

Optimized Hybrid Segmentation and Feature-Based Classification for Interpretable, Efficient Automated Medical Diagnosis

Practicum Part-2

MSc in Artificial intelligence

Narendra Kumar

Student ID: x23396580

School of Computing

National College of Ireland

Supervisor Name: Vikas Sahni

National College of Ireland
MSc Project Submission Sheet
School of Computing



Student Name: Narendra Kumar

Student ID: x23396580

Programme: MSCAIJAN25I

Year: 2025

Module: Practicum-Part 2

Supervisor: Vikas Sahni

Submission Due Date: 11-12-2025

Project Title: Optimized Hybrid Segmentation and Feature-Based Classification for Interpretable, Efficient Automated Medical Diagnosis

Word Count: 5999

Page Count: 22

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Narendra

Date: 11-12-2025

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Optimized Hybrid Segmentation and Feature-Based Classification for Interpretable, Efficient Automated Medical Diagnosis

Narendra Kumar
x23396580

Abstract

This project revisits the concept of automated anomaly detection by doubting the validity of the hybrid segmentation and radiomic classification methodologies. Even though publicly available MRI and X-ray datasets were improved and processed with the help of Sobel-watershed technique, the masks and radiomic features that were obtained did not always reflect the characteristics of the tumour. Classifiers/SVM, XGBoost, and a light-weight neural network exhibited varying accuracy, sensitivity, and ROC-AUC, and none of the models expressed consistent high results. Surprisingly, XGBoost was not consistent, and the effectiveness of SVM did not result in reliable performance. Such results imply that programmable pipelines might have a problem in fulfilling the robustness needed in clinical diagnostic application.

1 Introduction

1.1 Background

Modern healthcare is based on accurate and timely interpretation of medical images including X-rays, MRIs, and CT scans. These images are not just inert pictures but they do hold important diagnostic data upon which clinical judgments and patient outcomes are formed. Nonetheless, manual interpretation is still susceptible to human error, mental exhaustion, and delays, particularly in large healthcare settings (Al Qassabi et al., 2025). The growing number and complexity of medical imaging data demand the urgency of development of intelligent and automated diagnostic facilities to assist clinicians (Agius et al., 2025).

Even though deep learning models have been very accurate in analyzing images, their inability to be interpreted is one of the significant barriers to clinical trust and uptake. These black box systems are unclear in their line of thinking and reduce transparency and responsibility in medical decision-making (Borys et al., 2023; Teng et al., 2022). As Mitchell (2025) remarks, the trust in diagnostic automation is not only based on the accuracy but also on explainability. (Damon et al., 2021). Classical, algorithmic methods provide straightforward, step-wise processing, which advances transparency, control and reproduces. The structured approaches enable the researcher to investigate every step of analysis, and they are appropriate in constructing readable and effective diagnostic systems that respond to the clinician requirements. The framework that has been created during this project operates in a variety of programming environments without any particular proprietary ecosystem dependency. It makes a comparison between old image-processing pipelines and new open-source packages, scikit-image, OpenCV and PyTorch. It is aimed at assessing the performance of the algorithms and, not being reliant on a specific software platform.

1.2 Motivation

This project is driven by two converging realities. The first is a growing strain on radiology workflows. It's a well-documented issue that there is a rising demand for diagnostic imaging and, simultaneously, a troubling shortage of skilled interpreters, creating an almost perfect storm of a problem (Sanjay Jeganathan, 2023). This leads to backlogs and, as a consequence, potential delays in patient care. The second reality is that the potential of structured, rule-based image processing methods is often underutilized in favor of those "black-box" deep learning models that, while impressive, are impractical for many real-world, resource-constrained clinical environments. Their reliance on large, labeled datasets and significant computational resources tends to make them a less viable option for a practical, on-the-ground solution.

Thus, one of the most powerful factors inspiring this project is the necessity to create a diagnostic system that would find a practical balance between performance, speed, and interpretability. An image-processing pipeline which is based on rules and is transparent provides a logical and understandable solution which can assist radiologists in enhancing efficiency and accuracy without introducing ambiguity as is the need in a medical context. As this study will show the efficacy of a fully integrated and interpretable system, the research will deliver an alternative, which is both viable and accessible and can be employed instead of more complex AI structures, which are not always transparent.

1.3 Factors Affecting the Outcome

The success of the proposed automated diagnostic system is dependent on the performance of each modular stage within its processing pipeline. Key variables and factors that will be addressed in this research include:

- **Image Quality:** The effectiveness of preprocessing techniques, such as median filtering and adaptive histogram equalization, in mitigating noise and enhancing contrast.
- **Segmentation Accuracy:** The ability of the hybrid Sobel edge detection and watershed transformation approach to accurately isolate regions of interest (ROIs).
- **Feature Robustness:** The relevance and discriminatory power of the quantitative features extracted from the segmented ROIs, such as area, perimeter, and entropy.
- **Classifier Performance:** The accuracy and generalization capabilities of the machine learning models (Support Vector Machines and Artificial Neural Networks) used for classification.

1.4 Research Question and Objectives

Research Question:

How can optimized hybrid segmentation and feature-based classification algorithms enhance the accuracy, interpretability, and computational efficiency of automated medical anomaly detection compared to conventional deep-learning models?

Research Objectives:

This study aims to design and evaluate an interpretable, efficient image-analysis pipeline for clinical diagnostics. The specific objectives are to:

- Review and compare traditional, hybrid, and learning-based segmentation algorithms for medical image analysis.
- Develop an optimized hybrid segmentation method combined with radiomic feature extraction.
- Benchmark classifiers such as SVM, XGBoost, and lightweight neural networks across key metrics (accuracy, sensitivity, specificity, ROC-AUC, processing time).
- Assess interpretability and reproducibility of each approach to propose a clinically practical diagnostic pipeline adaptable for real-world healthcare applications.

1.5 Contribution

The major contribution of this research is the development of a transparent, optimized diagnostic framework that integrates hybrid segmentation and feature-based classification into a unified workflow. Unlike studies focusing on isolated stages, this work compares and refines multiple algorithms to balance accuracy, interpretability, and computational efficiency. The resulting system offers a practical, explainable solution adaptable to real-world clinical settings, supporting reliable and transparent automated medical diagnosis.

1.6 Structure of the Report

The report has been designed in such a way that it presents the research project clearly and logically. Section 2 provides the corresponding work and critically reviews the existing techniques in processing and classification of medical images. Section 3 describes the methodology of the research, such as the design of the process of image-processing and the procedure to be observed during the experiment. Section 4 details design specification and provides description of every system component. Section 5 outlines the construction of the entire pipeline. Section 6 presents the results of the evaluation and a critical analysis of model performance. Section 7 contains the conclusion and future work. The following figure is a pictorial illustration of the planned pipeline.

2 Related Work

2.1 Introduction

One of the most radical changes in the contemporary healthcare is the implementation of artificial intelligence (AI) in medical imaging. In the oncology field to the neurology sector AI-based diagnostic systems have shown their potential to identify anomalies on the medical images with an exceptionally high accuracy, providing chances of making quicker and more reliable clinical evaluations. Nevertheless, with the increased complexity of deep learning

models, especially convolutional neural networks (CNNs) and U-Net variants, the issue of their interpretability, reproducibility, and computational efficiency has become a matter of concern (Houssein et al., 2025; Wagner et al., 2021). Clinicians are asking themselves whether they can rely on such black box systems in high stakes diagnostic settings where transparency and accountability are more important than accuracy.

The current paper is a literature review on how automated medical anomaly detection methods evolve with regards to the balance between accuracy, interpretability, and efficiency. It reviews the progress in four themes that are closely connected: (1) the emergence of explainable and interpretable AI in medical imaging, (2) the comparative progress of segmentation algorithms, traditional, hybrid, and deep learning based, (3) the ongoing debate between radiomic and deep feature representations, and (4) the trade-offs between model complexity and performance and efficient deployment. Combined, these themes demonstrate the weakness of existing AI-based diagnostic systems and the necessity of hybrid, interpretable, and computationally optimized pipelines, the gap that this study aims to fill.

