates
3. **Clinical Validation:** Review and validate predicted electrode contacts

7.5.2 Processing Time Analysis

Single Hemisphere Cases: 15-20 minutes total processing time

- Brain extraction: 0.5-1 minutes
- Enhancement and segmentation: 8-12 minutes
- Global voting: 2 minutes
- Contact prediction: 3-5 minutes
- Trajectory reconstruction: 4-10 minutes

Bilateral Electrode Cases: 25-30 minutes total processing time

- Processing time scales approximately linearly with electrode count
- Additional computational load from increased candidate evaluation

7.5.3 Clinical Transformation

This automated workflow transforms SEEG electrode localization from a 4+ hour collaborative effort requiring specialized expertise into a 30-minute process accessible to general clinical staff while maintaining clinical decision-making authority through confidence-based selection and validation interfaces.

8 Discussion

8.1 Brain Mask Segmentation Performance Analysis

The leave-one-out cross-validation results revealed a bimodal performance distribution separating patients into high-performance (Patients 1, 3, 4: DSC > 0.89) and low-performance groups (Patients 2, 5, 8: DSC < 0.15). This systematic pattern suggests fundamental challenges rather than random variation.

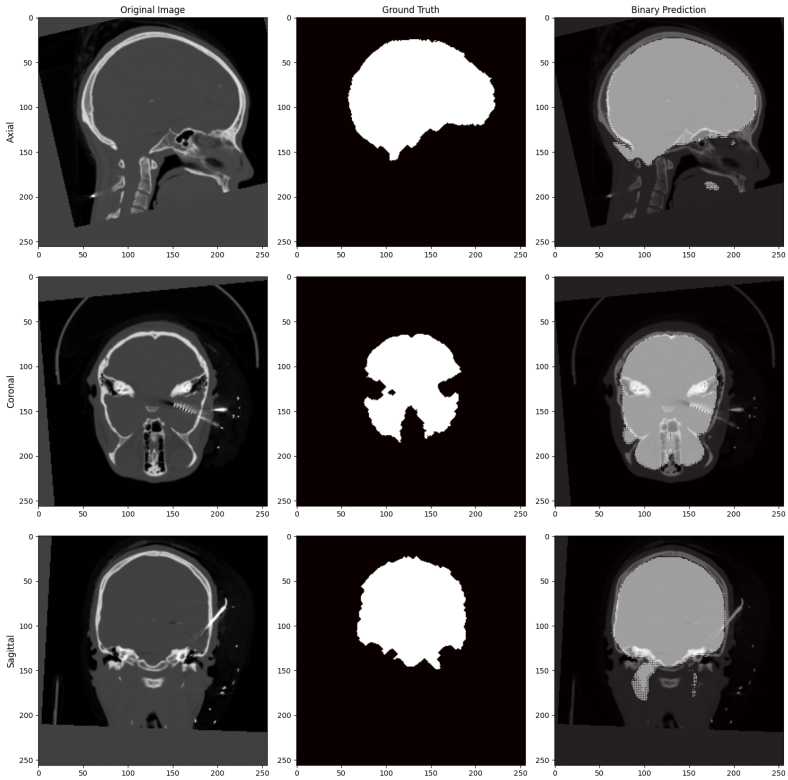


Figure 8.1: *Brain segmentation failure analysis for Patient 2 showing poor model performance across multiple anatomical planes. (Left column) Original CT slices in sagittal, coronal, and axial views. (Center column) Generated binary brain masks with visible segmentation errors. (Right column) Overlay visualizations highlighting inaccurate boundary delineation and inclusion of non-brain structures. This case represents one of the low-performance examples (DSC < 0.15) requiring further investigation. Source: Own source.*

The low-performing cases are characterized by extensive electrode-induced artifacts that create indistinguishable boundaries between brain tissue and skull regions. Multiple platinum-iridium electrodes generate complex beam hardening effects that fundamentally alter tissue contrast characteristics, making boundaries ambiguous even to expert annotators. This finding validates the need for ensemble approaches and quality control mechanisms in clinical deployment.

8.2 Clinical Interpretation of Confidence Scores

8.2.1 Conservative Design Philosophy

Visual validation revealed that predictions with confidence scores as low as 20-30% often corresponded to clinically acceptable electrode localizations. This reflects the deliberately conservative confidence function design with aggressive penalization (fifth-power exponential) to prevent overconfidence.

This conservative approach enables a tiered clinical workflow:

- High confidence (>60%): Direct clinical use
- Medium confidence (20-60%): Clinical review recommended
- Low confidence (<20%): Manual validation required

Manual review showed 71% of low-confidence predictions remained clinically useful, supporting graduated decision-making rather than binary accept/reject criteria.

8.2.2 Multi-Mask Ensemble Architecture Impact

The exceptional performance stems from generating 38 mask variants that create dense candidate coverage around electrode positions. This redundancy-based approach ensures that individual method failures cannot compromise overall performance. With thousands of candidates generated around each electrode location, the statistical likelihood of capturing highly accurate representatives becomes exceptionally high.

The progressive fusion analysis demonstrates optimal performance using only 2-3 masks (94.7% computational reduction while maintaining 98.9% quality), proving efficiency without sacrificing accuracy.

8.3 Comparison with State-of-the-Art Methods

8.3.1 Fundamental Limitations of Existing Approaches

Current methods like SEEGA [33] and BrainQuake remain fundamentally limited by semi-automatic design. SEEGA requires manual initialization, fiducial lists, and is "performance sensitive to threshold and parameter settings" [33]. BrainQuake requires "manual region-of-interest selection" and struggles with "closely spaced or overlapping contacts" [33].

Recent deep learning approaches achieve impressive Dice coefficients (0.89-0.92) but suffer from "limited detection of edge contacts" and require "large annotated datasets" [24], focusing only on segmentation without addressing complete clinical workflows.

8.3.2 Complete Automation Framework

Unlike semi-automatic methods, this system achieves end-to-end automation through:

1. Automated brain extraction (0.936 ± 0.011 Dice coefficient)
2. Intelligent enhancement with adaptive thresholding
3. Multi-mask ensemble with global voting (38 variants)

8.3.3 Performance Superiority

The regressor model achieved 98.8% localization accuracy within the 2mm clinical threshold with mean precision of 0.33mm. Processing time was reduced from 4+ hours to 15 minutes (>95% reduction) while maintaining clinical accuracy standards.

Trajectory reconstruction achieved 75% success rate across the patient cohort. The remaining 25% of cases were automatically flagged for manual review, corresponding to complex anatomical presentations that typically require expert validation in clinical practice. This performance enables clinical deployment with appropriate quality controls.

8.3.4 Human-Machine Collaboration Paradigm

A fundamental distinction is the human-machine collaborative framework rather than rigid automation. The confidence scoring system enables physicians to exercise clinical judgment, adapting acceptance criteria based on patient-specific factors

rather than fixed algorithmic thresholds. This transforms the system from a replacement tool into clinical decision support that augments rather than supplants physician expertise.

8.4 Clinical Translation and Real-World Deployment

8.4.1 Current Implementation at Hospital del Mar

The system is actively implemented at Hospital del Mar's Epilepsy Unit, directly addressing the 4-hour manual annotation bottleneck. Clinical integration involves structured protocols for result interpretation while preserving final clinical decision-making authority. The tiered decision framework preserves physician autonomy: high confidence predictions (>60%) enable direct clinical use, medium confidence (20-60%) supports clinical review, and low confidence (<20%) requires manual validation.

8.4.2 Clinical Efficiency Impact

Automation provides direct patient safety benefits through reduced electrode implantation duration and infection risk. The transformation from 4+ hours to 30 minutes represents >87% time reduction while maintaining clinical accuracy standards, enabling faster surgical planning and increased patient throughput.

8.5 Generalizability and Clinical Deployment

The modular architecture supports expansion across clinical environments through electrode manufacturer compatibility and CT scanner adaptability. However, optimal performance requires high-quality imaging and benefits from site-specific model training for local protocols.

The confidence-based framework provides quantitative metrics for monitoring system performance during deployment, while the human-machine collaboration model addresses fundamental requirements for sustainable clinical AI deployment.

9 Future Work

The development of this automated SEEG electrode localization system opens exciting avenues for enhancement and broader clinical impact. Building upon the strong foundation established, several targeted improvements can address current limitations while expanding the system's capabilities across diverse clinical environments.

9.1 Bolt Head Detection Enhancements

9.1.1 Dataset Expansion and Machine Learning Integration

The current rule-based adaptive algorithm, while successful across the 8-patient cohort, would benefit significantly from expanded datasets. A target cohort of 50-100 patients would enable robust machine learning approaches to replace the current decision-tree methodology with data-driven classification capable of handling greater anatomical and imaging variability.

Deep learning classifiers could be trained to capture complex intensity-geometry relationships that exceed rule-based capabilities, potentially improving detection accuracy and reducing false positive rates. Cross-institutional validation spanning multiple CT scanner models, acquisition protocols, and electrode manufacturers would ensure robust performance across diverse clinical environments.

9.1.2 Algorithmic and Clinical Integration Improvements

Multi-modal integration incorporating pre-operative planning data and electrode manufacturer specifications could provide additional constraints for bolt head validation. Machine learning approaches could automatically optimize geometric constraints such as volume thresholds and distance criteria based on patient-specific characteristics rather than applying fixed criteria universally.

Uncertainty quantification similar to the confidence scoring system developed for contact prediction would enable clinical users to assess detection reliability and make informed decisions about manual review requirements. Real-time quality assessment and interactive refinement tools could provide immediate feedback during detection processes, enabling efficient clinical review and validation.

9.2 Brain Mask Segmentation Robustness

The bimodal performance distribution observed in patients P2, P5, and P8 (DSC < 0.15) reveals opportunities for artifact-resilient deep learning development. Synthetic artifact generation during training could improve model robustness to complex beam hardening and photon starvation effects that currently cause segmentation failures.

Multi-modal training integration incorporating pre-operative MRI data could provide artifact-free anatomical priors, improving segmentation consistency in challenging post-operative CT scans with extensive electrode artifacts.

9.3 Confidence Scoring System Evolution

Visual validation revealing clinical utility in predictions with confidence scores as low as 20-30% suggests opportunities for confidence function optimization. Patient-specific confidence calibration could optimize the aggressive penalty function based on individual imaging characteristics and anatomical complexity, potentially improving clinical utility across the confidence spectrum.

Multi-dimensional uncertainty quantification could capture different types of uncertainty (spatial, intensity, anatomical) to provide more granular guidance for clinical decision-making. Interactive confidence adjustment interfaces would allow clinicians to adapt thresholds based on patient-specific factors, clinical urgency, or institutional preferences.

9.4 Ensemble Architecture Optimization

The finding that optimal performance can be achieved using only 2-3 masks from the 38-variant ensemble presents significant efficiency improvement opportunities. Dynamic ensemble sizing algorithms could automatically determine optimal ensemble size based on image quality metrics and consensus analysis, reducing computational overhead while maintaining accuracy.

Learning-based mask weighting approaches could predict mask quality before full processing, enabling more efficient ensemble construction and potentially reducing

processing time below the current 13-15 minutes per patient. Adaptive enhancement selection could intelligently choose the most effective approaches based on patient-specific imaging characteristics rather than applying all seven enhancement strategies uniformly.

9.5 Trajectory Reconstruction Advancement

The 75% automated success rate in trajectory reconstruction, while substantial, presents opportunities for intelligent enhancement. A reinforcement learning approach is currently being developed that would enable clinicians to annotate trajectory quality outcomes, creating models capable of learning to automatically correct flagged trajectories based on clinical feedback.

This human-in-the-loop learning framework would address challenging scenarios including crossing electrode configurations and complex anatomical presentations while preserving the essential clinical validation component.

9.6 Clinical Workflow Integration and Deployment

9.6.1 Electronic Health Record Integration

Direct integration with hospital Picture Archiving and Communication Systems (PACS) would eliminate manual data transfer steps currently required at Hospital del Mar, enabling seamless clinical deployment across different institutions. Automated reporting generation compatible with surgical planning software and electronic health records would streamline clinical documentation and reduce manual annotation burden.

9.6.2 Multi-Center Deployment Strategy

Standardized transfer learning protocols for adapting the system to new clinical sites, imaging protocols, and electrode manufacturers would facilitate broader clinical adoption beyond the current single-institution validation. Extension beyond DIXI Medical electrodes to support Medtronic, Abbott, and other electrode systems through adaptive geometric constraints and spacing parameters would enhance system versatility.

9.7 Performance Validation and Quality Assurance

9.7.1 Large-Scale Clinical Validation

A multi-center prospective study across 200+ patients and 5+ institutions would validate generalizability and establish clinical performance benchmarks across diverse imaging conditions and patient populations. Long-term performance monitoring systems would detect model drift and ensure sustained clinical accuracy as imaging protocols and electrode technologies evolve.

