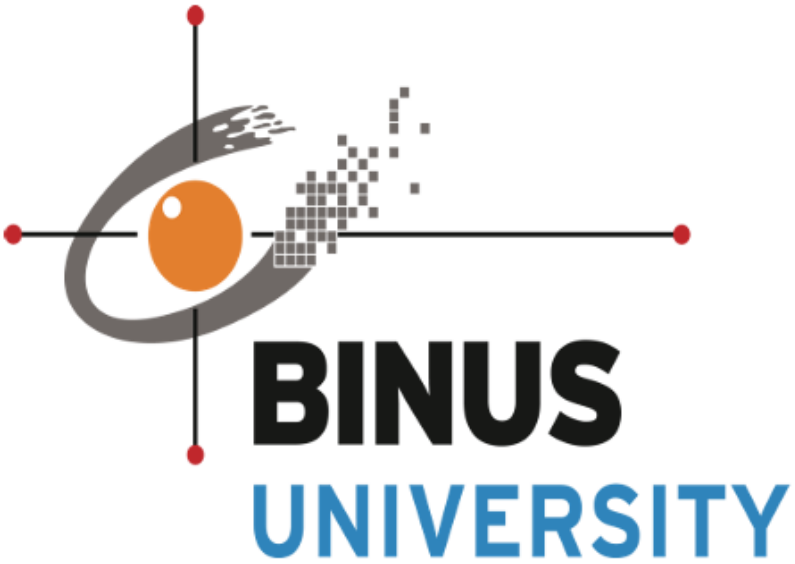
Final Project Computer Vision

Brain Tumor Classification on Low Resolution Images



Kelas: LC01

Kelompok ??:

Jason Evansaputra Sudargo - 2702243722

Moeh Adji Anggalaksana - 2702404323

Jason Prawira Tirtadjaja - 2702257500

UNIVERSITAS BINA NUSANTARA

SCHOOL OF COMPUTER SCIENCE

2025/2026

Abstract

This project develops a computer-vision system to classify brain MRI slices into healthy or some tumor types classes using handcrafted texture features (GLCM and LBP) and the best machine learning model between Random Forest, Support Vector Machine (SVM), and XGBoost for classification. The pipeline includes dataset acquisition, preprocessing (resize, grayscale, histogram equalization, denoising, sharpening, normalization), feature extraction (quantized GLCM properties and uniform LBP histograms), feature scaling, model training, hyperparameter tuning, and evaluation on a held-out test set. Key contributions are a reproducible preprocessing pipeline, a lightweight explainable feature set suitable for limited-data regimes, and a Streamlit demo for interactive inference. The result shows SVM as a better model for this case with Accuracy: 90%; Precision: 90%; Macro F1: 89%; Recall: 89%. The system is intended for research/educational use and not for clinical diagnosis.

Keywords— Machine Learning, Tumor, Texture-based feature detection, XGBoost, Random Forest, SVM, Tumor Classification, Model Comparison, Model Evaluation

I. Introduction

Brain tumor classification using MRI has become an essential area of research in medical imaging and clinical decision support. Accurate and early identification of tumor types can significantly improve treatment planning, prognosis assessment, and patient outcomes. However, MRI scans, especially obtained at a lower resolution often suffer from noise, contrast inconsistency, and overall image pollution, contributing to the already difficult manual imaging processes. To address these challenges this project investigates a texture-based machine learning pipeline that leverages handcrafted features combined with classical classifiers. These methods remain valuable in scenarios with limited data, constrained computational resources, or the need for explainability.

This study focuses on extracting discriminative texture descriptors from preprocessed MRI slices using "Gray Level Co-occurrence Matrix (GLCM) features and Local Binary Patterns (LBP). These features are then evaluated across three widely used non-deep learning classifiers: Support Vector Machine (SVM), Random Forest, and XGBoost. The goal is to determine which model best captures the structural patterns associated with healthy and tumor-affected brains. The result is a system that is designed to be reproducible, efficient, and adaptable for academic research.

II. Related Works

A. Texture-based MRI Classification

Traditional computer vision research has explored the use of hand made texture descriptors for medical imaging analysis for a very long time[1]. GLCM features, such as contrast, homogeneity, energy, and correlation are often used due to their viability to quantify spatial relationships between pixel color and/or intensities. Similarly, LBP has proven very effective for characterizing micro-texture variations, making it suitable for identifying irregular tissue structures in MRI scans. Previous studies have demonstrated that combining multiple texture descriptors can significantly improve model performance, especially when dealing with grayscale, low-resolution, or low-contrast images.

B. Deep Learning Approaches

In recent years, Convolutional Neural Networks (CNNs) and transformer based architectures have dominated brain tumor classification tasks, achieving substantial and relatively positive results by learning hierarchical feature representation directly from the data. However, these models typically require large annotated datasets, high computational resources, and complex hyperparameter tuning processes[2]. Beyond that, their "black box" nature often limits interpretability, which can hinder clinical acceptance. Despite their superior performance, deep learning methods may not be optimal in settings where data availability is limited or where transparent decision making by a human hand is prioritized.

C. Previous Work in Brain Tumor Classification

Existing literature shows successful attempts at

differentiating tumor types such as glioma, meningioma, and pituitary tumors using a variety of imaging techniques and classification strategies[2]. Many classical machine learning works focus on hand crafted features, feature selection techniques, and ensemble methods to improve predictive accuracy. Meanwhile, deep learning studies often rely on transfer learning or 3D CNNs to extract volumetric information from MRI sequences[3]. However, there are still gaps in studies focusing on low-resolution MRI classification using lightweight feature-based models, motivating the current project's emphasis on interpretability and computational efficiency.