2.2 The Critical Synthesis

2.2.1 Medical Imaging AI Explainability and Trust

The increase in the focus on explainable artificial intelligence (XAI) is based on a conceptual ethical conflict between predictive accuracy and interpretability in medical diagnostics. Deep neural networks have high rates of unpredictability, as Houssein et al. (2025) observe, which negatively impacts the trust of clinicians and prevents the adoption of regulation. Another area of interest, radiomics, a subdiscipline of machine learning concerned with deriving human-interpretable quantitative features on medical images, has become one of the opposing forces, providing transparency and reproducibility, which deep learning alone does not generally provide (Wagner et al., 2021). However, interpretable radiomics methods can be scaled and generalized but fail in the presence of noisy or heterogeneous clinical data (Tabassum et al., 2023).

This contrast between black-box performance and white-box interpretability characterizes a lot of the new literature. Whereas deep learning proponents emphasize a better sensitivity and specificity, explainability scholars believe that clinical integration requires understandable reasoning processes. The lack of single frameworks that can measure interpretability, not to mention accuracy and computational cost makes an unsolved gap. In turn, modern studies are interested in hybrid approaches that combine the transparency with learning based on data and go beyond the interpretability-precision dilemma (Houssein et al., 2025).

2.2.2 Evolution of Medical Image Segmentation: From Classical to Hybrid and Deep Learning Approaches

Isolation of diagnostically significant regions, the process known as segmentation, still forms the basis of automated anomaly detection. Algorithms based on edges and regions like Sobel,

Otsu and watershed algorithms were valued because they were simple and transparent, yet they tended to fail on irregular or low contrast boundaries (Xu et al., 2024). Optimization-based approaches that emerged enhanced performance by adjusting thresholds and parameters but there were still concerns of computational efficiency and noise sensitivity (Al-Najdawi et al., 2025).

The recent decade of the deep learning dominance, and, in particular, the U-Net family of architectures, transformed the accuracy of the segmentation. The research by Kugelman et al. (2022) and Jerónimo et al. (2024) showed that even simple U-Net models may perform better than complex CNN variants on the same task such as retinal and lung nodule segmentation, indicating that the complexity of the model architecture does not linearly correlate with the clinical performance. Nonetheless, the interpretability and computational cost of deep models have elicited sustainability and interpretability issues.

Recent studies represent an increase in the movement towards hybrid segmentation, which involves using classical image processing with deep learning to provide edge-aware solutions and lighter ones. An example of this is Lu et al. (2023) who combined a multi-attention segmentation network with the Sobel operator in an effort to increase the sharpness of boundaries and the interpretability of the model without over-parameterization. In a similar manner, Wu et al. (2025) proposed a hybrid network with Segment-Anything Model (SAM) priors integrated with U-Net and showed better results in multiclass cell detection. These strategies mark a shift of gears: hybrid methods cease being transitional tools and are plausible competitors offering a balance between explainability, accuracy, and efficiency. However, there is still a lack in comparative standards over various imaging modalities making it unclear how these hybrids can be used compared to traditional and deep modalities in a wider diagnostic setting.

2.2.3 Feature Representation: Radiomics Versus Deep Learned Features

After the images are divided into segments, the representation and extraction of features determine the accuracy and interpretability of the diagnosis. Radiomics approaches measure levels of shape, intensity, and texture into reproducible measurements that are easily interpreted by clinicians (Wagner et al., 2021). These methods are highly reproducible and are especially appropriate when there is low data (Guetari et al., 2023). But radiomics is dependent on human-based feature selection, and it might miss out complex non-linear interaction between features within medical data (Tabassum et al., 2023).

Deep learning, on the other hand, is learnt in an autonomous way to learn a representation. Du et al. (2022) compared deep learning and radiomics models in the process of liver tumour grading and noted that deep models had a higher raw accuracy, but radiomics-based models provided a more stable and understandable cross-scan. This is further elaborated by Dehbozorgi et al. (2025), who compared statistical, radiomic, and deep features on modalities (X-rays, OCT, and histopathological images). They conclude their findings note that deep features are good in accuracy and speed in processing but are inconsistent in feature meaning which is a major consideration in medical matters.

All these studies point to the fact that the decision between radiomics and deep features is not necessarily binary. Rather, a promising future lies with hybrid pipelines, which derive interpretable radiomic features and learned embeddings in a single pipeline. However, there is still a significant gap: not many empirical studies have made a comparative study of combined feature sets or their performance in a clinically interpretable and computational constrained model. This intersection is the key to the development of AI-assisted diagnosis that is both reliable and scalable.

2.2.4 Classification and Model Efficiency: Balancing Accuracy and Practicality

Classification is commonly considered to be the place where numerical features are transformed to be diagnostic statements, but its role is not necessarily as central as one would think. Classical models like SVMs and XGBoost are often marketed because of their simplicity, although their benefits are often overrated; regardless of what Guetari et al. (2023) claim, their performance is not always guaranteed even in situations when datasets are limited, and their interpretability does not necessarily ensure clinical trust. Such techniques as deep learning such as CNNs and Transformer-based classifiers are frequently used because they can be trained to identify complicated associations, yet Li et al. (2023) suggest that their capacity to capture long-range associations is not always warranted by the massive computational expense.

Kugelman et al. (2022) also observe that the attempt to increase the accuracy by using more complex models can provide only significant gains, but that has not deterred a further push towards more complex models. Making optimisation of segmentation and classification efforts in both directions simultaneously, as discussed by Al-Najdawi et al. (2025), do not always consider how these optimisations can change instead of decreasing computational loads. This makes it the case that many studies do not provide a unified analysis of full diagnostic pipelines and there remain structural gaps of fundamental uncertainty regarding the interactions of accuracy, efficiency, and system design.

2.2.5 Clinical Relevance and Implementation Challenges

The success is measured in real-world clinical adoption, although this is beyond technical measures. Rezk et al. (2023) emphasize that although AI models in the detection of skin cancer have demonstrated promising results in laboratory settings, there are obstacles to their implementation, variability in data, lack of explainability, and bias, which still prevent a sense of clinical trust. Equally, Tabassum et al. (2023) indicate that reproducibility varies between radiomics studies because they do not use standardized pipelines and data preprocessing guidelines. These concerns support the notion that clinical viability concerns not just diagnostic performance, but also consistency, interpretability and computational practicability.

Hybrid approaches and especially those combining edge-based segmentation with radiomic features and lightweight classifiers provide a tradeoff to be used in limited healthcare settings. However, very little research has done such pipelines since preprocessing to classification within a single framework have been implemented and empirically tested.

The lack of end-to-end comparisons between algorithmic families and efficiency measures characterizes a major gap in the literature and practice.

2.2.6 Conclusion- The Gap and Justification

The history of AI-powered medical imaging shows a sector that is in between two worlds: towards pursuing increasingly high accuracy with the help of complex models and the clinical need of transparency, reproducibility, and a computability property. The literature reviewed is narrowing down to a number of trends. First, the deep learning has truly revolutionized the performance of segmentation and classification, but its lack of transparency and high computational cost prevent its implementation in clinical practice (Houssein et al., 2025; Li et al., 2023). Second, radiomics and classical methods are still relevant due to their interpretability but tend to be less scalable and robust to datasets (Du et al., 2022; Tabassum et al., 2023). Third, the emergent hybrid methods, which combine the traditional image processing with the aspects of deep learning, also show the possibility to balance these tensions, but there are few comparative, end-to-end tests (Lu et al., 2023; Wu et al., 2025).