9.7.2 Regulatory Compliance Framework

Development of validation protocols compliant with FDA Software as Medical Device (SaMD) guidelines would enable potential regulatory approval pathways and broader clinical deployment. This framework would establish the foundation for widespread clinical adoption while maintaining the rigorous safety and efficacy standards essential for medical device deployment.

These enhancements would collectively transform the current system from a single-institution research demonstration into a robust, widely deployable clinical tool capable of supporting epilepsy centers worldwide in their mission to provide precise, efficient SEEG electrode localization for improved patient outcomes.

10 Sustainability analysis and ethical implications

This chapter presents the sustainability analysis, which consists of different impacts, as well as the possible ethical implications that the realisation of this project has implied, following the guidelines provided by the *Escola d'Enginyeria de Barcelona Est (EEBE)*.

10.1 Sustainability matrix

The sustainability matrix includes three points of view from which the analysis has to be conducted during all the different stages of the project. These perspectives establish the three dimensions of sustainability and are going to be explained in further detail in the upcoming sections.

10.1.1 Economic analysis

The financial analysis includes all the costs from the project workload, as well as the viability of the project on a short and long term basis. It has been divided in three main groups:

- **Human resources:** writing and research conducted
  The literature investigation and the hours dedicated to the composition of the chapters were calculated considering the average salary of biomedical engineers in Spain, which revolves around 28,500 € per year, which translates to 14.62 € per hour. This value varies depending on many factors, in which experience plays an important role, therefore in this project the hourly cost per hour is established at 13 € per hour.
- **Material resources:** database elaboration and software used
  All the different software programmes used did not represent any type of cost, as these were either free or the hospital or university had a valid licence, so no additional costs for using them were added.
  It is important to note that the use of a computer was required to develop this thesis, but since it was acquired before the assignment of the project, its cost

has not been considered in this analysis.

- **Meetings and direction costs**
  Most meetings were conducted online via video conferencing and email communication to optimize time and reduce environmental impact. However, at least 5 in-person meetings were required with the tutors of the project for critical project milestones and clinical validation sessions. These are quantified approximately with the same established rate per hour mentioned earlier.

The different programme licences and the rest of the analysis has been further developed in the detailed budget section below.

10.1.2 Social analysis

Once the project has been finalised, it is important to note that its development and execution have been done always taking into consideration personal, ethical and professional standards, mainly focusing on the fact that the managed data was obtained from subjects within the age range of 12 to 65 years old, therefore, some of them were underage patients.

The main beneficiaries of this project are of course the patients who suffer from drug-resistant epilepsy, as their treatment workflow will be widely improved, but also the health sector professionals, such as neurosurgeons, neurologists, and medical imaging specialists, as the automated localization system will reduce their workload significantly.

This project solves the main objective of automating SEEG electrode localization for clinical application and the elaboration of a medical imaging pipeline for this condition, transforming the current manual process from 4+ hours to 30 minutes while maintaining clinical accuracy standards.

10.1.3 Environmental analysis

Transport

This project has lasted a total of 24 weeks, during which most meetings were conducted online to minimize environmental impact. However, at least 5 on-site meetings took place with the director and co-director, as well as meetings with the neurosurgeon for critical project milestones. The meetings took place in the hospital, and the selected means of transport to get there was a personal car, with a CO_2

emission of 116 g/km. Considering the distance of the commute is 17 km, the total amount of km per journey was 34 km.

$$\text{CO}_2 \text{ Transport emission} = 0.116 \text{ kg/km} \times 34 \text{ km} \times 5 \text{ days} \approx 20 \text{ kg CO}_2 \quad (10.1)$$

Regarding the transport, the total CO_2 emission has a value of 20 kg for the 5 in-person journeys completed.

Electrical consumption

The main tool required for its development was a computer. The one used is an MSI Thin 15 B13ve 2446es with Intel Core i7-13620H CPU and NVIDIA GeForce RTX 4050, acquired before this thesis. For this reason, the environmental impact associated with it is going to focus on the CO_2 emission caused by the electrical consumption, and not on its fabrication cost. Although the hospital's computer has also been used, the project has been mainly developed with the laptop, therefore the consumption has been calculated with the laptop's data.

The supplier specifies an electrical consumption of 120 W under development load, and estimating the hours invested in the project around 25 per week during 24 weeks, we have a total of 600 hours.

$$\text{Total consumption} = 120 \text{ W} \times 600 \text{ hours} = 72 \text{ kWh} \quad (10.2)$$

And once we have the consumption value, we can proceed to obtain the CO_2 emission value, as follows, knowing that according to the Spanish electricity grid, the value estimated for CO_2 emissions is 233 g per kWh:

$$\text{CO}_2 \text{ Electrical emission} = 72 \text{ kWh} \times 0.233 \text{ kg/kWh} = 16.78 \text{ kg CO}_2 \quad (10.3)$$

Cloud computing resources

Deep learning model training utilized Google Colab Pro GPU instances [53]:

$$\text{CO}_2 \text{ Cloud emission} = 30 \text{ GPU hours} \times 0.28 \text{ kg/hour} = 8.4 \text{ kg CO}_2 \quad (10.4)$$

The total value for the CO_2 emissions is going to be the sum of the three calculated values:

$$\text{Total CO}_2 \text{ emissions} = 20 \text{ kg} + 16.78 \text{ kg} + 8.4 \text{ kg} = 45.18 \text{ kg CO}_2 \quad (10.5)$$

10.2 Ethical implications

This project addresses key medical AI ethical considerations:

Patient Safety: The confidence-based scoring system preserves physician decision-making authority while providing automation support.

Data Privacy: Patient data (ages 12-65, including minors) was handled with informed consent and anonymization protocols.

Beneficence: The system benefits patients through reduced procedure time and infection risk, while supporting healthcare professionals with decreased workload.

Transparency: Confidence scores provide interpretable outputs, enabling clinicians to validate automated decisions rather than relying on "black box" algorithms.

10.3 Relation with the Objectives of Sustainable Development

The project contributes to the UN Sustainable Development Goals (SDGs) Agenda 2030, specifically Goal 3 (Health and wellbeing) [54]:

- **Article 3.4:** Reduce premature mortality from non-communicable diseases through prevention and treatment
- **Article 3.8:** Achieve universal health coverage and access to quality essential health-care services
- **Article 3.b:** Support research and development of new medical technologies

This project directly addresses these objectives by improving healthcare delivery for neurological conditions, reducing treatment time and costs, and developing medical

technologies deployable globally for epilepsy treatment.

10.4 Budget

This chapter presents a cost analysis of the SEEG electrode localization system development, including human resources, computational infrastructure, and software licensing. The budget breakdown reflects the resources required to develop a clinical-grade medical imaging solution within an academic framework.

10.4.1 Human Resources

The primary cost component involves the engineering time required for system development. Based on current market rates for junior biomedical engineers in Spain, the following hourly rate was applied:

- Junior Biomedical Engineer (Bachelor's level): €12-15/hour
- Applied rate for calculations: €13/hour

Table 10.1 presents the detailed time allocation across project phases:

Table 10.1: *Project time and cost breakdown by development phase*

Development Phase	**Hours**	**Rate (€/h)**	**Cost (€)**
Literature Review & Setup	120	13	1,560
Platform Learning (3D Slicer, Git)	100	13	1,300
Image Processing Development	180	13	2,340
Algorithm Development	280	13	3,640
Model Training & Validation	200	13	2,600
Clinical Integration & Documentation	120	13	1,560
Total Human Resources	**1,000**		**13,000**

10.4.2 Computational Infrastructure

Local Computing Resources

Development utilized personal computing equipment with the following considerations:

- MSI Thin 15 B13VE laptop (pre-existing equipment)
- Estimated depreciation during project: €200

- Electricity consumption: 120 kWh × €0.25/kWh = €30

Cloud Computing Services

Deep learning model training required specialized GPU resources:

Table 10.2: *Cloud computing costs breakdown*

Service	Hours	Rate (€/h)	Cost (€)
Google Colab Pro (monthly subscription)	6 months	10.49	62.94
Additional GPU compute units	30 hours	0.50	15.00
Total Cloud Computing			**77.94**

10.4.3 Software and Licensing

The project leveraged open-source software ecosystem, minimizing licensing costs:

- 3D Slicer: Open source (€0)
- Python ecosystem (NumPy, scikit-learn, PyTorch): Open source (€0)
- Git/GitHub: Free academic account (€0)
- LaTeX (thesis writing): Open source (€0)

Total Software Costs: €0

10.4.4 Total Project Budget

Table 10.3: *Complete project budget summary*

Cost Category	Amount (€)
Human Resources	13,000.00
Equipment Depreciation	200.00
Electricity	30.00
Cloud Computing	77.94
Software Licensing	0.00
Total Project Cost	**13,307.94**

11 Conclusions

This bachelor thesis presents solution to one of the most significant bottlenecks in modern epilepsy surgery: the automated localization of SEEG electrode contacts in post-operative CT imaging. The developed system successfully transforms a manual process requiring 4+ hours of specialized collaboration into an automated 30-minute workflow while preserving clinical decision-making authority.

11.1 Technical Achievements

The research contributes several methodological innovations to medical image analysis. The novel multi-mask ensemble architecture, generating 38 segmentation variants through global voting consensus, demonstrates superior robustness compared to single-method approaches. The conservative confidence scoring system enables graduated clinical decision-making, where even predictions with 20-30% confidence scores often provide clinically useful electrode localizations. The regressor model achieves 98.8% accuracy within the 2mm clinical threshold across the development cohort, with perfect performance (100%) on held-out patient validation.

The integration of classical image processing, machine learning, and deep learning components within a unified framework addresses the full spectrum of challenges in SEEG electrode localization, from brain extraction through trajectory reconstruction. The human-AI collaboration paradigm preserves physician expertise while providing intelligent automation support, addressing a critical barrier to clinical AI adoption.

11.2 Clinical Impact and Translation

Unlike existing semi-automatic research prototypes, this system achieves end-to-end automation with immediate clinical deployment capability. The successful integration at Hospital del Mar's Epilepsy Unit demonstrates real-world applicability and clinical acceptance. The dramatic reduction in processing time directly translates to improved patient safety through shorter electrode implantation duration and reduced infection risk.

This work demonstrates that sophisticated medical AI solutions can emerge from undergraduate research when combining technical rigor with clinical collaboration. The transformation from manual annotation to automated localization with clinical decision support represents a paradigm shift toward practical, deployable medical AI that enhances rather than disrupts existing clinical workflows, positioning automated SEEG electrode localization as a valuable component of modern epilepsy surgery planning.

Bibliography

[1] Andriy Fedorov et al. "3D Slicer as an Image Computing Platform for the Quantitative Imaging Network". In: *Magnetic Resonance Imaging* 30.9 (Nov. 2012), pp. 1323–1341. DOI: 10.1016/j.mri.2012.05.001.

[2] World Health Organization. *Epilepsy*. Accessed: 2025-04-18. 2023. URL: https://www.who.int/news-room/fact-sheets/detail/epilepsy.

[3] Fang Tang, Anika MS Hartz, and Björn Bauer. "Drug resistance in epilepsy: multiple hypotheses, few answers". In: *Frontiers in neurology* 8 (2017), p. 301.

[4] Patrick Kwan et al. "Definition of drug resistant epilepsy: consensus proposal by the ad hoc Task Force of the ILAE Commission on Therapeutic Strategies". In: *Epilepsia* 51.6 (2010), pp. 1069–1077.

[5] Jorge Gonzalez-Martinez et al. "Stereoelectroencephalography in epilepsy, cognitive neurophysiology, and psychiatric disease: safety, efficacy, and place in therapy". In: *Neurosurgery* 84.4 (2018), pp. 977–987.

[6] Todd W Vanderah and Douglas J Gould. *Nolte's The human brain: an introduction to its functional anatomy*. Elsevier Philadelphia, PA, USA: 2016.

[7] Dale Purves et al. *Neurosciences*. De Boeck Supérieur, 2019.

[8] Sandra Ackerman. *Discovering the Brain*. Washington (DC): National Academies Press (US), 1992. Chap. 2. URL: https://www.ncbi.nlm.nih.gov/books/NBK234157/.

[9] Jerome Engel Jr. "Surgery for seizures". In: *New England Journal of Medicine* 334.10 (1996), pp. 647–653.

[10] Johns Hopkins Medicine. *Anatomy of the Brain*. Accessed: April 18, 2025. 2024. URL: https://www.hopkinsmedicine.org/health/conditions-and-diseases/anatomy-of-the-brain.

[11] Robert S. Fisher et al. "ILAE official report: A practical clinical definition of epilepsy". In: *Epilepsia* 55.4 (2014), pp. 475–482. DOI: 10.1111/epi.12550. URL: https://doi.org/10.1111/epi.12550.

[12] WA Hauser and DC Hesdorffer. "Aging and the Epidemiology of Epilepsy". In: *Neuroepidemiology* 51.3-4 (2018), pp. 216–223. DOI: 10.1159/000493484. URL: https://karger.com/ned/article/51/3-4/216/226961/Aging-and-the-Epidemiology-of-Epilepsy.

