National College of Ireland

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

Practicum Part-2

MSc in Artificial intelligence

Narendra Kumar

Student ID: 23396580

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: 884 **Page Count**: 3

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: 09-12-2025

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):	

Configuration Manual

Narendra Kumar
x23396580

1 Section 1: System Architecture Overview

This section details the design and flow of the diagnostic system, which operates as a sequential, modular pipeline. The architecture is designed to prioritize **interpretability** and **computational efficiency** by utilizing classical image processing and feature-based classification.

1.1 System Flow (Pipeline)

The system executes a linear process to transform a raw MRI image into a final tumor classification:

1. **Input:** Reads MRI slices from labeled folders.
2. **Preprocessing:** Standardizes image format, contrast, and noise.
3. **Segmentation:** Isolates the Region of Interest (ROI) using a hybrid approach.
4. **Feature Extraction:** Quantifies structural and intensity properties of the ROI.
5. **Classification:** Predicts one of the four tumor classes using a tuned machine learning model.

2 Section 2: Data Handling and Preprocessing Configuration

2.1 Data Source and Description

Parameter	Configuration
Source	Kaggle
Modality	Brain MRI Scans
Classes	Glioma, Meningioma, Pituitary, No-Tumor
Data Split	80% Training / 20% Testing
Training Records	5712 feature records
Testing Records	~1311 images (e.g., Pituitary: 300, No-tumor: 405, Meningioma: 306, Glioma: 300)

Parameter	Configuration
Sampling Method	Stratified Sampling
Input Feature Count	5 radiomic features

2.2 Preprocessing Pipeline Parameters

The Preprocessing Pipeline: It is a sequence of reusable functions that generate a homogeneous and standardized dataset, which is essential to develop a strong analysis of medical images (Diaz et al., 2021). The logic can be applied to all images, which will allow one to maintain consistency and reduce the number of errors in the input data (Lathigara et al., 2025). The preprocessing is crucial to enhance reproducibility, which is a problem of radiomics research (Tabassum et al., 2023). The pictures are processed by four consecutive functions: they are first Resized to 256 x 256 pixels in order to make them input-standardized, and converted to Grayscale to store only the essential intensity data. The high-frequency noise is removed using an Adaptive Median Filter in order to deal with noise, but maintain structural edges. Lastly, Adaptive Histogram Equalization is used to increase the local contrast and sharpen blurred edges to produce clear and standardized inputs to the hybrid segmentation module.

3 Section 3: Segmentation and Feature Configuration

In the pipeline, there is a two-step procedure that starts with Segmentation and then Radiomic Feature Extraction. A Hybrid Sobel-Watershed algorithm was used to isolate the Region of Interest (ROI) and is an improved variant of classical segmentation methods (Xu et al., 2024). This technique was adopted to precisely outline the tumor boundaries, which has been already linked to high accuracy in the definition of brain tumors (Ejaz et al., 2022). The hybrid algorithm requires Sobel operator to identify strong edges first and Watershed procedure to use the gradient data to divide the image into clean and focused areas that will be further analyzed. After the isolation of the ROI, the Feature Extraction module transforms the visual properties into a five interpretable radiomic feature, namely the area, mean intensity, eccentricity, solidity, and perimeter (Wagner et al., 2021). Radiomics also offers the necessary quantitative values that are easy to reproduce and understand among clinicians, and it can be used to address the lack of transparency that is seen with deep learning models (Tabassum et al., 2023). These last numerical feature vectors are put to standard to the StandardScaler in order to have a uniform distribution, a requirement of the stability and performance of the classification models.