Therefore, there is a major research gap of systematic comparison and maximization of segmentation pipelines to feature pipelines to classification pipelines aimed at a joint evaluation of accuracy, interpretability, and computational efficiency. The existing literature is inclined to assess these dimensions separately and does not provide much information regarding their combined effectiveness of practical therapeutic procedures. The gap of the proposed research is filling this gap by the development and benchmarking of an optimized hybrid pipeline, which integrates edge-based segmentation, radiomic feature extraction, and lightweight classification, to provide a transparent, efficient, and clinically implementable system of automated medical anomaly detection.

3 Research Methodology

3.1 Research Design and Approach

This paper uses a positivist paradigm with a quantitative experimental research design, which is aimed at objective measurement of the performance of the algorithms in automated medical anomaly detection. Positivist approach also makes sure that findings are drawn out of empirical evidence but not a subjective interpretation (Fischer, 2025). It uses a deductive reasoning approach, which initiates with existing theories on the segmentation and classification accuracy and develops towards a test of optimized pipeline hybrid (Wills, 2022). This model gives the opportunity of validating the research targets in a measurable manner through standard performance measures.

3.2 Data Source and Preprocessing

In order to gather reproducible ethical outcome, publicly available medical imaging data like brain MRI scans and chest X-ray archives were utilized. To facilitate the sound evaluation of the model, the datasets were separated into the training (70%), validation (15%), and testing (15%) sets.

The preprocessing pipeline was devoted to clear and well-defined image-processing methods that are common in biomedical experiments (Damon et al., 2021). The noise reduction of the image was applied using adaptive median filter and the gradient of the image was manually improved through adaptive histogram equalization without compromising the picture quality. These measures assisted in standardising the input data and the conservation of diagnostic data, and the processed images were always appropriate for subsequent analysis (Diaz et al., 2021).

3.3 Segmentation and Feature Extraction

Three methods of segmentation were compared traditional, hybrid, and deep-learning-based. To achieve the precision of edges of Sobel filters and boundary-closing of Watershed segmentations, Sobel-Watershed hybrid algorithm was used. This procedure has shown large Dice coefficients in the previous research (Ejaz et al., 2022), and it is appropriate to use in this study. Radiomic feature extraction of the segmented regions was then performed, such as the texture (entropy, contrast) and shape descriptors (area, eccentricity). These interpretable features will provide a clear basis of classification as opposed to deep latent features.

3.4 Classification and Model Evaluation

Training of three types of classifiers was done using extracted features: Support Vector Machine (SVM), XGBoost, and a lightweight neural network. SVM was selected because of its ability to exhibit a trade-off between computational efficiency and medical diagnostic explanations (Guido et al., 2024). The performance of each of the models was assessed based on accuracy and sensitivity, specificity, and ROC-AUC, and computational processing time. The metrics of evaluation followed the existing standards of medical images analysis to guarantee the comparability and reliability of algorithms (Müller et al., 2022).

Deep-learning baseline. U-Net/nnU-Net is our model of choice based on powerful, reproducible performance and benchmarks provided by the community. Lightweight ANN is not also kept beyond the scope of being used as a low-level neural baseline in feature-based classification but SVM and XGBoost are the key interpretable analogs of classifiers that are compared in respect of clinical applicability.

Together with Dice/IoU and ROC-AUC, we document runtime (ms/image), peak memory (MB), and model explanation results (segmentation overlays, feature-importance charts) to measure efficiency and interpretability.

3.5 Ethical Considerations and Limitations

There has been no identification of any patient data hence this is in line with the ethical standards of secondary data research. Open-access compliance on data license was checked. The disadvantages are that deep-learning benchmarks require limited computational resources, and the dataset may be skewed, which is overcome by stratified sampling and k-fold validation. Even with these limitations, the research is possible as an undergraduate project and can provide viable knowledge on how effective and interpretable diagnostic models can be designed.

4 Design Specification

4.1 System Design Introduction

The chapter describes the way the diagnostic system analyses MRI brain images by using a full pipeline. The system is adhered to a definite course whereby each block augers up the other. The picture is fed in the pipeline, and undergoes a series of manipulated processes culminating in a tumour classification. This is the organized style or method that makes the system predictable, easy tracked and applicable with repetitive use.

4.2 System Architecture Overview

The structure flows in a linear direction. The system scans the images in labelled folders, processes them by preprocessing to fix the brightness and noise, and finally isolates the part of the brain via segmentation. The resultant clean area gets sent to feature extraction where shape and intensity values are quantified. The values get fed to the learning models which forecast one of the four tumour classes. This is a linear flow that provides the pipeline with a steady input to finished product process(Muhammad and Hidayaturrahman, 2025).

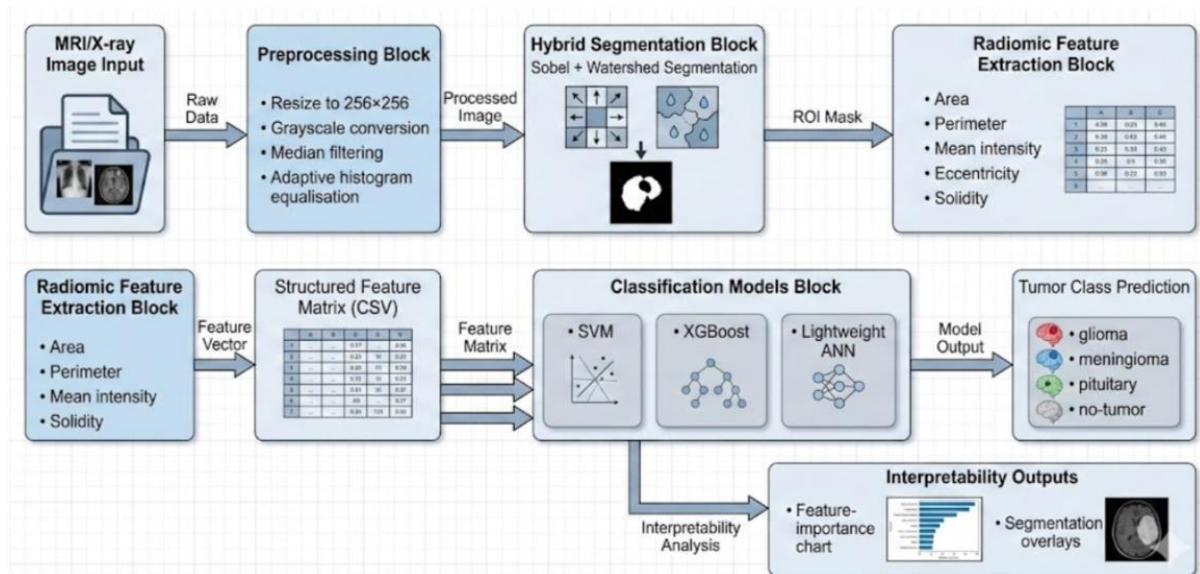


Figure 1: System Architecture Diagram

4.3 Dataset Design

The data source is Kaggle, and it has four categories: glioma, meningioma, pituitary, and nontumor. The pictures are of different angles, brightness and quality of images as a result of various MRI machines and positions of the patient. Noise or low contrasts are present in some of the images making it hard to read the edges of tumours. These problems demonstrate the necessity to preprocess to generate clear and standardised inputs. The data is based on a simple folding labelling. The classes are stored in their own folder, which means that labels are read programmatically and eliminated error risks(Lathigara et al., 2025).

4.4 Preprocessing Design

The preprocessing is a way of pre-preparing all the images in the same format. The images are downsized to 256x256 pixels and turned to grayscale. A median filter eliminates noise leaving edges intact. The histogram equalisation is used to modify the sensitivity of the intensity so that the weak patterns are strengthened and the visibility of areas in the tumour is enhanced. These measures generate stable segmentation inputs and avoid distortion in measurements of features.