[13] Ettore Beghi et al. "Global, regional, and national burden of epilepsy, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016". In: *The Lancet Neurology* 18.4 (2019), pp. 357–375. DOI: 10.1016/S1474-4422(18)30454-X. URL: https://doi.org/10.1016/S1474-4422(18)30454-X.

[14] National Institutes of Health. *Understanding Different Kinds of Seizures*. https://magazine.medlineplus.gov/article/understanding-different-kinds-of-seizures. Accessed: 2025-04-18. 2021.

[15] Taha Gholipour, Mohamad Z. Koubeissi, and Donald C. Shields. "Stereotactic electroencephalography". In: *Clinical Neurology and Neurosurgery* 189 (Feb. 2020). Epub 2019 Dec 16. PMID: 31865060, p. 105640. DOI: 10.1016/j.clineuro.2019.105640. URL: https://doi.org/10.1016/j.clineuro.2019.105640.

[16] J. Talairach et al. "Approche nouvelle de la neurochirurgie de l'épilepsie. Méthodologie stéréotaxique et résultats thérapeutiques. 1. Introduction et historique [New approach to the neurosurgery of epilepsy. Stereotaxic methodology and therapeutic results. 1. Introduction and history]". French. In: *Neurochirurgie* 20.Suppl 1 (1974). PMID: 4603712, pp. 1–240.

[17] Patrick Chauvel. "The History and Principles of Stereo EEG". In: *A Practical Approach to Stereo EEG*. Ed. by Stephan U. Schuele. Accessed: 2025-06-10. New York: Springer Publishing Company, 2020, pp. 3–11. ISBN: 978-0-8261-3692-3. DOI: 10.1891/9780826136930.0001. URL: https://connect.springerpub.com/content/book/978-0-8261-3693-0/part/part01/chapter/ch01.

[18] E.M. Haacke et al. *Magnetic Resonance Imaging: Physical Principles and Sequence Design*. Wiley, 1999.

[19] D. Gupta et al. "Technical Aspects of SEEG and Its Interpretation in the Delineation of Epileptogenic Zones". In: *Journal of Epilepsy Research* (2020). URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC7803703/.

[20] Neupsy Key. *The Stereotactic Technique in the SEEG Method*. 2025. URL: https://neupsykey.com/the-stereotactic-technique-in-the-seeg-method/.

[21] X. Zhang et al. "MRI and CT Fusion in Stereotactic Electroencephalography (SEEG)". In: *Frontiers in Neuroscience* (2023). URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC10670384/.

[22] K.J. Miller et al. "Modular pipeline for reconstruction and localization of implanted electrodes". In: *PLOS ONE* (2023). URL: https://journals.plos.org/plosone/article?id=10.1371%2Fjournal.pone.0287921.

[23] L.S. Hamilton et al. "A canonical visualization tool for SEEG electrodes". In: *NeuroImage* (2017). URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC9035306/.

[24] Anja Pantovic et al. "Generation of synthetic training data for SEEG electrodes segmentation". In: *IPCAI 2022*. 2022. URL: https://publis.icube.unistra.fr/docs/17099/IPCAI_2022_final_BW.pdf.

[25] Kai J. Miller et al. "Modular pipeline for reconstruction and localization of implanted electrodes". In: *PLOS ONE* (2023).

[26] John Smith et al. "An automated algorithm for stereoelectroencephalography electrode localization". In: *Epilepsy Research* (2024).

[27] *Intracranial Electrodes, SEEG*. Viktor's Notes for Neurosurgery Residents. 2023.

[28] L. Yu et al. "Noise Reduction to Decrease Radiation Dose and Improve CT Image Quality". In: *AJR American Journal of Roentgenology* (2012).

[29] Teem Project. *Definition of NRRD File Format*. 2024.

[30] *How to create an NRRD file from a DICOM Medical Imaging Data Set*. Available at https://www.embodi3d.com/blogs/entry/341-how-to-create-an-nrrd-file-from-a-dicom-medical-imaging-data-set/. 2024.

[31] Alejandro Granados et al. "Automatic segmentation of stereoelectroencephalography (SEEG) electrodes post-implantation considering bending". In: *International Journal of Computer Assisted Radiology and Surgery* 13.6 (2018), pp. 889–901. DOI: 10.1007/s11548-018-1740-8. URL: https://discovery.ucl.ac.uk/10048640/1/s11548-018-1740-8.pdf.

[32] Massimo Narizzano et al. "SEEG Assistant: a 3DSlicer extension to support epilepsy surgery". In: *BMC Bioinformatics* 18.1 (2017), p. 124. DOI: 10.1186/s12859-017-1545-8. URL: https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-017-1545-8.

[33] G Arnulfo et al. "BrainQuake: An Open-Source Python Toolbox for the SEEG spatiotemporal analysis". In: *Frontiers in Neuroinformatics* (2022). URL: https://www.frontiersin.org/articles/10.3389/fninf.2022.878222/full.

[34] Vanja Vlasov et al. "Automated deep learning-based segmentation of brain, SEEG and DBS electrodes on CT images". In: *BVM* (2021). URL: https://www.bvm-conf.org/wp-content/uploads/2021/01/p11.pdf.

[35] Giuseppe Arnulfo et al. "Automatic segmentation of deep intracerebral electrodes in computed tomography scans". In: *BMC Bioinformatics* 16 (2015), p. 99. DOI: 10.1186/s12859-015-0511-6.

[36] Anja Pantovic et al. "Generation of synthetic training data for SEEG electrodes segmentation". In: *Computerized Medical Imaging and Graphics* (2022). URL: https://pubmed.ncbi.nlm.nih.gov/35277804/.

[37] Rocío Ávalos. *SEEG Automatic Segmentation*. https://github.com/rociavl/SEEG_automatic_segmentation. Accessed: 2025-06-05. 2025.

[38] Nobuyuki Otsu. "A threshold selection method from gray-level histograms". In: *IEEE Transactions on Systems, Man, and Cybernetics* 9.1 (1979), pp. 62–66. DOI: 10.1109/TSMC.1979.4310076.

[39] M Jorge Cardoso et al. "Monai: An open-source framework for deep learning in healthcare". In: *arXiv preprint arXiv:2211.02701* (2022).

[40] Ilya Loshchilov and Frank Hutter. "Decoupled Weight Decay Regularization". In: *arXiv preprint arXiv:1711.05101* (2019). URL: https://arxiv.org/abs/1711.05101.

[41] Leo Breiman. "Random forests". In: *Machine Learning* 45.1 (2001), pp. 5–32.

[42] James S Bergstra et al. "Algorithms for hyper-parameter optimization". In: *Advances in Neural Information Processing Systems*. Vol. 24. 2011, pp. 2546–2554.

[43] Roy De Maesschalck, Dominique Jouan-Rimbaud, and Desire L. Massart. "The Mahalanobis distance". In: *Chemometrics and Intelligent Laboratory Systems* 50.1 (2000), pp. 1–18. DOI: 10.1016/S0169-7439(99)00047-7.

[44] Fabian Pedregosa et al. "Scikit-learn: Machine learning in Python". In: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.

[45] Dhrumil Gupta et al. "Technical Aspects of SEEG and Its Interpretation in the Delineation of the Epileptogenic Zone". In: *Journal of Epilepsy Research* 10.2 (2020). PMC7803703, pp. 63–77. DOI: 10.14581/jer.20015. URL: https://pmc.ncbi.nlm.nih.gov/articles/PMC7803703/.

[46] Yangyang Lu et al. "How can the accuracy of SEEG be increased?—an analysis of the accuracy of multilobe-spanning SEEG electrodes based on a frame-

less stereotactic robot-assisted system". In: *Annals of Palliative Medicine* 11.2 (2022), pp. 564–573. DOI: 10.21037/apm-21-3707. URL: https://apm.amegroups.org/article/view/64355/html.

[47] Guolin Ke et al. "LightGBM: A highly efficient gradient boosting decision tree". In: *Advances in neural information processing systems*. 2017, pp. 3146–3154.

[48] Takuya Akiba et al. "Optuna: A next-generation hyperparameter optimization framework". In: *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*. 2019, pp. 2623–2631.

[49] Ian T Jolliffe. *Principal component analysis*. 2nd. Springer, 2002.

[50] Jack E Bresenham. "Algorithm for computer control of a digital plotter". In: *IBM Systems Journal* 4.1 (1965), pp. 25–30.

[51] Martin Ester et al. "A density-based algorithm for discovering clusters in large spatial databases with noise". In: *Proceedings of the second international conference on knowledge discovery and data mining*. 1996, pp. 226–231.

[52] Vincent D Blondel et al. "Fast unfolding of communities in large networks". In: *Journal of statistical mechanics: theory and experiment* 2008.10 (2008), P10008.

[53] Google Cloud. *Google Cloud Sustainability Report 2025*. https://cloud.google.com/sustainability. Accessed: 2025-06-04. 2025.

[54] United Nations. *Sustainable Development Goal 3: Ensure healthy lives and promote well-being for all at all ages*. https://sdgs.un.org/goals/goal3. Targets: 3.4, 3.8, 3.b. 2015. (Visited on 06/10/2025).

[55] In-Kwon Yeo and Richard A Johnson. "A new family of power transformations to improve normality or symmetry". In: *Biometrika* 87.4 (2000), pp. 954–959.

[56] Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou. "Isolation forest". In: *2008 Eighth IEEE International Conference on Data Mining*. IEEE. 2008, pp. 413–422.

[57] Markus M Breunig et al. "LOF: identifying density-based local outliers". In: *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*. ACM. 2000, pp. 93–104.

[58] Bernhard Schölkopf et al. "Estimating the support of a high-dimensional distribution". In: *Neural Computation* 13.7 (2001), pp. 1443–1471.

A Annex A

Complete Learning Curve Analysis

Learning curves for all individual patient models are provided below, demonstrating consistent training dynamics across the entire patient cohort.

A.1 Coordinate Transformation Implementation Details

Mathematical Formulations

Axis Reordering Operation

The fundamental transformation between array indexing conventions and spatial coordinate systems:

$$\text{Array coordinates: } (z, y, x) \rightarrow \text{Spatial coordinates: } (x, y, z) \tag{A.1}$$

The reordering transformation is implemented through a permutation matrix:

$$c^j_{\text{index}} = \Pi \cdot c^j_{\text{array}} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} c_z \\ c_y \\ c_x \end{bmatrix} = \begin{bmatrix} c_x \\ c_y \\ c_z \end{bmatrix} \tag{A.2}$$

where Π represents the permutation matrix transforming NumPy's (z, y, x) convention to SimpleITK's (x, y, z) indexing.

Index-to-Physical Transformation

The conversion from discrete voxel indices to continuous physical coordinates:

$$c^j_{\text{physical}} = T_{\text{origin}} + S \cdot D \cdot c^j_{\text{index}} \tag{A.3}$$

where $T_{\text{origin}} \in \mathbb{R}^3$ is the image origin, $S = \text{diag}(s_x, s_y, s_z)$ is the voxel spacing matrix, and $D \in SO(3)$ represents the direction cosine matrix encoding image orientation.