4 Section 4: Model Configurement and Performance Benchmarking

Three models with features, namely, XGBoost, Support Vector Machine (SVM), and a small Artificial Neural Network (ANN) were compared to identify the most viable and high-performing model to be used in the

pipeline. The XGBoost Classifier was the best set-up as it performed best on the test set with an accuracy of 80.49% and a Macro ROC-AUC of 0.946. This high performance shows that the model can model non-linear interactions between features which are complex (Anitha et al., 2025). Moreover, it was highly efficient, and the inference time was about 0.01 ms/image. The analysis of the feature importance established that the perimeter and area were the most important features used to distinguish between the types of tumors. Compared to it, the optimized SVM model, which was set with the RBF kernel (C=10, γ=1) had an accuracy of 79.79% and the lowest memory footprint of 0.22 Mb, which corresponded with the objective of resources-constrained environment (Guido et al., 2024). The accuracy was lowest with the lightweight ANN at 76.82, indicating that the model was constrained by dimensionality of radiomic feature set. The XGBoost Classifier is preferred to be used because of the high balance of diagnostic accuracy and computing speed in the ultimate deployment.

Comparative Performance Metrics (on Test Set)

Metric	**XGBoost (Optimal)**	**SVM (Optimal)**	**ANN (Lightweight)**
Accuracy	**80.49%**	79.79%	76.82%
Macro F1-Score	**0.80**	0.79	N/A (implied lower)
ROC-AUC	**0.946**	N/A (implied lower)	N/A (implied lower)
Inference Time	**~0.01 ms/image**	~0.13 ms/image	N/A (implied higher than XGB)
Model Size	2.83 MB	**0.22 Mb**	N/A (implied larger than SVM)
Optimal Hyperparameters	Tuned for best score	RBF Kernel, C=10, Gamma=1	3 Hidden Layers, ReLU, SoftMax

Conclusion: The **XGBoost Classifier** is the recommended model for the final diagnostic pipeline due to its superior accuracy (80.49%) and exceptional inference speed (~0.01 ms/image), providing the best practical balance between performance and computational efficiency. The SVM remains a viable alternative if model size is the absolute priority (0.22 Mb).

5 References

1. Anitha, P., Surya, V., Geetha, R., Loganathan, E., Singaravel, G. and Prakash, N. (2025) 'High-Accuracy Breast Cancer Classification Utilizing XGBoost and Support Vector Machine', in *Proceedings of the ICDICI Conference*, pp. 885–889. Available at: https://doi.org/10.1109/icdici66477.2025.11135150 (Accessed: 9 December 2025).
2. Ejaz, K., Mohd Rahim, M. S., Arif, M., Izdrui, D., Craciun, D. M. and Geman, O. (2022) 'Review on hybrid segmentation methods for identification of brain tumor in MRI', *Contrast Media & Molecular Imaging*, 2022, pp. 1–16. Available at: https://doi.org/10.1155/2022/1541980 (Accessed: 9 December 2025).
3. Guido, R., Ferrisi, S., Lofaro, D. and Conforti, D. (2024) 'An overview on the advancements of support vector machine models in healthcare applications: A review', *Information*, 15(4), p. 235. Available at: https://doi.org/10.3390/info15040235 (Accessed: 9 December 2025).
4. Lathigara, A., Upodhyay, D., Shukla, P. K., Akula, A., Meenu and 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)*, pp. 1–7. Available at: https://doi.org/10.1109/worldsuas66815.2025.11199050 (Accessed: 9 December 2025).
5. Tabassum, M., Al Suman, A. A., Suero Molina, E., Pan, E., Di Ieva, A. and Liu, S. (2023) 'Radiomics and machine learning in brain tumors and their habitat: A systematic review', *Cancers*, 15(15), p. 3845. Available at: https://doi.org/10.3390/cancers15153845 (Accessed: 9 December 2025).
6. Wagner, M. W., Namdar, K., Biswas, A., Monah, S., Khalvati, F. and Ertl-Wagner, B. B. (2021) 'Radiomics, machine learning, and artificial intelligence—What the neuroradiologist needs to know', *Neuroradiology*. Available at: https://doi.org/10.1007/s00234-021-02813-9 (Accessed: 9 December 2025).
7. Xu, Y., Quan, R., Xu, W., Huang, Y., Chen, X. and Liu, F. (2024) 'Advances in medical image segmentation: A comprehensive review of traditional, deep learning and hybrid approaches', *Bioengineering*, 11(10), p. 1034. Available at: https://doi.org/10.3390/bioengineering11101034 (Accessed: 9 December 2025).