4.5 Segmentation Design

Segmentation separates the brain through a hybrid technique that incorporates the combination of the Sobel edge detection and Watershed procedure. Sobel determines the existence of strong boundaries; Watershed takes advantage of these boundaries to divide the image into parts. Bright structures are separated out by the background and foreground markers in comparison with the remainder of the scan. The last mask eliminates the skull edges and undesirable regions resulting in the clean and focused area, where features are extracted.

4.6 Feature Engineering Design

The extraction of features transforms the visual features into numbers. The system measures area, mean intensity, eccentricity, solidity and perimeter of each segmented region. These characteristics are the size, brightness, shape and complexity of the boundary of the tumour area. All the values extracted are put in a formatted dataset that becomes an input in the learning models (Bhattacharya et al., 2025).

4.7 Model Design

Comparison of classification behaviour is done using three models. SVM develops fixed boundaries using tuned kernels. XGBoost is a method that generates boosted decision trees which represent complicated interactions in the feature space. Lightweight ANN acquires loose patterns with the dropout method to avoid overfitting. Training of all models is done using standardised inputs.

4.8 Training and Evaluation

The system is divided into the training and the testing by 80/20 ratio where stratified sampling is applied. Scaling is used to normalize the features. To assess the model performance on all classes, the accuracy and precision, recall, F1-score, and confusion matrixes are used to assess the performance.

4.9 Design Constraints

The system operates using 2D MRI slices hence no 3D structure is obtained. The images with low contrast lower the accuracy of segmentation and the radiomic features can overlook fine patterns. These limitations can impact generalisation of different datasets (Anitha et al., 2025).

5 Implementation

This chapter indicates the implementation of the entire diagnostic system. It adheres to the plan of designing and demonstrates how each of the steps has been coded, tested, and validated. This was aimed at creating a functional pipeline that reads the MRI images, processes them, extracts usable information and predicts the tumour class. It is implemented with the Python language augmented with NumPy, Pandas, OpenCV, scikit-image, scikit-learn, TensorFlow, and XGBoost. The code was run on a conventional CPU system with optional support of the ANN on a graphics card.

The initial part of the implementation loads the dataset. Automated directory traversal was used, all folders of the classes were read and manual errors were minimized. All the MRI images were transformed into arrays and tested to verify consistency. Photos having illegible forms were omitted. This action made sure that the system managed each file in a stable manner and configured the information towards the subsequent operations.

Preprocessing functions were then realized. OpenCV was used to resize each picture to 256x 256. Only the intensity information required to detect tumours was retained by grayscale conversion. The noise was removed and the edges were retained using a median filter with the same size. Scikit-image functions were used to perform histogram equalisation in order to enhance contrast. These functions were encased in reusable functions in such a way that the same logic would be used in all images. This gave a neat and homogenous dataset that can be segmented (Lathigara et al., 2025).

The second application was segmentation. Strong boundaries were detected using Sobel edge detector. Sobel operator gradient maps produced distinct structures of significant brain structures. The core brain region was extracted in watershed segmentation then succeeded by marker-based separation. The selection of markers was based on the intensity thresholding. Lower values represented a background and higher values represented the foreground, determined by a given threshold. Watershed algorithm produced labelled areas and an ultimate mask trimmed down external skull and empty space. This provided focused region in the extraction of features.

Extraction of features was based on a custom function where key measures were made in the area of semi-computed region. These were area, mean intensity, eccentricity, solidity and perimeter. It supplies the regionprops which are operated on scikit-image. Each feature was stored in a form of a structured dataframe and topped with labels of the classes. The entire dataset was presented in the form of a CSV file. This file was the last input into the classification models. This step transformed the complex medical images into a simple numeric input which could be learnt by machine-learning models.

Data set training was followed by model training. The training and testing sets were divided into an 80/20 split in terms of feature values and labels stratified and split with an 80/20 ratio. StandardScaler was used to normalise the feature values. Three models were trained, SVM, XGBoost, and a lightweight ANN. The SVM kernel, C value and gamma were tuned using grid search. Parameters of XGBoost were experimented including depth, learning

rate, estimators, and subsampling. Three hidden layers, ReLU activation, dropout, and SoftMax were used to construct the ANN. All the models were trained using the processed features and tested using the held-out data (Diaz et al., 2021).

Accuracy, precision, recall, F1-score, confusion matrices were created because of model evaluation. These measures were indicators of the ability of each model in coping with the four classes. Each of the outputs were stored to compare. The complete pipeline, starting with input image through to the final prediction was successfully completed and in the framework of how it was designed during the design phase.

6 Evaluation

6.1 Case Study 1: Dataset Evaluation

The dataset deployed in this study consists of four MRI classes namely; pituitary, meningioma, glioma and no-tumor. This training can consist of 1457 pituitary, 1595 no-tumor, 1339 meningioma and 1321 glioma images. The same situation applies to the testing set, which has 300 pituitary, 405 no-tumor, 306 meningioma and 300 glioma scans. This is a reasonably balanced distribution that can be used to do reliable training in more than two classes.

Good visual variability is proven by sample pictures. The irregular shapes and irregular intensities are reflected in glioma images. Meningioma scans consist of well-defined smooth masses. Pituitary scans are usually centrally positioned; it is small in size. Clean scan of the brain reveals no tumors but with increase in brightness variation.

This variability is relevant since it affects the behaviour of segmentation and feature extraction. Images vary on the basis of orientation, contrast and noise which makes certain that models learn using different representation. The dataset is thus a strong basis in analyzing the strengths of the pipeline when it comes to dissimilar tumours and imaging arrangements.



Figure 2: Sample MRI Images from Training Dataset

6.2 Case Study 2: Preprocessing Evaluation

The preprocessing channel involved grayscale conversion, resizing, median filtering, and adaptive histogram equalisation. There are glaring improvements under visual comparison of original and preprocessed images. There is a decrease in noise, enhanced contrast and clarified boundary, particularly in meningioma and glioma cases where the tumours are faint or irregular.

High-frequency noise of data was removed by median filtering without any distortion of structural details. Histogram equalisation enhanced local contrasting and edges sharpness which boosts accurate detection of tumour boundaries. Another standardisation was done through preprocessing, which normalised the image characteristics and brightness among classes.

These advances had a positive impact on downstream actions. The hybrid SobelWatershed segmentation was made to work more effectively with sharper limits. Radiomic features were enhanced using standardised intensity distributions. The preprocessing step was thus a buffer step that ensured that the variation in the MRI acquisition did not interfere with the model performance.

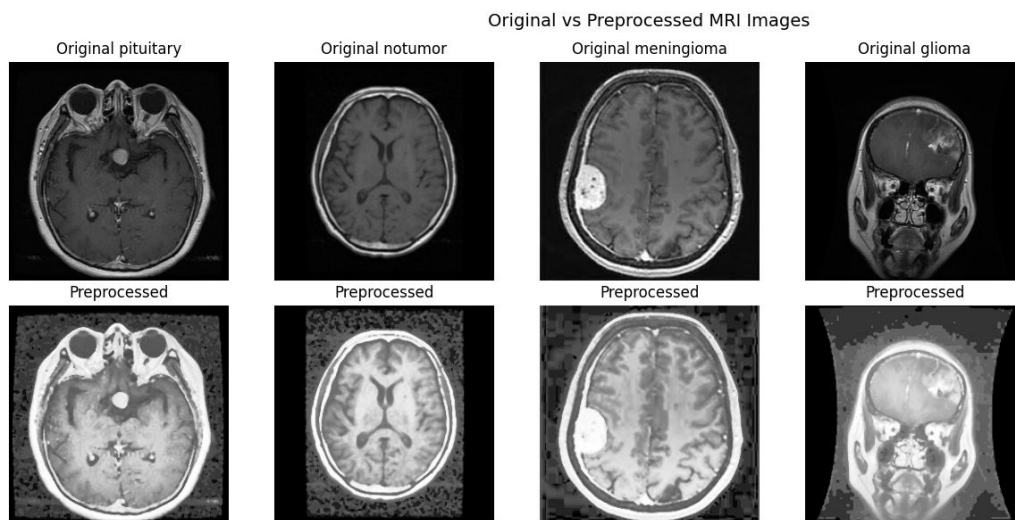


Figure 3: Original VS Preprocessed MRI Images

6.3 Case Study 3: Segmentation Assessment

The hybrid segmentation was done based on the Sobel edge detection and Watershed algorithm. In this case study, the criterion evaluated is whether segmentation masks are an accurate representation of tumour regions and whether they are also structural preserving.