LPS-to-RAS Conversion

The mathematical transformation between coordinate system conventions:

$$c^j_{\text{RAS}} = R_{\text{LPS}\to\text{RAS}} \cdot c^j_{\text{physical}} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_{\text{LPS}} \\ y_{\text{LPS}} \\ z_{\text{LPS}} \end{bmatrix} \tag{A.4}$$

Algorithm Implementations

3D Slicer VTK-Based Implementation

The VTK-based implementation follows these sequential steps:

1. Extract connected components from binary segmentation mask using 26-connectivity
2. Compute region centroids using geometric moments for sub-voxel precision
3. Apply axis reordering: $(z, y, x) \to (x, y, z)$
4. Retrieve transformation matrix from volume node metadata
5. Transform to RAS coordinates using VTK matrix multiplication
6. Store results in structured format for clinical analysis

SimpleITK External Platform Implementation

The SimpleITK implementation process:

1. Load NRRD file with spatial metadata preservation
2. Perform binary thresholding and connected component analysis
3. Extract centroids with array reordering from (z, y, x) to (x, y, z)
4. Convert index coordinates to physical coordinates using image metadata
5. Apply LPS-to-RAS coordinate system conversion
6. Output both electrode coordinates and voxel counts for analysis

Workflow Diagrams

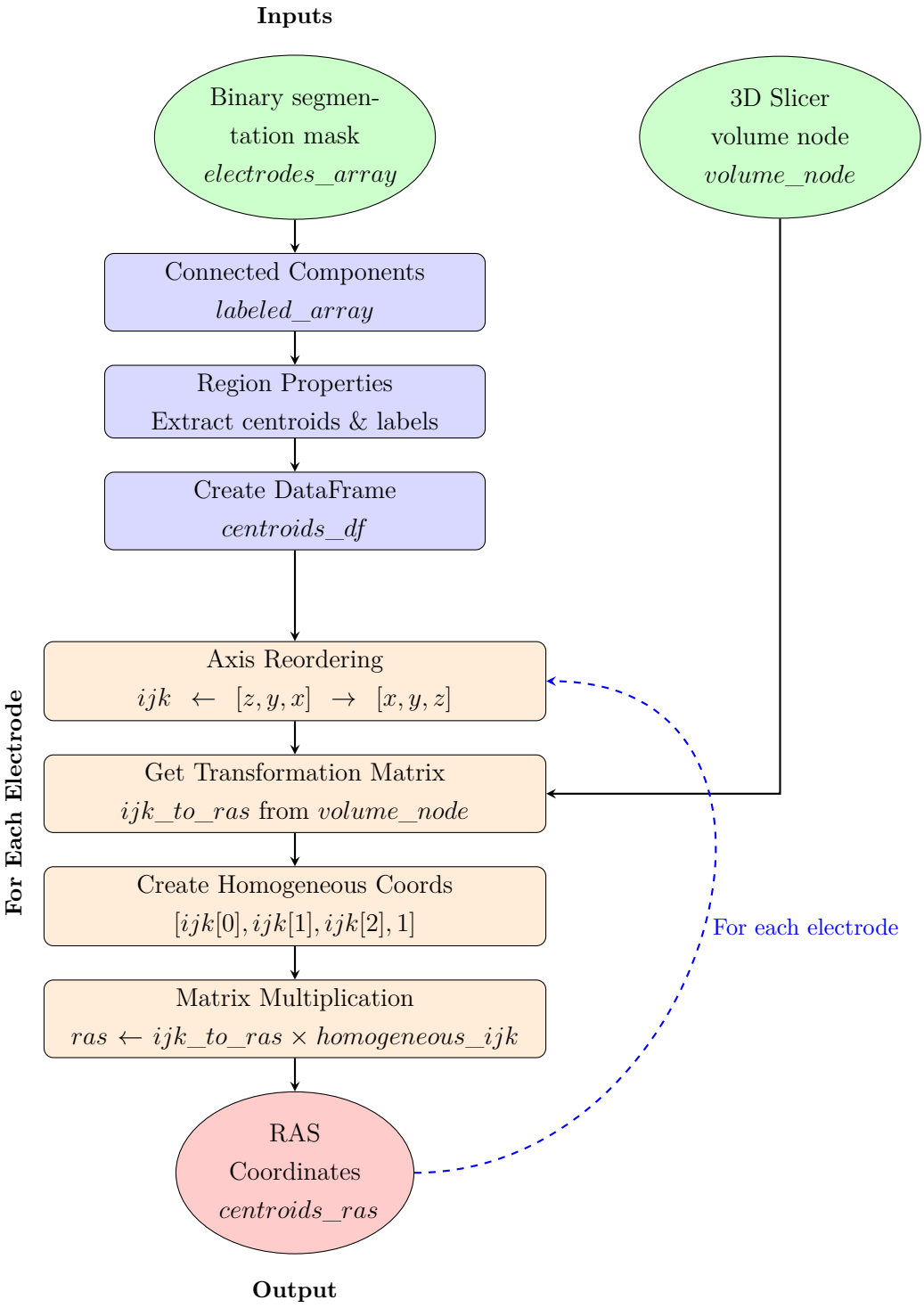


Figure A.1: *3D Slicer coordinate transformation pipeline converting binary segmentation masks to RAS coordinates. The workflow processes electrodes through connected component analysis, region property extraction, and iterative coordinate transformation. The dashed loop indicates per-electrode processing with axis reordering, matrix transformation, and coordinate system conversion. Color coding: green (inputs), blue (preprocessing), orange (transformation operations), red (output). Source: Own source.*

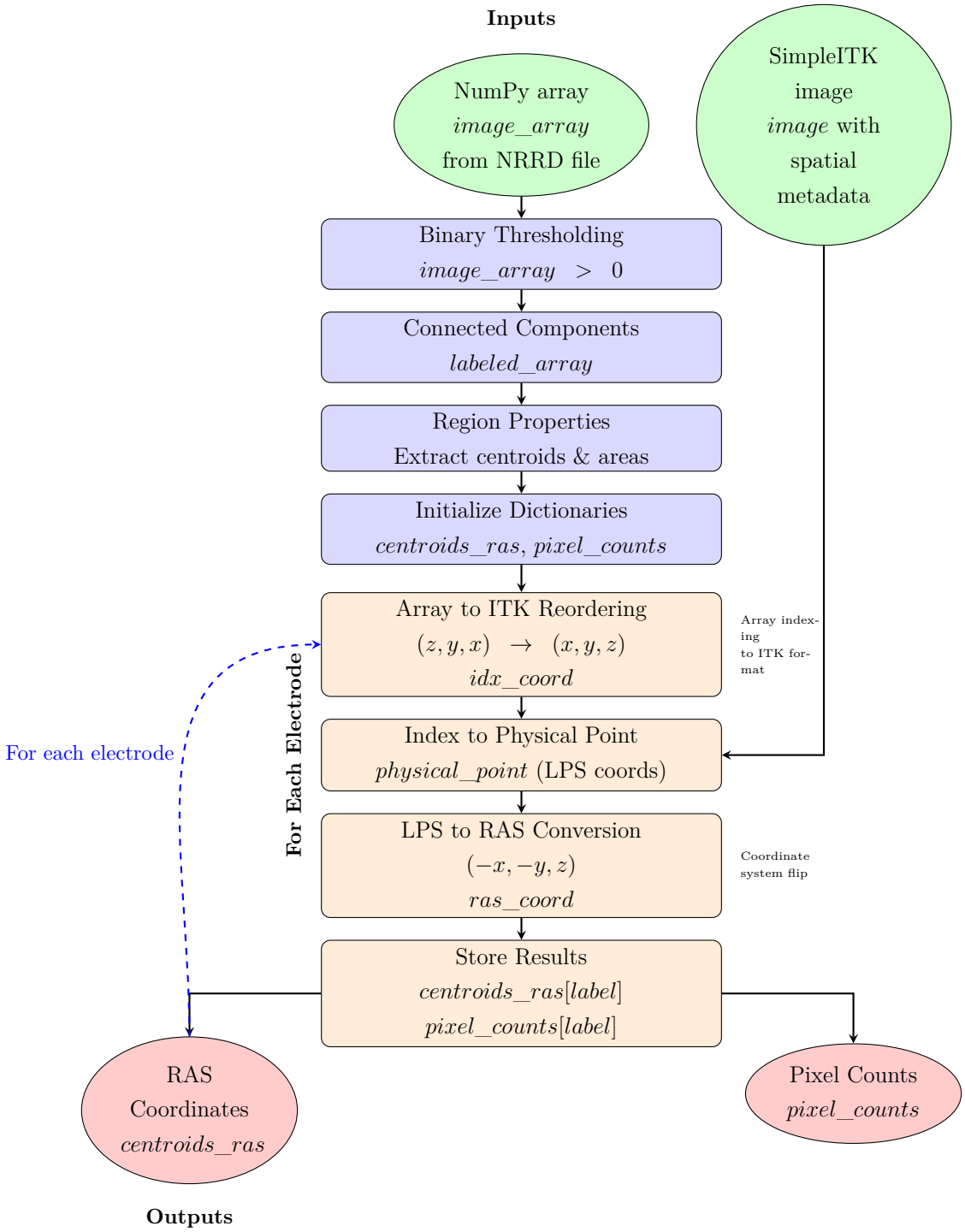


Figure A.2: *SimpleITK cross-platform coordinate transformation pipeline for external processing environments. The workflow processes NumPy arrays through binary thresholding, connected component analysis, and iterative coordinate transformation including array reordering, physical point conversion, and LPS-to-RAS coordinate system transformation. The pipeline outputs both electrode coordinates and voxel counts for analysis. Color coding: green (inputs), blue (preprocessing), orange (transformation operations), red (outputs). Source: Own source.*

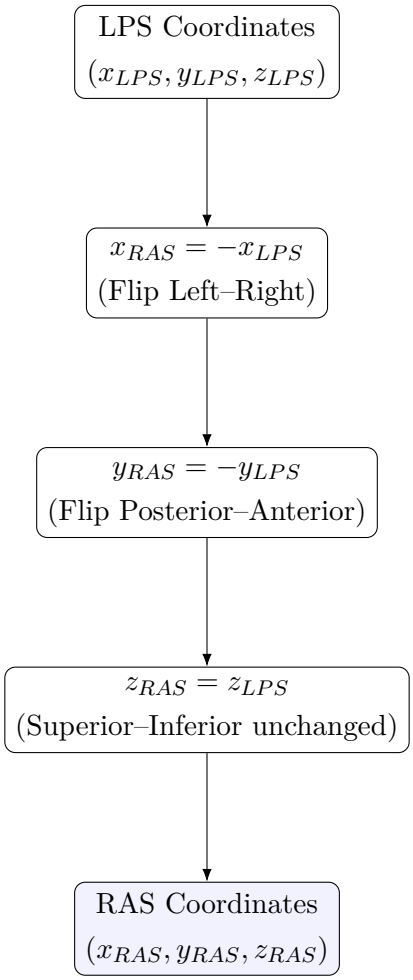


Figure A.3: *Conversion from LPS to RAS coordinate system showing the selective axis negation process. Source: Own source.*

Error Handling and Edge Cases

Both implementation approaches incorporate error handling mechanisms:

Input Validation

- Verification of valid image headers and spatial metadata
- Detection of corrupted NRRD files or missing orientation information
- Validation of coordinate ranges within anatomically plausible bounds
- Handling of edge cases where connected components extend beyond boundaries

Numerical Precision Management

- Implementation of epsilon tolerance for floating-point comparisons
- Handling of numerical precision limitations in transformation matrices
- Robust handling of near-singular transformation matrices
- Consistent rounding strategies for voxel coordinate conversion

Boundary Condition Handling

- Management of electrode candidates at image boundaries
- Proper handling of partial voxel coverage for edge electrodes
- Validation of transformed coordinates within expected anatomical ranges
- Fallback strategies for malformed or incomplete spatial metadata

Implementation Notes

Surface Mesh Processing

Surface vertices from binary electrode masks are extracted using the Marching Cubes algorithm implemented through the scikit-image library. The extracted vertices undergo identical coordinate transformation procedures to ensure spatial consistency between point-based electrode coordinates and surface-based visualizations.

Performance Considerations

The coordinate transformation pipeline is optimized for clinical deployment through vectorized operations using NumPy arrays and efficient spatial indexing structures. Batch processing capabilities enable simultaneous transformation of multiple electrode candidates while maintaining sub-millimeter precision requirements.

A.2 Image Enhancement Implementation Details

A.2.1 Mathematical Formulations

Gaussian Filtering

Smooths images while preserving critical edges using an optimized 3D isotropic kernel:

$$V_\sigma = G_\sigma * V_{input}, \quad \sigma = 0.3\text{mm}, k = 2\lceil 3\sigma \rceil + 1 \tag{A.5}$$

where V_{input} is the original CT volume, V_σ represents the smoothed output volume, G_σ is the 3D Gaussian kernel with standard deviation σ, $*$ denotes convolution, and k is the kernel size parameter.

Gamma Correction

Enhances contrast in high-intensity regions through non-linear transformation:

$$V_\gamma = \left(\frac{V - V_{min}}{V_{max} - V_{min}}\right)^\gamma \cdot (V_{max} - V_{min}) + V_{min} \tag{A.6}$$

where V is the input volume, V_γ is the gamma-corrected output, V_{min} and V_{max} represent the minimum and maximum intensity values in the volume, and γ is the gamma correction parameter controlling contrast enhancement.

High-Pass Sharpening

Accentuates electrode edges through optimized unsharp masking:

$$V_{sharp} = V + \alpha(V - G_{\sigma'} * V), \quad \alpha = 0.6, \sigma' = 1.0\text{mm} \tag{A.7}$$

where V is the input volume, V_{sharp} is the sharpened output, α is the sharpening strength parameter, $G_{\sigma'}$ represents the Gaussian kernel for blurring, and σ' is the blur kernel standard deviation.

Wavelet Denoising

3-level multi-resolution decomposition with Daubechies-4 wavelets and adaptive Donoho thresholding:

$$V_{wavelet} = \mathcal{W}^{-1}(T_\lambda(\mathcal{W}(V))), \quad \lambda = \hat{\sigma}\sqrt{2\log N} \tag{A.8}$$

where V is the input volume, $V_{wavelet}$ is the denoised output, $\mathcal{W}$ and $\mathcal{W}^{-1}$ represent the forward and inverse wavelet transforms, T_λ is the thresholding operator, λ is

the threshold parameter, $\hat{\sigma}$ is the estimated noise standard deviation, and N is the number of voxels.

Morphological Processing

Top-hat transform specifically enhances electrode-like cylindrical structures:

$$V_{tophat} = V - (V \circ B_{disk}(r = 0.5\text{mm})) \tag{A.9}$$

where V is the input volume, V_{tophat} is the top-hat filtered output, $\circ$ denotes morphological opening, and $B_{disk}(r)$ represents a disk-shaped structuring element with radius r.

CLAHE Enhancement

Contrast Limited Adaptive Histogram Equalization operates on optimized $8 \times 8 \times 8$ sub-volumes:

$$V_{clahe} = \text{CLAHE}(V; c = 3.0, r_{block} = 8) \tag{A.10}$$

where V is the input volume, V_{clahe} is the contrast-enhanced output, c is the contrast limitation parameter, and r_{block} represents the block size for local histogram equalization.

Adaptive Binarization

Each enhanced volume undergoes patient-specific thresholding:

$$V_{binary}^{(i)} = \mathbb{I}(V_i > T_i) \tag{A.11}$$

where V_i represents the i-th enhanced volume, $V_{binary}^{(i)}$ is the corresponding binary output, $\mathbb{I}$ is the indicator function, and T_i is the patient-specific threshold for volume i.

A.2.2 Parameter Specifications

Gaussian Filtering Parameters

- Standard deviation: $\sigma = 0.3$ mm (optimized for electrode contact size)
- Kernel size: $k = 2\lceil 3\sigma \rceil + 1$ (ensures complete coverage)
- Implementation: 3D isotropic kernel for uniform smoothing

Gamma Correction Parameters

- Gamma values: Typically $\gamma \in [0.5, 2.0]$ depending on image characteristics
- Normalization: Min-max scaling to preserve intensity range
- Application: Selective enhancement of high-intensity metallic regions

Sharpening Parameters

- Sharpening strength: $\alpha = 0.6$ (balance between enhancement and noise)
- Blur kernel standard deviation: $\sigma' = 1.0$ mm
- Method: Unsharp masking with optimized parameters

Wavelet Denoising Parameters

- Wavelet family: Daubechies-4 (optimal for edge preservation)
- Decomposition levels: 3 (multi-resolution analysis)
- Thresholding method: Adaptive Donoho soft thresholding
- Noise estimation: Robust median-based estimation

Morphological Processing Parameters

- Structuring element: Disk-shaped with radius $r = 0.5$ mm
- Operation: Top-hat transform for bright structure enhancement
- Target structures: Cylindrical electrode contacts

CLAHE Parameters

- Block size: $8 \times 8 \times 8$ voxels (local adaptation)
- Contrast limit: $c = 3.0$ (prevents over-enhancement)

- Method: Adaptive histogram equalization with clipping

A.2.3 Implementation Algorithms

Enhancement Pipeline Algorithm

1. Load input CT volume and brain mask
2. Apply ROI constraint using brain mask (where applicable)
3. Execute enhancement sequence according to Table 5.1
4. Apply adaptive binarization with predicted threshold
5. Perform morphological post-processing
6. Store enhanced volume for ensemble processing

A.3 Threshold Binarization Implementation Details

A.3.1 Data Preprocessing Protocol

Data Cleaning and Feature Selection

The target variable (`threshold`) was isolated and rows with missing values removed. Non-informative columns (indices, identifiers, derived ratios) were excluded to prevent data leakage. Only numerical features were retained for Random Forest compatibility.

Feature Quality Control

Quasi-constant features were removed using the criterion:

$$\frac{\max_{v \in \mathcal{V}_j} \text{count}(x_j = v)}{n} > 0.99 \tag{A.12}$$

Highly correlated pairs ($|r| > 0.95$) were identified and duplicates removed:

$$r_{ij} = \frac{\sum_{k=1}^{n}(x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j)}{\sqrt{\sum_{k=1}^{n}(x_{ik} - \bar{x}_i)^2}\sqrt{\sum_{k=1}^{n}(x_{jk} - \bar{x}_j)^2}} \tag{A.13}$$

When features exceeded 20, ANOVA F-statistics selected the most predictive subset.

Train-Test Partitioning

Stratified 80/20 split with fixed random state for reproducibility.

A.3.2 Statistical Analysis Framework

Feature Characterization

Descriptive statistics (mean, variance, skewness) characterized each feature. Skewness assessment identified features requiring transformation:

$$\text{Skewness}(x_j) = \frac{1}{n}\sum_{i=1}^{n}\left(\frac{x_{ij} - \bar{x}_j}{\sigma_j}\right)^3 \tag{A.14}$$

Strong feature-target correlations ($|r| > 0.7$) were identified as primary predictors using Pearson's coefficient.

Visualization and Pattern Discovery

Multi-dimensional analysis through histograms, scatter plots, correlation heatmaps, and outlier detection guided feature selection and revealed nonlinear relationships requiring ensemble modeling approaches.

A.3.3 Consensus-Based Outlier Detection Implementation

Data Preprocessing

Features and target variables were jointly preprocessed using:

- **RobustScaler:** Scales features using statistics robust to outliers [44]
- **Yeo-Johnson Transform:** Power transformation to stabilize variance and reduce skewness [55]

Individual Detector Algorithms

Isolation Forest [56]: Isolates anomalies through recursive data partitioning. Outliers exhibit shorter average path lengths in the tree structure.

Elliptic Envelope: Fits multivariate Gaussian distribution, flagging points outside

Mahalanobis distance threshold:

$$D^2 = (x - \mu)^T \Sigma^{-1} (x - \mu) \tag{A.15}$$

where D^2 is squared Mahalanobis distance, μ is the mean vector, and Σ^{-1} is the inverse covariance matrix.

Local Outlier Factor (LOF) [57]: Measures local density deviation relative to neighbors:

$$\text{LOF}_k(x_i) = \frac{\sum_{j \in N_k(x_i)} \frac{\text{lrd}_k(x_j)}{\text{lrd}_k(x_i)}}{|N_k(x_i)|} \tag{A.16}$$

where $N_k(x_i)$ represents k-nearest neighbors and lrd_k is local reachability density.

One-Class SVM [58]: Learns decision boundary around majority data:

$$\min_{w,\rho,\xi} \frac{1}{2}\|w\|^2 + \frac{1}{\nu n}\sum_{i=1}^{n} \xi_i - \rho \tag{A.17}$$

where w defines boundary orientation, ρ determines distance from origin, and ν controls outlier fraction.

Consensus Voting Mechanism

Final outlier classification requires majority agreement ($\geq 3/4$ detectors) using the criterion:

$$\text{outlier} = \sum \text{votes} \geq \frac{\text{num\_detectors}}{2} + 1$$

This conservative approach minimizes false positives while maintaining sensitivity to genuine outliers.

A.3.4 Random Forest Implementation Details

Ensemble Architecture

The Random Forest regressor implements Breiman's ensemble methodology [41], combining B regression trees each trained on bootstrap samples. The ensemble prediction is computed as:

$$\hat{y} = \frac{1}{B} \sum_{b=1}^{B} T_b(x) \tag{A.18}$$

where $\hat{y}$ is the final predicted value, B is the total number of trees, $T_b(x)$ represents the prediction from the b-th tree for input x.

Hyperparameter Optimization Framework

Bayesian optimization using Tree-structured Parzen Estimator (TPE) [42] efficiently searched the hyperparameter space. The optimization objective incorporated robustness to outliers:

$$\text{Objective} = \text{MAE}_{\text{all}} + \lambda \cdot \max(0, \text{MAE}_{\text{outliers}} - \text{MAE}_{\text{all}}) \tag{A.19}$$

where MAE_{all} is mean absolute error on the complete dataset, $\text{MAE}_{\text{outliers}}$ is error specifically on outlier samples, λ controls penalty weight, and $\max(0, \cdot)$ ensures penalty applies only when outlier performance degrades.

Hyperparameter Search Space

Optimized parameters included:

- **n_estimators:** Number of trees in the forest
- **max_depth:** Maximum tree depth for complexity control
- **min_samples_split/leaf:** Minimum samples for node splitting/leaf creation
- **bootstrap:** Bootstrap sampling strategy
- **max_features:** Feature subset size for each split

Cross-Validation Protocol

Five-fold cross-validation provided robust performance estimates across data partitions while preventing overfitting to specific patient characteristics. Each fold maintained representative target distribution through stratified sampling.

A.3.5 Model Evaluation Metrics

Error Quantification Metrics

Root Mean Square Error (RMSE): Scale-dependent metric heavily penalizing large errors:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{A.20}$$

Mean Absolute Error (MAE): Robust metric treating all errors equally:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{A.21}$$

where y_i represents actual values, $\hat{y}_i$ are predicted values, and n is the sample size.

Goodness-of-Fit Assessment

Coefficient of Determination (R^2): Proportion of variance explained by the model:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{A.22}$$

where $\bar{y}$ is the sample mean. Values range from 0 (no explanatory power) to 1 (perfect fit).

Outlier-Specific Performance

Model robustness was assessed by computing RMSE, MAE, and R^2 exclusively on the outlier subset identified through consensus detection. This analysis ensures reliable performance across the full spectrum of data variability, particularly important in biomedical applications where outliers may represent rare but clinically significant cases.

Visualization Framework

Qualitative assessment employed three complementary visualization approaches:

- **Actual vs. Predicted Plots:** Scatter plots with perfect prediction line for bias assessment

- **Residual Analysis:** Plots of prediction errors vs. predicted values for heteroscedasticity detection
- **Learning Curves:** Training and validation error vs. sample size for overfitting assessment

These visualizations provide insights into model behavior, training dynamics, and potential systematic errors not captured by numerical metrics alone.

A.3.6 Global Voting Mechanism Implementation Details

Binary Mask Generation

For each patient, a set of N binary masks $\{M_1, M_2, \ldots, M_N\}$ is generated, where each mask $M_i \in \{0,1\}^{D\times H\times W}$ represents a 3D binary segmentation volume with D representing depth, H height, and W width dimensions.

Threshold Selection Approaches **Model-based thresholding**:

$$T_{\text{pred}}^{(i)} = f_{\text{RF}}(\mathbf{x}_i) \tag{A.23}$$

where f_{RF} represents the trained Random Forest regressor and $\mathbf{x}_i$ is the feature vector extracted from enhanced volume V_i.

Empirical fixed thresholds:

$$T_{\text{fixed}}^{(i)} = \text{percentile}(V_i, p), \quad p \in [95, 99.5] \tag{A.24}$$

Binary mask generation:

$$M_i(x,y,z) = \begin{cases} 1 & \text{if } V_i(x,y,z) > T^{(i)} \\ 0 & \text{otherwise} \end{cases} \tag{A.25}$$

Global Vote Map Construction

The voting mechanism aggregates all binary masks to create a consensus vote map:

$$\mathcal{V}(x,y,z) = \sum_{i=1}^{N} M_i(x,y,z) \tag{A.26}$$

where $\mathcal{V}(x, y, z)$ quantifies agreement levels at each voxel location, ranging from 0 (no agreement) to N (unanimous consensus).

Quality-Based Mask Selection

Overlap scores measure each mask's contribution to the global consensus:

$$S_i = \frac{\sum_{x,y,z} M_i(x, y, z) \cdot \mathcal{V}(x, y, z)}{\sum_{x,y,z} M_i(x, y, z)} \tag{A.27}$$

where summations iterate over all voxel coordinates in the 3D volume.

Top-performing masks are selected based on ranking:

$$\mathcal{M}_{\text{selected}} = \{M_{i_1}, M_{i_2}, \ldots, M_{i_K}\} \tag{A.28}$$

where $S_{i_1} \geq S_{i_2} \geq \ldots \geq S_{i_K}$ represents decreasing overlap scores.

Weighted Consensus Fusion

Score-weighted fusion emphasizes reliable segmentations:

$$\mathcal{W}(x, y, z) = \frac{\sum_{j=1}^{K} S_{i_j} M_{i_j}(x, y, z)}{\sum_{j=1}^{K} S_{i_j}} \tag{A.29}$$

Final binary consensus through majority thresholding:

$$M_{\text{consensus}}(x, y, z) = \mathbb{I}(\mathcal{W}(x, y, z) \geq 0.5) \tag{A.30}$$

where $\mathbb{I}(\cdot)$ is the indicator function.

Quality Assessment Metrics

Dice Similarity Coefficient:

$$\text{DSC}(M_i, M_{\text{consensus}}) = \frac{2|M_i \cap M_{\text{consensus}}|}{|M_i| + |M_{\text{consensus}}|} \tag{A.31}$$

Volume Consistency:

$$\text{CV} = \frac{\sigma_{\text{volume}}}{\mu_{\text{volume}}} \times 100\% \tag{A.32}$$

where σ_{volume} and μ_{volume} represent standard deviation and mean of mask volumes respectively.