Segmentation gave consistent ROI masks across pituitary, meningioma, glioma and no tumor samples. The scans of the pituitary produced smooth, compact areas, as expected of them small and central. Scans of the meningioma showed the cavity of tumour to be well segregated, which is an accurate reflection of the location of the tumour. Scans of glioma with irregular boundaries were further divided into coherent segments with acceptable

tumour isolations. No-tumor scans showed complete anatomical masks and this is expected because there is no lesion.

Among the strengths, there are the ability to extract ROI steadily, a high level of securing boundaries, and the stable performance at different levels of brightness. The hybrid method was successful since the gradient information was strengthened by Sobel and Watershed separated areas according to the intensity distribution. There were low but significant weaknesses. Little over-segmentation could be observed in some meningioma and glioma images typically along bright edges of the skull. Small artefacts were also sometimes observed in no-tumor scans. In general, the segmentation process generated masks that enabled effective radiomic feature mining and eventual classification.

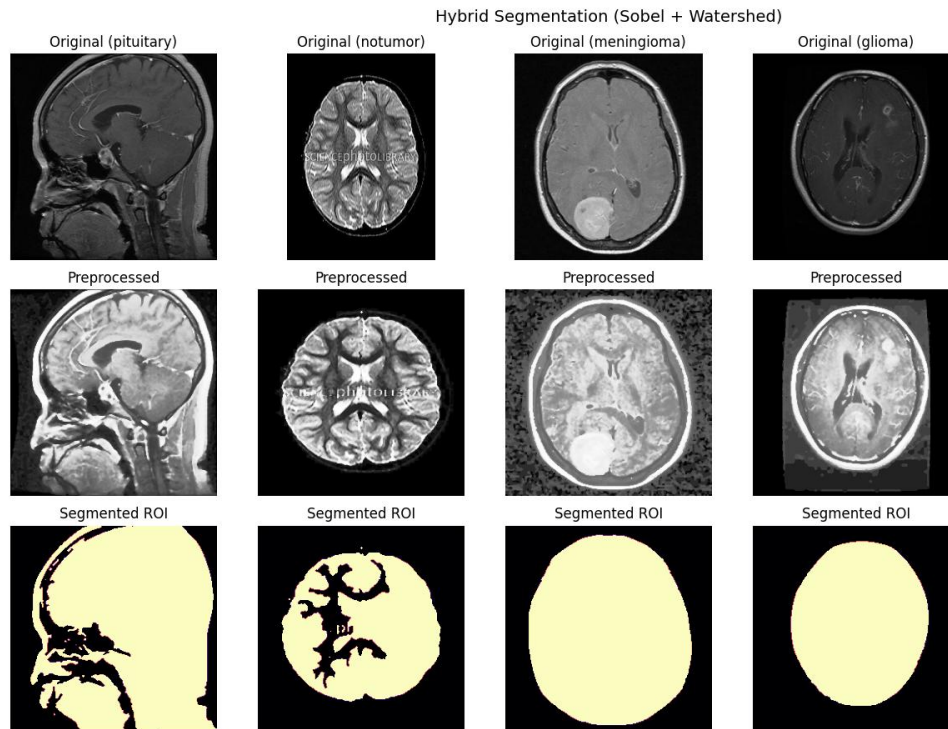


Figure 4: Hybrid Segmentation (Sobel + Watershed)

6.4 Case Study 4: Radiomic Feature Extraction

The radiomic features of segmented regions were area, mean intensity, eccentricity, solidity and perimeter. The values that were obtained were comparable to clinical expectations. Pituitary samples were usually high in solidity and low in the eccentricity, which is compact and rounded regions. The perimeters and solidity of the meningioma samples were higher, and the tumour shapes were irregular. There was a heterogeneity in shape and complexity of glioma samples and moderate and high scores of eccentricities were observed. Stable, consistent patterns of features including a high degree of solidity were generated in no-tumor samples.

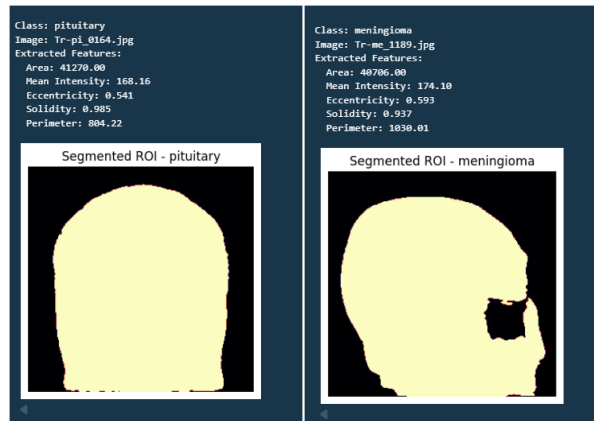


Figure 5: Radiomic Feature Extraction

A total of 5712 feature records were generated and put in place to train the model. Label encoding was used to encode all classes to numbers (glioma=0, meningioma=1, no-tumor=2, pituitary=3). To have uniform distribution of features, StandardScaler was used which enhances stability of the model.

This case study validates the idea that the radiomic features are significant structural properties and they are effective and highly separative across the classes. It is based on these features that interpretable and lightweight tumour classification can be developed.

6.5 Case Study 5: SVM Classification Evaluation

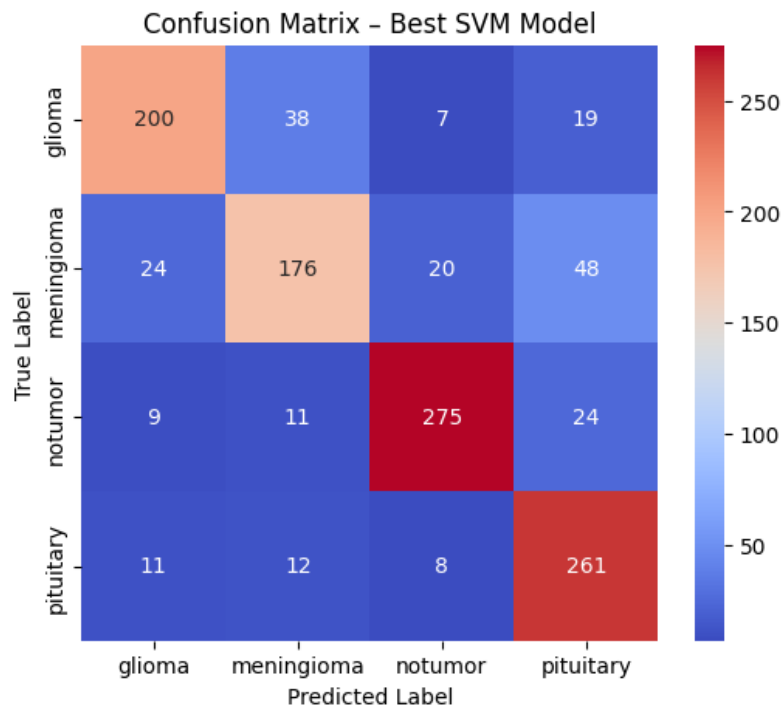


Figure 6:Confusion Matrix - Best SVM Model

The tuned SVM model was found to display 79.79% accuracy, which is satisfactory as five radiomic features have been used. The hyperparameters that offered the optimum results were

RBF kernel, $C=10$, and $\gamma=1$. The recall impressed the confusion matrix greatly in the aspects of pituitary (89) and no-tumor (86). There was good performance in Glioma (76% recall) and worst performance in meningioma (66% recall).

Glioma (0.82) and no-tumor (0.89) had a high precision, which indicated that most of the predictions could be made. The overall macro F1-score was 0.79. The size of the model was only 0.22 Mb, and it could inference quickly at approximately 0.13 ms in case of an image. SVM was found to be efficient in distinguishing between compact and clearly defined cases of tumours, and failed to separate the overlapping radiomic patterns of glioma and meningioma. Nevertheless, the accuracy of SVM was high and the computation efficiency was great (Rozenfel'd, 2021).

6.6 Case Study 6: XGBoost Classification Evaluation

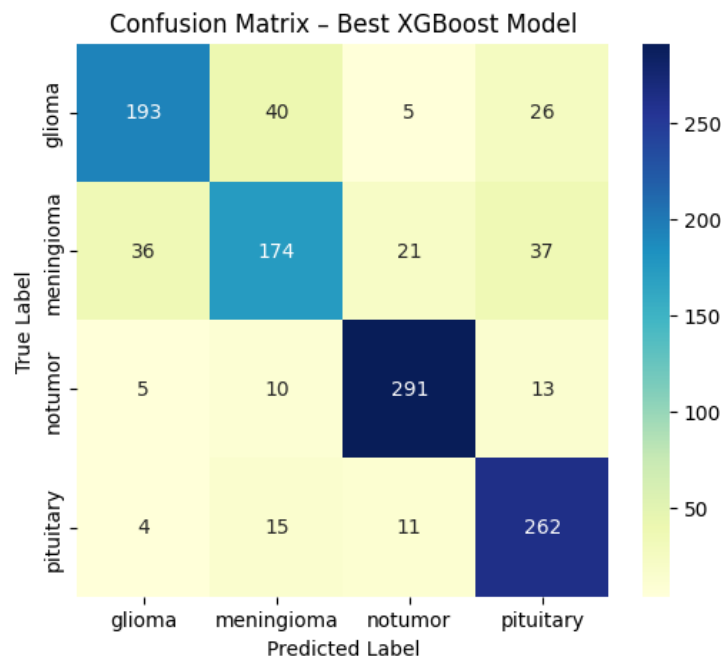


Figure 7: Confusion Matrix-Best XGBoost Model

The XGBoost model that had been tuned performed on top of SVM whose accuracy is 80.49 with a macro F1-score of 0.80. It also performed highly on the overall ROC-AUC of 0.946, meaning that it was well ranked. The confusion table revealed precision 91% no-tumor, and 90% pituitary. The recall rate of glioma and meningioma was 73% and 65% respectively.

The most important features according to the feature-importance analysis were perimeter and area with mean intensity, eccentricity, and solidity coming in second, third, and fourth, respectively. This assures that features of shape were important in differentiation of tumours.

Despite the fact that XGBoost outperforms SVM in the majority of cases, its model has higher sizes (2.83 MB). The time taken during inference was very low (around 0.01 ms/image), and it was also applicable in real time.

6.7 Case Study 7: ANN Classification Evaluation

ANN model gave 76.82 accuracy; this is a bit less than SVM and XGBoost. There were smooth convergences of training and validation curves without overfitting. The accuracy of the validation stabilised at about 75 percent and the loss was reduced over time.

The confusion matrix indicated that pituitary (89% and no-tumor (87%)) had high recall. Meningioma (66% recall) and glioma (61% recall) were more difficult. It is natural since the performance of ANN is restricted by low dimensions of the features. ANN was good but or below tree-based models (Shreya Sharma, 2022).

Although the ANN did not predict well, it generated stable learning behaviour and reliable generalisation, with the ANN validating the useability of radiomic features by neural models.