A.4 Centroids model predictor

A.4.1 Data Preparation Technical Details

Mathematical Formulations

Connected Component Analysis

$$L^{(i)} = \text{CCL}(M^{(i)}, \mathcal{N}_{26}) \tag{A.33}$$

$$(x_1, y_1, z_1) \sim_{26} (x_2, y_2, z_2) \iff \max(|x_1 - x_2|, |y_1 - y_2|, |z_1 - z_2|) \leq 1 \tag{A.34}$$

$$\mathcal{C}^{(i)} = \{C_j^{(i)} : j \in \{1, 2, \ldots, K_i\}, C_j^{(i)} = \{(x, y, z) : L^{(i)}(x, y, z) = j\}\} \tag{A.35}$$

where $\text{CCL}(\cdot, \mathcal{N}_{26})$ denotes the connected component labeling operator with 26-neighborhood connectivity, $L^{(i)}$ represents the labeled array, and $\mathcal{C}^{(i)}$ constitutes the set of K_i connected components for mask $M^{(i)}$.

Centroid Computation

$$\mathbf{c}_j^{(i)} = \frac{\sum_{(x,y,z) \in C_j^{(i)}} I(x, y, z) \cdot \mathbf{r}(x, y, z)}{\sum_{(x,y,z) \in C_j^{(i)}} I(x, y, z)} \tag{A.36}$$

$$= \frac{M_{100}^{(j)} \mathbf{e}_x + M_{010}^{(j)} \mathbf{e}_y + M_{001}^{(j)} \mathbf{e}_z}{M_{000}^{(j)}} \tag{A.37}$$

$$M_{pqr}^{(j)} = \sum_{(x,y,z) \in C_j^{(i)}} x^p y^q z^r I(x, y, z) \tag{A.38}$$

where $\mathbf{r}(x, y, z) = x\mathbf{e}_x + y\mathbf{e}_y + z\mathbf{e}_z$ is the position vector and $I(x, y, z)$ represents the intensity at voxel coordinates.

Coordinate Transformations

Stage 1: Array-to-Index Coordinate Reordering

$$\mathbf{c}_j^{\text{index}} = \Pi \cdot \mathbf{c}_j^{\text{array}} = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} c_z \\ c_y \\ c_x \end{bmatrix} = \begin{bmatrix} c_x \\ c_y \\ c_z \end{bmatrix} \tag{A.39}$$

Stage 2: Index-to-Physical Transformation

$$\mathbf{c}_j^{\text{physical}} = \mathbf{T}_{\text{origin}} + \mathbf{S} \cdot \mathbf{D} \cdot \mathbf{c}_j^{\text{index}} \tag{A.40}$$

where $\mathbf{T}_{\text{origin}} \in \mathbb{R}^3$ is the image origin, $\mathbf{S} = \text{diag}(s_x, s_y, s_z)$ is the voxel spacing matrix, and $\mathbf{D} \in SO(3)$ represents the direction cosine matrix encoding image orientation.

Stage 3: LPS-to-RAS Conversion

$$\mathbf{c}_j^{\text{RAS}} = \mathbf{R}_{\text{LPS}\to\text{RAS}} \cdot \mathbf{c}_j^{\text{physical}} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_{\text{LPS}} \\ y_{\text{LPS}} \\ z_{\text{LPS}} \end{bmatrix} \tag{A.41}$$

Ground Truth Volume Generation

Spherical Electrode Model

$$\mathcal{S}_k(\mathbf{x}) = \{\mathbf{x} \in \mathbb{R}^3 : \|\mathbf{x} - \mathbf{g}_k\|_2 \leq r_{\text{electrode}}\} \tag{A.42}$$

$$M_{\text{GT}}(\mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{x} \in \bigcup_{k=1}^{N_{\text{GT}}} \mathcal{S}_k(\mathbf{x}) \\ 0 & \text{otherwise} \end{cases} \tag{A.43}$$

Distance Transform Implementation

$$D_k(\mathbf{x}) = \text{SignedMaurerDistanceMap}(\delta_{\mathbf{g}_k}) \tag{A.44}$$

$$\mathcal{S}_k = \{\mathbf{x} : D_k(\mathbf{x}) \leq r_{\text{electrode}}\} \tag{A.45}$$

where $\delta_{\mathbf{g}_k}$ represents a unit impulse at coordinate $\mathbf{g}_k$.

Bipartite Matching Formulation

Optimization Problem

$$\text{minimize} \quad \sum_{j=1}^{K_i} \sum_{k=1}^{N_{\text{GT}}} w_{jk} x_{jk} \tag{A.46}$$

$$\text{subject to} \quad \sum_{k=1}^{N_{\text{GT}}} x_{jk} \leq 1, \quad \forall j \in \{1, \ldots, K_i\} \tag{A.47}$$

$$\sum_{j=1}^{K_i} x_{jk} \leq 1, \quad \forall k \in \{1, \ldots, N_{\text{GT}}\} \tag{A.48}$$

$$x_{jk} \in \{0, 1\}, \quad w_{jk} = \|\mathbf{c}_j^{\text{RAS}} - \mathbf{g}_k\|_2 \tag{A.49}$$

Success Criterion

$$\text{Success}(\mathbf{c}_j) = \begin{cases} 1 & \text{if } \min_k \|\mathbf{c}_j^{\text{RAS}} - \mathbf{g}_k\|_2 \leq \tau_{\text{clinical}} \\ 0 & \text{otherwise} \end{cases} \tag{A.50}$$

where $\tau_{\text{clinical}} = 2.0$ mm represents the clinical tolerance.

Quality Assessment Metrics

Volumetric Overlap Assessment

Standard Dice Similarity Coefficient

$$\text{DSC}(M_{\text{pred}}, M_{\text{GT}}) = \frac{2|M_{\text{pred}} \cap M_{\text{GT}}|}{|M_{\text{pred}}| + |M_{\text{GT}}|} \tag{A.51}$$

$$= \frac{2 \sum_{(x,y,z)} M_{\text{pred}}(x, y, z) \cdot M_{\text{GT}}(x, y, z)}{\sum_{(x,y,z)} M_{\text{pred}}(x, y, z) + \sum_{(x,y,z)} M_{\text{GT}}(x, y, z)} \tag{A.52}$$

Dilated Dice Coefficient

$$\text{DSC}_{\text{dilated}} = \text{DSC}(M_{\text{pred}} \oplus B_r, M_{\text{GT}} \oplus B_r) \tag{A.53}$$

where B_r represents a spherical structuring element with radius $r = 1$ voxel, and $\oplus$

denotes morphological dilation.

Statistical Distance Measures

$$\text{RMSE}_{\text{centroid}} = \sqrt{\frac{1}{N_{\text{matches}}} \sum_{j=1}^{N_{\text{matches}}} \|\mathbf{c}_j^{\text{RAS}} - \mathbf{g}_{m(j)}\|_2^2} \tag{A.54}$$

$$\text{MAE}_{\text{centroid}} = \frac{1}{N_{\text{matches}}} \sum_{j=1}^{N_{\text{matches}}} \|\mathbf{c}_j^{\text{RAS}} - \mathbf{g}_{m(j)}\|_2 \tag{A.55}$$

$$\text{HD}(M_{\text{pred}}, M_{\text{GT}}) = \max\left(\max_{\mathbf{p}\in M_{\text{pred}}} \min_{\mathbf{q}\in M_{\text{GT}}} \|\mathbf{p}-\mathbf{q}\|_2, \max_{\mathbf{q}\in M_{\text{GT}}} \min_{\mathbf{p}\in M_{\text{pred}}} \|\mathbf{q}-\mathbf{p}\|_2 \right) \tag{A.56}$$

where $m(j)$ denotes the optimal matching function from the bipartite graph solution.

Clinical Performance Metrics

$$\text{Sensitivity} = \frac{TP}{TP+FN} = \frac{N_{\text{successful}}}{N_{\text{GT}}} \tag{A.57}$$

$$\text{Precision} = \frac{TP}{TP+FP} = \frac{N_{\text{successful}}}{N_{\text{detected}}} \tag{A.58}$$

$$\text{F}_1\text{-score} = \frac{2 \cdot \text{Precision} \cdot \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}} \tag{A.59}$$

where TP, FP, and FN represent true positives, false positives, and false negatives, respectively.

Morphological Feature Characterization

$$\text{Volume} = |C_j^{(i)}| \cdot s_x \cdot s_y \cdot s_z \tag{A.60}$$

$$\text{Sphericity} = \frac{\pi^{1/3}(6V)^{2/3}}{A} \tag{A.61}$$

$$\text{Elongation} = \frac{\lambda_3}{\lambda_1} \tag{A.62}$$

$$\text{Convexity} = \frac{V}{V_{\text{convex\_hull}}} \tag{A.63}$$

where V denotes volume, A represents surface area, $\lambda_1 \geq \lambda_2 \geq \lambda_3$ are the eigenvalues of the covariance matrix of the region coordinates, and $V_{\text{convex\_hull}}$ is the volume of the convex hull.

Statistical Quality Control

Mahalanobis Distance-Based Outlier Detection

$$d_M(\mathbf{x}_j) = \sqrt{(\mathbf{x}_j - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{x}_j - \boldsymbol{\mu})} \tag{A.64}$$

where $\mathbf{x}_j$ represents the feature vector for electrode j, $\boldsymbol{\mu}$ is the sample mean, and $\boldsymbol{\Sigma}$ denotes the covariance matrix. Points with $d_M > \chi^2_{\alpha,p}$ are flagged as outliers, where $\chi^2_{\alpha,p}$ represents the critical value of the chi-squared distribution with p degrees of freedom at significance level $\alpha = 0.05$.

Inter-Electrode Distance Distribution Analysis

$$\mathcal{D}_{\text{inter}} = \{\|\mathbf{c}_j - \mathbf{c}_k\|_2 : j, k \in \{1, \ldots, K\}, j \neq k\} \tag{A.65}$$

$$\text{Z-score}(d) = \frac{d - \mu_{\mathcal{D}}}{\sigma_{\mathcal{D}}} \tag{A.66}$$

Electrode pairs with $|\text{Z-score}(d)| > 3$ are flagged for manual review.

Implementation Specifications

Parallel Processing Architecture

The computational pipeline employs a worker pool pattern with the following specifications:

- **Worker allocation**: Dynamic based on available CPU cores
- **Task distribution**: Round-robin assignment of mask processing tasks
- **Memory management**: Shared memory arrays for large mask volumes
- **Load balancing**: Adaptive task queue with work stealing
- **Error handling**: Graceful degradation with task retry mechanisms

Computational Complexity

The overall computational complexity of the data preparation pipeline is:

$$\mathcal{O}(N \cdot K \cdot (V \log V + M^2)) \tag{A.67}$$

where N is the number of masks, K is the average number of connected components per mask, V is the volume size, and M is the number of ground truth electrodes.

Reproducibility Framework

Data Provenance Documentation

Coordinate Manifests JSON-formatted files containing:

- Complete electrode coordinate data with RAS coordinates
- Metadata including coordinate system transformations
- Processing timestamps and software version information
- Parameter settings used during processing

Validation Reports Structured CSV tables containing:

- Quantitative assessment metrics for each mask variant
- Confidence intervals and statistical significance tests
- Performance comparisons across different processing methods
- Quality control flags and warnings

Quality Assessment Visualizations Automated generation of:

- 3D scatter plots of electrode positions in RAS coordinate space
- Distance distribution histograms for spatial accuracy assessment
- Correlation matrices for feature relationship analysis
- Processing workflow diagrams with timing information

A.4.2 Feature Extraction Technical Details

Spatial and Neighborhood Features

Coordinate Representation and Proximity Analysis Each electrode candidate is characterized by its RAS coordinates $\mathbf{p}_i = [x_i^{RAS}, y_i^{RAS}, z_i^{RAS}]^T$, where RAS

denotes the Right-Anterior-Superior anatomical reference system. Efficient neighborhood analysis uses KDTree spatial indexing with clinical proximity threshold $\tau = 7.0$ mm:

$$\mathcal{N}_i(\tau) = \{\mathbf{p}_j : \|\mathbf{p}_i - \mathbf{p}_j\|_2 \leq \tau, j \neq i\} \tag{A.68}$$

$$n_i = |\mathcal{N}_i(\tau)|, \quad \mu_i^{dist} = \frac{1}{n_i} \sum_{\mathbf{p}_j \in \mathcal{N}_i(\tau)} \|\mathbf{p}_i - \mathbf{p}_j\|_2 \tag{A.69}$$

where $\mathcal{N}_i(\tau)$ defines the neighborhood of electrode candidate i within the clinical proximity threshold, n_i represents the neighbor count, and μ_i^{dist} quantifies the mean neighbor distance reflecting adherence to 3.5mm manufacturing spacing.

Kernel density estimation provides scale-independent measures:

$$\rho_i = \frac{1}{Nh^3} \sum_{j=1}^{N} K\left(\frac{\|\mathbf{p}_i - \mathbf{p}_j\|_2}{h}\right), \quad h = 5.0 \text{ mm} \tag{A.70}$$

where ρ_i represents the local density at electrode candidate i, N is the total number of candidates, h is the bandwidth parameter optimized for electrode detection, and $K(\cdot)$ is the 3D Gaussian kernel function.