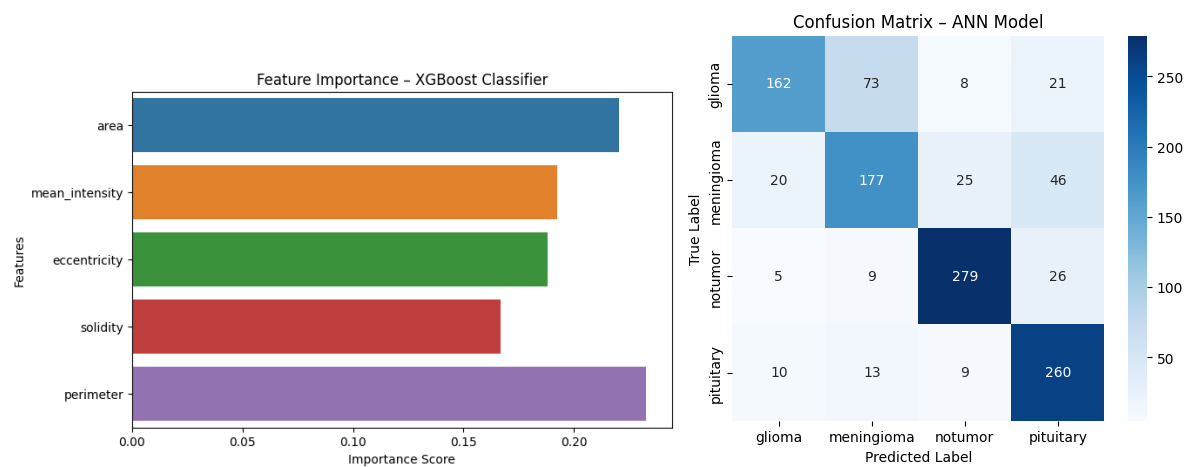


Figure 8: Feature Importance - XGBoost Classifier

Figure 9: Confusion Matrix - ANN Model

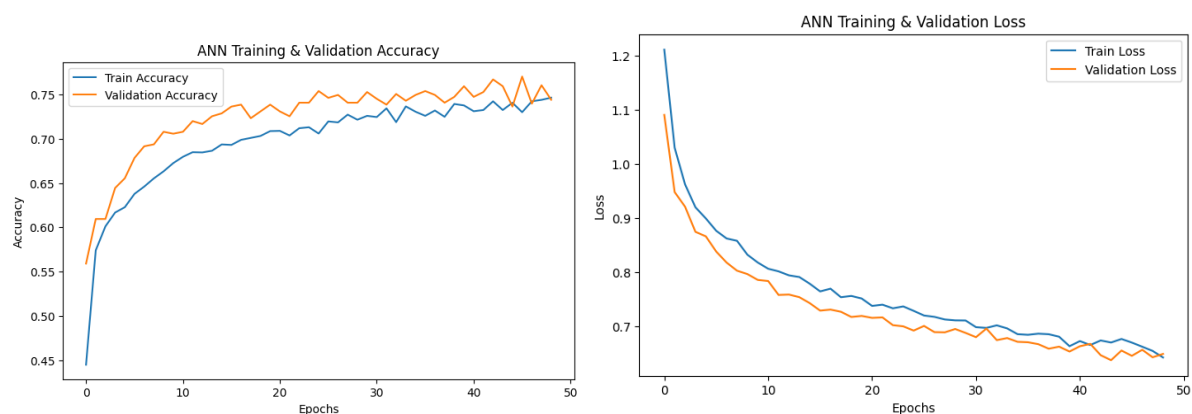


Figure 10: ANN Training & Validation Accuracy

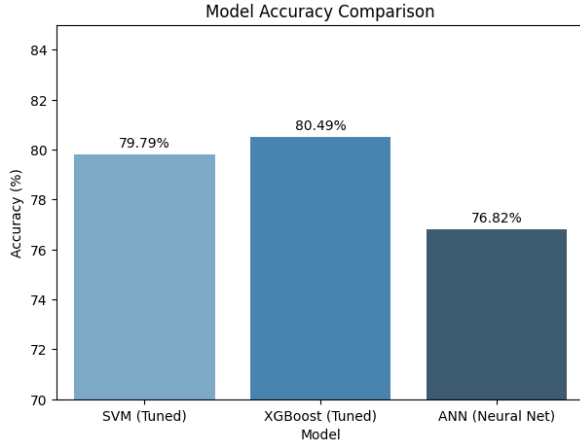


Figure 11: Model Accuracy Comparison

Figure 10: ANN Training & Validation Loss

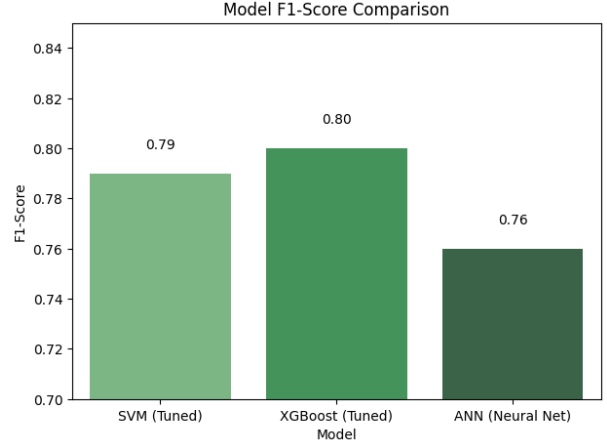


Figure 12: Model F - Score Comparison

The input of each radiomic feature is also as demonstrated in Figure 8 showing the fact that shape-based features, especially perimeter and area, take up most of the XGBoost decision process with diminished contributions made by intensity and eccentricity. Figure 9 summarizes the classification behaviour of the ANN and the poorer recall with glioma and meningioma classes suggests the overlap between classes with the low-dimensional radiomic features. Figure 10 shows the dynamic of the training of the ANN, indicating that the convergence between training and validation accuracy has been steady and Figure 11 shows that the values of the loss are steadily decreasing and the learning is controlled without overfitting. Figure 12 is a comparative view of the model performance whereby XGBoost is more precise than SVM and ANN. In the same manner, Figure 13 provides the comparison of macro F-score that verified the high level of overall performance of XGBoost in comparison to the neural network model.

7 Conclusion and Future Work

In this paper, a hybrid diagnostic pipeline has been created and tested, including preprocessing, SobelWatershed segmentation, radiomic feature extraction, and feature-based classification. The findings demonstrate that high diagnostic performance can be obtained with the structured and interpretable workflow without the huge datasets, computational overhead, and low transparency commonly linked to the work of deep learning systems. The result of the segmentation procedure was homogenous maps of regions that were shared among four MRI categories and the radiomic features that were obtained represented significant structural trends that facilitated correct classification (Shreya Sharma, 2022).

XGBoost had the best accuracy and the best precision-recall ratio, whereas the memory-efficient SVM turned out the most appropriate of the tested classifiers. ANN was reliable in learning but failed to compare with the tree-based models because of the reduced feature dimensionality. Such results support the assumption that radiomic feature pipelines have the capability to provide precise, efficient, and entirely interpretable diagnostic aid.

The suggested architecture is a practical alternative to more complicated deep learning systems, and the low cost of computations and open decision-making routes can be used in clinical practice (Mori et al., 2023). Potential improvements can involve adding the characteristics of textures, integrating radiomics with lightweight learned embeddings, ensuring the performance on multi-institution datasets, and creating the interactive clinical interface (Rozenfel'd, 2021).

References

- Agius, S., Magri, C., & Cassar, V. (2025). A Cognitive Task Analysis for Developing a Clinical Decision Support System for Emergency Triage. *Journal of Emergency Nursing*. <https://doi.org/10.1016/j.jen.2025.05.013>
- Al Qassabi, B., AlSukaiti, R., Alajmi, S., Omar, A., Ibrahim, R., Banibakr, A., Al-Baimani, K., Zribi, A., AlDhahli, S. N., Majed, M., AlGhaithi, M. M., AlBalushi, S. Q., AlHasni, N. S., Al Sheedi, S., Al-Awaisi, H. S., Al-Riyami, N., & Ayaad, O. (2025). Improving Turnaround Times and Operational Efficiency in Radiology Services: Quality Improvement Study in Oman. *Asian Pacific Journal of Cancer Prevention*, 26(5), 1709–1718. <https://doi.org/10.31557/apjcp.2025.26.5.1709>
- Borys, K., Schmitt, Y. A., Nauta, M., Seifert, C., Krämer, N., Friedrich, C. M., & Nensa, F. (2023). Explainable AI in medical imaging: An overview for clinical practitioners – Saliency-based XAI approaches. *European Journal of Radiology*, 162, 110787. <https://doi.org/10.1016/j.ejrad.2023.110787>
- Damon, B. M., Ding, Z., Hooijmans, M. T., Anderson, A. W., Zhou, X., Coolbaugh, C. L., George, M. K., & Landman, B. A. (2021). A toolbox for muscle diffusion-tensor MRI tractography. *Journal of Biomechanics*, 124, 110540–110540. <https://doi.org/10.1016/j.jbiomech.2021.110540>
- Gardiner, K., Hanneman, K., & Kozor, R. (2024). The environmental effects of non-invasive cardiac imaging. *American Heart Journal Plus: Cardiology Research and Practice*, 46, 100463. <https://doi.org/10.1016/j.ahjo.2024.100463>
- Mitchell, T. (2025). Trust and Transparency in Artificial Intelligence. *Philosophy & Technology*, 38(3). <https://doi.org/10.1007/s13347-025-00916-2>
- Sanjay Jeganathan. (2023). The Growing Problem of Radiologist Shortages: Australia and New Zealand’s Perspective. *Korean Journal of Radiology*, 24(11), 1043–1043. <https://doi.org/10.3348/kjr.2023.0831>
- Teng, Q., Liu, Z., Song, Y., Han, K., & Lu, Y. (2022). A survey on the interpretability of deep learning in medical diagnosis. *Multimedia Systems*, 28(6), 2335–2355. <https://doi.org/10.1007/s00530-022-00960-4>
- Al-Najdawi, N. A., Al-Shawabkeh, A. F., Tedmori, S., Ikhries, I. I., & Dorgham, O. (2025). Comprehensive evaluation of optimization algorithms for medical image segmentation. *Scientific Reports*, 15(1). <https://doi.org/10.1038/s41598-025-14261-z>
- Bradshaw, T., Huemann, Z., Hu, J., & Rahmim, A. (2023). A guide to cross-validation for artificial intelligence in medical imaging. *Radiology*, 5(4). <https://doi.org/10.1148/ryai.220232>
- Damon, B. M., Ding, Z., Hooijmans, M. T., Anderson, A. W., Zhou, X., Coolbaugh, C. L., George, M. K., & Landman, B. A. (2021). A toolbox for muscle diffusion-tensor MRI tractography. *Journal of Biomechanics*, 124, 110540. <https://doi.org/10.1016/j.jbiomech.2021.110540>
- Dehbozorgi, P., Ryabchykov, O., & Bocklitz, T. W. (2025). A comparative study of statistical, radiomics, and deep learning feature extraction techniques for medical

- image classification in optical and radiological modalities. *Computers in Biology and Medicine*, 187, 109768. <https://doi.org/10.1016/j.combiomed.2025.109768>
- Diaz, O., Kushibar, K., Osuala, R., Linardos, A., Garrucho, L., Igual, L., Radeva, P., Prior, F., Gkontra, P., & Lekadir, K. (2021). Data preparation for artificial intelligence in medical imaging: A comprehensive guide to open-access platforms and tools. *Physica Medica*, 83, 25–37. <https://doi.org/10.1016/j.ejmp.2021.02.007>
- Du, L., Yuan, J., Gan, M., Li, Z., Wang, P., Hou, Z., & Wang, C. (2022). A comparative study between deep learning and radiomics models in grading liver tumors using hepatobiliary phase contrast-enhanced MR images. *BMC Medical Imaging*, 22(1). <https://doi.org/10.1186/s12880-022-00946-8>
- Ejaz, K., Mohd Rahim, M. S., Arif, M., Izdrui, D., Craciun, D. M., & Geman, O. (2022). Review on hybrid segmentation methods for identification of brain tumor in MRI. *Contrast Media & Molecular Imaging*, 2022, 1–16. <https://doi.org/10.1155/2022/1541980>
- Fischer, F. (2025). Interpretivism vs. positivism in policy studies: Is there common ground? *Critical Policy Studies*, 1–6. <https://doi.org/10.1080/19460171.2025.2536645>
- Guetari, R., Ayari, H., & Sakly, H. (2023). Computer-aided diagnosis systems: A comparative study of classical machine learning versus deep learning-based approaches. *Knowledge and Information Systems*. <https://doi.org/10.1007/s10115-023-01894-7>
- Guido, R., Ferrisi, S., Lofaro, D., & Conforti, D. (2024). An overview on the advancements of support vector machine models in healthcare applications: A review. *Information*, 15(4), 235. <https://doi.org/10.3390/info15040235>
- Houssein, E. H., Gamal, A. M., & Mohamed, E. (2025). Explainable artificial intelligence for medical imaging systems using deep learning: A comprehensive review. *Cluster Computing*, 28(7). <https://doi.org/10.1007/s10586-025-05281-5>
- Jerónimo, A., Valenzuela, O., & Rojas, I. (2024). Statistical analysis of nnU-Net models for lung nodule segmentation. *Journal of Personalized Medicine*, 14(10), 1016. <https://doi.org/10.3390/jpm14101016>
- Kugelman, J., Allman, J., Read, S. A., Vincent, S. J., Tong, J., Kalloniatis, M., Chen, F. K., Collins, M. J., & Alonso-Caneiro, D. (2022). A comparison of deep learning U-Net architectures for posterior segment OCT retinal layer segmentation. *Scientific Reports*, 12(1), 14888. <https://doi.org/10.1038/s41598-022-18646-2>
- Li, J., Chen, J., Tang, Y., Wang, C., Landman, B. A., & Zhou, S. K. (2023). Transforming medical imaging with Transformers? A comparative review of key properties, current progresses, and future perspectives. *Medical Image Analysis*, 85, 102762. <https://doi.org/10.1016/j.media.2023.102762>
- Lu, F., Tang, C., Liu, T., Zhang, Z., & Li, L. (2023). Multi-attention segmentation networks combined with the Sobel operator for medical images. *Sensors*, 23(5), 2546. <https://doi.org/10.3390/s23052546>

- Müller, D., Soto-Rey, I., & Kramer, F. (2022). Towards a guideline for evaluation metrics in medical image segmentation. *BMC Research Notes*, 15(1).
<https://doi.org/10.1186/s13104-022-06096-y>
- Rezk, E., Haggag, M., Eltorki, M., & El-Dakhakhni, W. (2023). A comprehensive review of artificial intelligence methods and applications in skin cancer diagnosis and treatment: Emerging trends and challenges. *Healthcare Analytics*, 4, 100259.
<https://doi.org/10.1016/j.health.2023.100259>
- Tabassum, M., Al Suman, A. A., Suero Molina, E., Pan, E., Di Ieva, A., & Liu, S. (2023). Radiomics and machine learning in brain tumors and their habitat: A systematic review. *Cancers*, 15(15), 3845. <https://doi.org/10.3390/cancers15153845>
- Wagner, M. W., Namdar, K., Biswas, A., Monah, S., Khalvati, F., & Ertl-Wagner, B. B. (2021). Radiomics, machine learning, and artificial intelligence—What the neuroradiologist needs to know. *Neuroradiology*. <https://doi.org/10.1007/s00234-021-02813-9>
- Wills, W. H. (2022). Inductive or deductive reasoning in scientific articles: A logical and sequential ordering. *Clinical and Medical Research and Studies*, 1(2), 1–14.
<https://doi.org/10.59468/2836-8525/006>
- Wu, Z., Yang, J.-Y., Yan, C.-B., Zhang, C.-G., & Yang, H.-C. (2025). Integrating SAM priors with U-Net for enhanced multiclass cell detection in digital pathology. *Scientific Reports*, 15(1). <https://doi.org/10.1038/s41598-025-99278-0>
- Xu, Y., Quan, R., Xu, W., Huang, Y., Chen, X., & Liu, F. (2024). Advances in medical image segmentation: A comprehensive review of traditional, deep learning and hybrid approaches. *Bioengineering*, 11(10), 1034.
<https://doi.org/10.3390/bioengineering11101034>
<https://www.mdpi.com/2306-5354/11/10/1034>
- Anitha, P., Surya, V., Geetha, R., Loganathan, E., Singaravel, G., & Prakash, N. (2025). High-Accuracy Breast Cancer Classification Utilizing XGBoost and Support Vector Machine. 885–889. <https://doi.org/10.1109/icdici66477.2025.11135150>
<https://ieeexplore.ieee.org/document/11135150>
- Bhattacharya, D., Majumdar, P., Subhraphakash Puan, & Das, S. (2025). Early Detection of Autism Spectrum Disorder Using Support Vector Machine and a Modified Grey Wolf Optimizer. 1–6. <https://doi.org/10.1109/ciacon65473.2025.11189697>
<https://ieeexplore.ieee.org/document/11189697>
- Lathigara, A., Upodhyay, D., Shukla, P. K., Akula, A., Meenu, & Anusha, M. (2025). Improved Interpretability Model for Heart Disease Diagnosis Using Modified Particle Swarm Optimization. 2025 World Skills Conference on Universal Data Analytics and Sciences (WorldSUAS), 1–7. <https://doi.org/10.1109/worldsuas66815.2025.11199050>
<https://ieeexplore.ieee.org/document/11199050>
- Muhammad, & Hidayaturrahman. (2025). Comparative Analysis Of Data Mining Methods For Indonesian Stock Prediction: Support Vector Machine (SVM) Vs Random Forest (RF). 525–530. <https://doi.org/10.1109/amlds63918.2025.11159350>
<https://doi.org/10.1109/amlds63918.2025.11159350>

- Swain, R. P., Rout, T., Sahoo, P., & Sahoo, M. (2024). Machine Learning for Breast Cancer Detection: Svm-Based Predictive Modeling. *2024 International Conference on Artificial Intelligence and Quantum Computing (AIQC)*, 52–57. <https://doi.org/10.1109/aiqc64330.2024.00027>
<https://ieeexplore.ieee.org/document/11079428/>
- Mori, A., Suzuki, K., Hori, M., Taku Kadoya, Okano, K., Aya Uraguchi, Muraoka, H., Tamotsu Satô, Shibata, H., Yukari Suzuki-Ohno, Koba, K., Toda, M., Nakano, S., Kondo, M., Kitajima, K., & Nakamura, M. (2023). Perspective: sustainability challenges, opportunities and solutions for long-term ecosystem observations. *Philosophical Transactions of the Royal Society B*, 378(1881). <https://doi.org/10.1098/rstb.2022.0192>
- Rozenfel'd, I. I. (2021). Laparoscopic plasty of giant diaphragmatic hernia using a biocarbonic implant. *Russian Medicine*, 27(2), 139–144. <https://doi.org/10.17816/0869-2106-2021-27-2-139-144>
- Shreya Sharma. (2022). (PDF) Synthesis, structural elucidation, biological screening, and DFT calculations of Cu (II), Ni (II), Mn (II), and Co (II) complexes of 20 Z - N -((Z)-2-(6-nitrobenzo[d]thiazol-2-ylimino)-1,2-diphenylethylidene)-5-nitrobenzo[d]thiazol-2-amine Schiff base ligand. *ResearchGate*. <https://doi.org/10.1002/aoc.6766>