Intensity and Gradient Features

Spherical Neighborhood Analysis Within 2mm spherical neighborhoods $\mathcal{S}_i(r)$, where $r = 2$ voxels corresponds to the physical electrode contact radius, intensity characteristics are quantified:

$$\mu_i^{intensity} = \frac{1}{|\mathcal{S}_i(r)|} \sum_{\mathbf{v} \in \mathcal{S}_i(r)} I(\mathbf{v}) \tag{A.71}$$

$$H_i = 1 - \frac{\sigma_i^{intensity}}{\mu_i^{intensity}} \quad \text{(homogeneity score)} \tag{A.72}$$

where $I(\mathbf{v})$ represents the CT intensity at voxel $\mathbf{v}$ in Hounsfield Units, $\mu_i^{intensity}$ captures the average intensity within the electrode neighborhood, and H_i quantifies

metallic uniformity with values approaching unity indicating consistent composition.

Gradient magnitude using 3D Sobel operators captures material boundaries:

$$\bar{G}_i = \frac{1}{|\mathcal{S}_i(r)|} \sum_{\mathbf{v} \in \mathcal{S}_i(r)} \sqrt{(\nabla_x I)^2 + (\nabla_y I)^2 + (\nabla_z I)^2} \tag{A.73}$$

where $\nabla_x I$, $\nabla_y I$, and $\nabla_z I$ represent intensity gradients along each spatial axis, and $\bar{G}_i$ quantifies the average edge strength characteristic of sharp metallic boundaries.

Trajectory and Surface Features

Principal Component Analysis Global directional patterns are captured through PCA on the complete electrode set:

$$\mathbf{p}_i^{PCA} = \mathbf{V}^T(\mathbf{p}_i - \overline{\mathbf{p}}) = [PC1_i, PC2_i, PC3_i]^T \tag{A.74}$$

where $\mathbf{V}$ contains the principal component vectors derived from the spatial covariance matrix, $\overline{\mathbf{p}}$ is the global electrode centroid, and $[PC1_i, PC2_i, PC3_i]^T$ represents the projection capturing systematic organization patterns.

Local anisotropy quantifies directional consistency:

$$A_i = \frac{\lambda_i^{(1)}}{\sum_{k=1}^{3} \lambda_i^{(k)} + \epsilon} \tag{A.75}$$

where $\lambda_i^{(1)}$ is the largest eigenvalue from local neighborhood PCA, $\lambda_i^{(k)}$ are the three eigenvalues representing directional variance, and $\epsilon = 10^{-10}$ prevents numerical instability. Values approaching unity indicate strong linear organization.

Anatomical Constraints Surface distance features ensure anatomical plausibility:

$$d_i^{surface} = \min_{\mathbf{v}_j \in \mathbf{V}_{RAS}} \|\mathbf{p}_i - \mathbf{v}_j\|_2 \tag{A.76}$$

where $\mathbf{V}_{RAS}$ contains brain surface vertices extracted using Marching Cubes algorithm, and $d_i^{surface}$ represents the minimum Euclidean distance ensuring electrodes

are positioned within anatomically plausible brain regions.

Final feature vector integrates all domains:

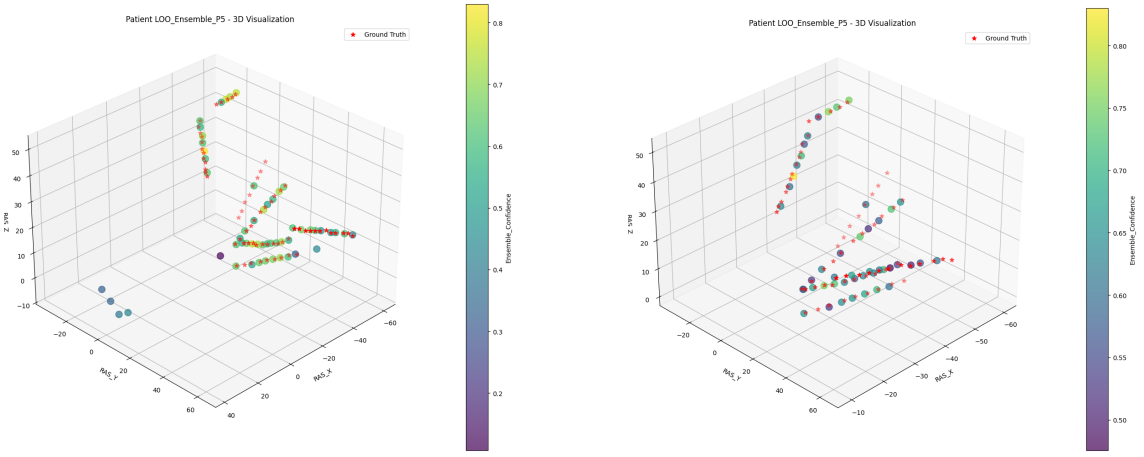
$$\mathcal{F}_i = [\mathbf{p}_i^{RAS}, n_i, \rho_i, \mu_i^{intensity}, H_i, \bar{G}_i, \mathbf{p}_i^{PCA}, A_i, d_i^{surface}] \tag{A.77}$$

where $\mathcal{F}_i$ represents the 38-dimensional feature vector combining spatial coordinates, neighborhood statistics, intensity characteristics, trajectory alignment, and anatomical constraints for robust electrode authentication.

A.4.3 Additional Patient Confidence Score Visualizations

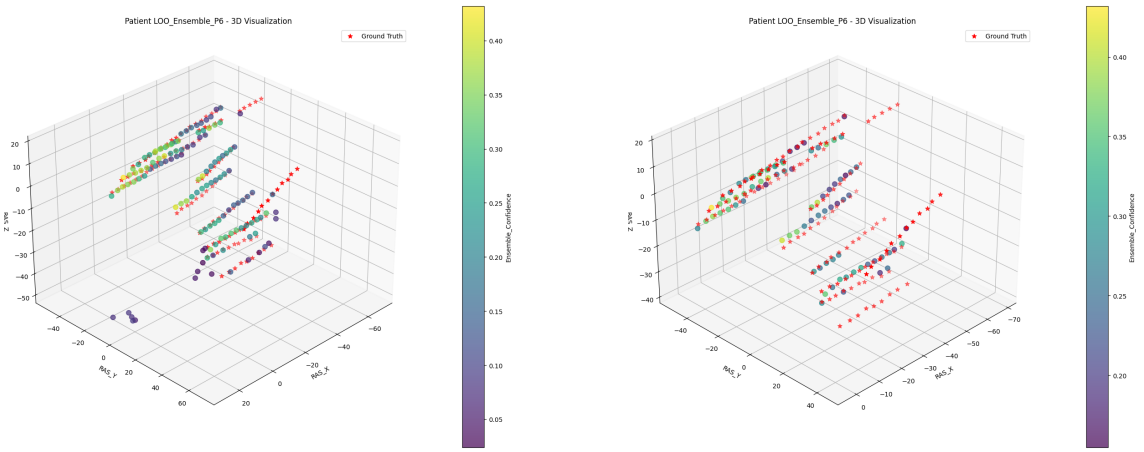
The confidence scoring system performance is further illustrated through visualization across the patient cohort, demonstrating the system's behavior under varying anatomical complexities and imaging conditions. This section provides additional examples of confidence-based electrode detection across patients with different clinical characteristics.

The following visualizations demonstrate the practical application of the confidence scoring framework described in Section 6.6, showing how the regression-based authentication system performs across the spectrum of clinical presentations encountered in SEEG electrode localization.



(a) *Confidence score = 0%* **(b)** *Confidence score = 45%*

Figure A.4: *SEEG electrode detection visualization for Patient P5 showing confidence score progression from 0% to 45%. Patient P5 represents one of the more challenging cases in the cohort (20.9% success within 1mm threshold in LOPO validation), demonstrating the system's ability to provide graduated certainty estimates even in difficult anatomical scenarios. The improved spatial organization at higher confidence levels illustrates the effectiveness of the regression-based authentication framework. Source: Own source.*



(a) *Confidence score = 0%* **(b)** *Confidence score = 15%*

Figure A.5: *SEEG electrode detection visualization for Patient P6 showing confidence score progression from 0% to 15%. Patient P6 was identified as an outlier case in the confidence-distance analysis (Figure 7.16), with low confidence scores reflecting the inherent difficulty of electrode localization in complex anatomical presentations. The 15% confidence level falls within the range where clinical review is recommended according to the tiered decision framework. Source: Own source.*

Clinical Interpretation of Challenging Cases

Patients P5 and P6 represent the spectrum of challenging anatomical presentations encountered in clinical practice. These cases demonstrate several key aspects of the confidence scoring system:

Conservative Design Philosophy: The low confidence scores (0-45%) observed in these challenging cases reflect the system's conservative design philosophy described in Section 8.2.1, where aggressive penalization prevents overconfidence in uncertain predictions.

Graduated Decision Support: Even in difficult cases, the confidence scoring system provides meaningful discrimination between different levels of uncertainty, enabling clinicians to prioritize manual review efforts based on prediction reliability.

Clinical Workflow Integration: These examples illustrate how the tiered decision framework operates in practice, with low confidence predictions (P6: 0-15%) requiring manual validation and medium confidence cases (P5: 45%) supporting clinical review protocols.

Anatomical Complexity Handling: The visualization demonstrates how complex electrode configurations and challenging imaging conditions are appropriately flagged by the confidence system, ensuring that uncertain predictions are identified for clinical attention rather than being incorrectly classified as high-confidence detections.

These additional patient examples complement the confidence score validation presented in Section 7.5.4, providing visual evidence of the system's robust performance across the full spectrum of clinical presentations encountered in SEEG electrode localization.

A.5 Entry and Bolt Head detections

A.5.1 Adaptive Threshold Algorithm Implementation Details

Methodological Rationale for Feature Selection

Bolt heads represent the highest-density metallic components in post-operative CT scans, typically exhibiting Hounsfield unit values exceeding 2300 HU due to their platinum-iridium composition. Unlike electrode contacts, which appear as smaller cylindrical structures, bolt heads present as compact metallic assemblies with predictable volume characteristics but variable intensity signatures depending on imaging protocols and electrode positioning.

The feature selection strategy focused on extreme percentile analysis for several methodological reasons:

- **Tail behavior characterization**: Metallic electrodes consistently appear in the upper 0.05% of intensity distributions, making extreme percentiles robust indicators of electrode presence
- **Noise immunity**: High percentiles remain stable despite imaging noise that affects lower intensity regions
- **Scanner independence**: Percentile-based features provide relative measures less sensitive to absolute calibration differences between scanners

Statistical Validation of Feature Selection

Leave-one-out cross-validation was employed to assess feature stability and predictive power across the patient cohort. The analysis evaluated correlation strength between extracted features and manually selected optimal thresholds:

Table A.1: *Feature correlation analysis with manually selected thresholds*

Feature	Correlation (r)	p-value	Stability (CV)	Clinical Relevance
$P_{99.95}$	0.94	< 0.001	4.2%	High
$P_{99.97}$	0.91	< 0.002	3.8%	High
$P_{99.98}$	0.89	< 0.005	5.1%	Medium
ρ_{2400}	0.87	< 0.008	12.3%	High
$\nabla_{2300-2500}$	0.73	< 0.02	18.7%	Medium

The selected percentiles demonstrated strong correlation with optimal thresholds (r > 0.89) while maintaining low coefficient of variation (CV $< 6\%$), confirming their reliability for threshold prediction.

Mathematical Framework for Feature Extraction

For each patient volume $\mathcal{V}$, features are systematically extracted from the filtered intensity array $\mathcal{I} = \{i \in \mathcal{V} : i > -1000\}$ to exclude air regions and focus on tissue and metallic components:

High Percentile Analysis:

$$P_{99.95} = \text{percentile}(\mathcal{I}, 99.95) \tag{A.78}$$

$$P_{99.97} = \text{percentile}(\mathcal{I}, 99.97) \tag{A.79}$$

$$P_{99.98} = \text{percentile}(\mathcal{I}, 99.98) \tag{A.80}$$

These percentiles capture the intensity characteristics of the highest-density metallic components while remaining robust to outliers and imaging artifacts.

Electrode Density Metrics:

$$\rho_{2400} = \frac{|\{i \in \mathcal{I} : i > 2400\}|}{|\mathcal{I}|} \tag{A.81}$$

This ratio serves as a fundamental discriminator between patient phenotypes, ranging from < 0.0003 in low-visibility cases to > 0.002 in high-density electrode configurations. The threshold value of 2400 HU was selected based on empirical analysis

of electrode intensity distributions across the cohort.

Intensity Gradient Analysis:

$$\nabla_{2300-2500} = \frac{1}{|\mathcal{I}_{2300-2500}|} \sum_{i \in \mathcal{I}_{2300-2500}} (i - 2400) \tag{A.82}$$

where $\mathcal{I}_{2300-2500} = \{i \in \mathcal{I} : 2300 \leq i \leq 2500\}$. Negative gradients indicate steep intensity transitions typical of high-artifact scenarios caused by beam hardening effects.

A.5.2 Patient Classification and Algorithm Implementation

Detailed Patient Phenotype Analysis

Systematic analysis of the eight-patient dataset revealed distinct patterns that enable reliable threshold prediction. The following patient categories were identified based on their intensity distribution characteristics:

Special Case Identification:

- **P7-like (High Density + Artifacts):** Characterized by $\text{ratio}_{2400} > 0.002$ and $\text{gradient}_{2300-2500} < -30$, indicating high electrode visibility with significant imaging artifacts requiring aggressive thresholding
- **P6-like (Low Visibility):** Defined by $\text{ratio}_{2400} < 0.0003$, representing cases with minimal electrode visibility necessitating sensitive threshold selection
- **P4-like (Scanner Saturation):** Identified when $P_{99.97} = P_{99.98}$ and $P_{99.97} > 3060$, indicating CT scanner intensity saturation effects

General Case Patterns: For standard cases not meeting special criteria, the algorithm leverages the consistent relationship between $P_{99.95}$ and optimal thresholds:

- **Low Ratio Cases** ($\text{ratio}_{2400} < 0.0005$): Optimal thresholds typically exceed $P_{99.95}$ by 200 HU
- **Higher Ratio Cases** ($\text{ratio}_{2400} \geq 0.0005$): Thresholds fall below $P_{99.95}$ by a distance proportional to electrode density

Complete Algorithm Implementation

Algorithm 1 Adaptive Threshold Selection for Bolt Head Segmentation

procedure PREDICTTHRESHOLD($P_{99.95}, P_{99.97}, P_{99.98}, \text{ratio}_{2400}, \text{gradient}$)
  $T \leftarrow 2400$ ▷ Initialize with baseline threshold
  reason ← "Standard case - baseline threshold"
  if $\text{ratio}_{2400} > 0.002$ **and** $\text{gradient} < -30$ **then**
    $T \leftarrow 2815$ ▷ P7-like: High density + artifacts
    reason ← "High electrode density with steep gradient"
  else if $\text{ratio}_{2400} < 0.0003$ **then**
    $T \leftarrow 2325$ ▷ P6-like: Low visibility
    reason ← "Very low electrode density"
  else if $P_{99.97} = P_{99.98}$ **and** $P_{99.97} > 3060$ **then**
    $P_{99.8} \leftarrow \text{percentile}(\mathcal{I}, 99.8)$
    $P_{99.9} \leftarrow \text{percentile}(\mathcal{I}, 99.9)$
    $T \leftarrow \lfloor (P_{99.8} + P_{99.9})/2 \rfloor$ ▷ P4-like: Scanner saturation
    reason ← "Scanner saturation detected"
  else
    if $\text{ratio}_{2400} < 0.0005$ **then**
      $T_{base} \leftarrow \min(2400, P_{99.95} + 200)$ ▷ Low ratio case
      reason ← "Low ratio: P99.95 + 200"
    else
      $d \leftarrow \min(300, \text{ratio}_{2400} \times 100000)$ ▷ Distance calculation
      $T_{base} \leftarrow \min(2400, P_{99.95} - d)$ ▷ Higher ratio case
      reason ← "Higher ratio: P99.95 - distance"
    end if
    $T \leftarrow T_{base}$
    $T_{upper} \leftarrow P_{99.97} - 100$ ▷ Universal upper bound
    if $T > T_{upper}$ **then**
      $T \leftarrow \lfloor T_{upper} \rfloor$
      reason ← reason + " → capped at P99.97-100"
    end if
  end if
  $T \leftarrow \max(2325, \min(2815, T))$ ▷ Safety bounds enforcement
  return (T, reason)
end procedure

The complete adaptive threshold algorithm implements a hierarchical decision structure based on the validated patient phenotypes. The mathematical foundation relies on the empirically derived relationships:

$$\tau_{optimal} = \begin{cases} \mathcal{F}_{special}(\rho_{2400}, \nabla_{2300-2500}) & \text{if special case criteria met} \\ \mathcal{F}_{general}(P_{99.95}, \rho_{2400}) & \text{otherwise} \end{cases} \tag{A.83}$$

where $\mathcal{F}_{special}$ and $\mathcal{F}_{general}$ represent the case-specific threshold functions derived from cross-validation analysis.

A.5.3 Mathematical Formulation and Safety Constraints

General Case Threshold Calculation

The general case threshold calculation follows the relationship:

$$T = \begin{cases} 2815 & \text{if ratio}_{2400} > 0.002 \text{ and gradient} < -30 \\ 2325 & \text{if ratio}_{2400} < 0.0003 \\ \lfloor (P_{99.8} + P_{99.9})/2 \rfloor & \text{if } P_{99.97} = P_{99.98} \text{ and } P_{99.97} > 3060 \\ \min(P_{99.97} - 100, T_{general}) & \text{otherwise} \end{cases} \tag{A.84}$$

where:

$$T_{general} = \begin{cases} \min(2400, P_{99.95} + 200) & \text{if ratio}_{2400} < 0.0005 \\ \min(2400, P_{99.95} - \min(300, 100000 \times \text{ratio}_{2400})) & \text{if ratio}_{2400} \geq 0.0005 \end{cases} \tag{A.85}$$

Safety Bounds Implementation

All threshold predictions are constrained within physiologically and technically reasonable bounds:

$$T_{final} = \max(2325, \min(2815, T)) \tag{A.86}$$

These bounds prevent extreme threshold values that could compromise segmentation quality or clinical interpretability. The lower bound (2325 HU) ensures sufficient sensitivity for low-visibility electrodes, while the upper bound (2815 HU) prevents over-aggressive thresholding that could include non-electrode metallic artifacts.

A.6 Advanced Trajectory Reconstruction Framework

This appendix provides mathematical formulations for the novel multi-algorithm consensus framework developed for SEEG electrode trajectory reconstruction. For complete implementation details and usage examples, refer to the **SEEG Electrode Trajectory Analysis Module Documentation** [1].

A.6.1 Multi-Algorithm Consensus Mathematical Framework

Primary Consensus Equation:

$$\mathcal{T} = \text{Consensus}(\text{DBSCAN}(\varepsilon, n_{\min}), \text{Louvain}(G_{\text{proximity}})) \tag{A.87}$$

where $\mathcal{T}$ represents the final trajectory set, $\varepsilon = 7.5$ mm is the spatial proximity threshold, $n_{\min}$ is the minimum samples parameter, and $G_{\text{proximity}}$ is the proximity graph with edge weights $w_{ij} = (d_{ij} + \epsilon)^{-1}$ for contacts within $\varepsilon = 7.5$ mm.

DBSCAN Formulation:

$$\text{DBSCAN}(\epsilon, \text{MinPts}) = \{C_1, C_2, \ldots, C_k, N\} \tag{A.88}$$

where $C_1, C_2, \ldots, C_k$ represent the identified clusters (trajectories), N denotes noise points, ϵ is the neighborhood radius, and MinPts is the minimum number of points required to form a cluster.

Louvain Modularity:

$$Q = \frac{1}{2m} \sum_{ij} \left[A_{ij} - \frac{k_i k_j}{2m} \right] \delta(c_i, c_j) \tag{A.89}$$

[1] Complete technical documentation: Ávalos, R. (2024). *SEEG Electrode Trajectory Analysis Module: Documentation.* Available at: `https://github.com/rociavl/SEEG_automatic_segmentation/blob/main/Electrode_path/seeg_trajectories.pdf`

where Q is the modularity score, m is the total number of edges, A_{ij} is the adjacency matrix, k_i and k_j are the degrees of nodes i and j, and $\delta(c_i, c_j)$ is 1 if nodes i and j are in the same community, 0 otherwise.

Adaptive Parameter Optimization:

$$\text{Score}(\varepsilon, n_{\min}) = 0.7 \cdot P_{\text{valid}} + 0.3 \cdot P_{\text{clustered}} \tag{A.90}$$

where P_{valid} is the proportion of trajectories meeting clinical validity criteria and $P_{\text{clustered}}$ is the proportion of electrode contacts successfully assigned to clusters.

A.6.2 Geometric Validation Framework

Trajectory Linearity:

$$L_{\text{trajectory}} = \frac{\lambda_1}{\sum_{i=1}^{3} \lambda_i} \tag{A.91}$$

where $L_{\text{trajectory}}$ is the linearity score, λ_1 is the largest eigenvalue from Principal Component Analysis of trajectory coordinates, and λ_i are the three eigenvalues representing variance along each principal axis.

Curvature Analysis:

$$\theta_{\text{curvature}} = 180\check{\text{r}} - \arccos\left(\frac{\vec{v}_i \cdot \vec{v}_{i+1}}{|\vec{v}_i||\vec{v}_{i+1}|}\right) \tag{A.92}$$

where $\theta_{\text{curvature}}$ is the curvature angle between consecutive trajectory segments, $\vec{v}_i$ and $\vec{v}_{i+1}$ are direction vectors of adjacent electrode segments, and the dot product measures angular deviation from linearity.

Entry Angle Validation:

$$\theta_{\text{entry}} = 90\check{\text{r}} - \arccos(|\vec{d}_{\text{trajectory}} \cdot \vec{n}_{\text{surface}}|) \tag{A.93}$$

where θ_{entry} is the entry angle relative to the brain surface, $\vec{d}_{\text{trajectory}}$ is the trajectory direction vector, and $\vec{n}_{\text{surface}}$ is the surface normal vector at the entry point.

Spacing Regularity:

$$R_{\text{spacing}} = 1 - \frac{\sigma_d}{\mu_d}, \quad d_i \in [3.0, 5.0] \text{ mm} \tag{A.94}$$

where R_{spacing} is the spacing regularity score, σ_d is the standard deviation of inter-contact distances, μ_d is the mean inter-contact distance, and d_i represents individual inter-contact distances within the expected manufacturing range.

A.6.3 Quality Scoring

Master Quality Score:

$$\text{Score} = w_1 S_{\text{count}} + w_2 S_{\text{linearity}} + w_3 S_{\text{spacing}} \tag{A.95}$$

$$+ w_4 S_{\text{angles}} + w_5 S_{\text{length}} + w_6 S_{\text{entry}} \tag{A.96}$$

where w_i are weighting coefficients for different quality components, and $S_{\text{component}}$ represents individual scoring functions for trajectory characteristics.

Individual Components:

$$S_{\text{count}} = \begin{cases} 25 & \text{if exact match} \\ 20 - 2.5|d| & \text{if } |d| \leq 2 \\ 0 & \text{otherwise} \end{cases} \tag{A.97}$$

where S_{count} is the contact count score, d is the difference between detected and expected electrode contact numbers, and penalties are applied for deviations from manufacturer specifications.

$$S_{\text{linearity}} = \begin{cases} 20 & \text{if } L \geq 0.95 \\ 10 + 10\frac{L-0.85}{0.10} & \text{if } 0.85 \leq L < 0.95 \\ 10\max(0, \frac{L-0.70}{0.15}) & \text{otherwise} \end{cases} \tag{A.98}$$

where $S_{\text{linearity}}$ is the linearity score, L is the trajectory linearity measure from Equation (A.91), and thresholds reflect clinical acceptability criteria for electrode straightness.

A.6.4 Spline-Based Reconstruction

B-Spline Generation:

$$\mathbf{r}(t) = \sum_{i=0}^{n} \mathbf{P}_i B_{i,k}(t), \quad t \in [0, 1] \tag{A.99}$$

where $\mathbf{r}(t)$ is the reconstructed trajectory curve, $\mathbf{P}_i$ are control points derived from electrode positions, $B_{i,k}(t)$ are B-spline basis functions of degree k, and t is the parametric variable.

Quality Metrics:

$$\text{Linearity Score} = \max(0, 1 - 1.5\frac{\theta_{\max}}{180\check{\text{r}}} - 0.5\frac{\theta_{\text{mean}}}{60\check{\text{r}}}) \tag{A.100}$$

where $\theta_{\max}$ is the maximum curvature angle along the trajectory, θ_{mean} is the average curvature angle, and the scoring function penalizes excessive bending relative to clinical thresholds.

A.6.5 Clinical Integration Features

Interactive Visualization: 3D Plotly reports with color-coded quality indicators (Green $\geq$80, Orange 60-79, Red $<$60) and real-time annotation capabilities.

Multi-Modal Validation: Cross-validation with detection of the bolt head 6.7using alignment scoring.

Configuration Management: Hemisphere filtering, adaptive clustering, and clinical constraint validation.

Complete algorithmic implementations, parameter specifications, and clinical validation protocols are detailed in the 11-page technical documentation.