III. METHODOLOGY

This chapter describes the materials and methods utilized to compare the performance of each model architecture in classifying brain tumor types. The methodology involves selecting a suitable dataset, preparing the dataset through preprocessing, and selecting the best models designed for stacking.

A. Dataset

The dataset used in this research is publicly available and can be obtained from Kaggle. It contains four classes, glioma, meningioma, pituitary tumor, and healthy brains. Each sample consists of a single MRI Slice with substantial variation in resolution, illumination, and varying noise levels. These characteristics make the dataset an ideal candidate for assessing the usefulness of texture-based classification methods. The dataset is divided into training, validation, and test sets using random sampling to preserve class distribution across partitions.

B. Data Preprocessing

A standardized preprocessing pipeline was applied to all images to ensure consistency and improve feature extraction quality. The transformations included:

1. **Resizing** all images to a uniform resolution for better comparability.
2. **Grayscale conversion**, as texture features rely purely on intensity variations.
3. **Histogram equalization** to enhance contrast and reveal subtle texture patterns.
4. **Denoising** using Gaussian filtering to reduce high-frequency noise artifacts.
5. **Sharpening** to emphasize edges and structural boundaries within tumor regions.
6. **Normalization** of intensity values to stabilize feature distribution.

This sequence produced cleaner and more standardized images, enabling more reliable extraction of texture descriptors.

C. Modelling

Two groups of handcrafted features being GLCM and LBP were extracted from the preprocessed images. GLCM properties were computed using quantized gray levels and multiple directional offsets, while LBP histograms were generated with similar patterns to reduce redundancies. The combined feature sets were standardized using z-score normalization before being fed to each model.

Three Machine Learning models were evaluated:

1. Support Vector Machine (SVM) with RBF kernel.
2. Random Forest with tuned depth and estimator patterns.
3. XGBoost with optimized tree-based parameters.

Each model went through hyperparameter tuning using grid search or bayesian optimization, depending on the model complexity.

D. Model Evaluation

Model performance was assessed using a held-out test set. The evaluation metrics included:

1. Accuracy
2. Precision
3. Recall

4. F1-Score

These metrics were chosen to identify class imbalance and to measure the model's ability to generalize across all tumor categories. Confusion matrices and classification reports were also analyzed to identify misclassification trends and strengths of each model.

IV. Results and Discussions

The experimental results indicate that SVM achieved the highest overall performance, outperforming Random Forest and XGBoost across most metrics. SVM obtained an outstanding accuracy of 90%, precision of 90%, recall of 90%, and Macro F1-Score of 89%. This suggests that the RBF kernel effectively captured the nonlinear relationships present in the texture descriptors.

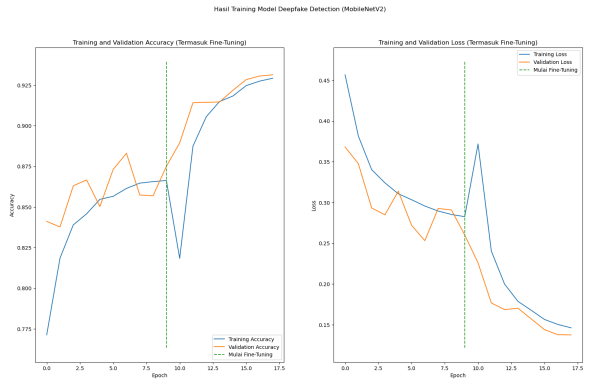
SVM Accuracy

```
Accuracy SVM Tuned: 0.9000762776506483
Recall SVM Tuned: 0.8918046477850399
Precision SVM Tuned: 0.8983875869476236
F1 SVM Tuned: 0.8917303228517206
```

XGBoost accuracy

```
Accuracy XGBoost Tuned: 0.8977879481311976
Recall XGBoost Tuned: 0.889720406681191
Precision XGBoost Tuned: 0.8962506085036279
F1 XGBoost Tuned: 0.8891561881340004
```

In contrast, Random Forest and XGBoost, while still performing very well, appeared less capable of modeling the fine-grained texture variations inherent in low-resolution MRI images. The confusion matrices revealed that misclassifications primarily occurred between structurally similar tumor types, such as glioma, and meningioma, consistent with observations from prior literature.



This graph shows the training loss and accuracy of the models.

These findings demonstrate that lightweight, feature-based models can remain competitive in environments with limited data and without reliance on deep learning architectures. The results highlight the importance of strong preprocessing and carefully engineered features in classical computer vision pipelines.

V. Conclusion

This project successfully developed a complete texture-based classification system for brain tumor MRI images through machine learning advancements. The combination of GLCM and LBP descriptors, when paired with effective preprocessing and feature scaling, produced strong discriminative performance. Among the evaluated models, SVM demonstrated superior classification capability, confirming its suitability for tasks requiring nonlinear boundary modeling.

Future improvements may include an expansion to the dataset, including multi-slice volumetric information, or experimenting with hybrid deep learning + hand crafted feature architectures to further enhance accuracy. Additional exploration into interpretability tools, such as saliency mapping or feature attribution techniques, may also support clinical applicability

Author's Contribution

All authors contributed to the project through data prep, preprocessing design, feature engineering, modeling, evaluation, and documentation. Collaborative effort ensured methodological success and reproducibility through the project cycle.

Availability Data and Materials

The Brain Tumor MRI dataset used in this study is publicly available on https://www.kaggle.com/datasets /masoudnickparvar/brain-tumor-mri-dataset.

References

[1] R. N. Asif et al., "Brain tumor detection empowered with

ensemble deep learning approaches from MRI scan images," Nature News, https://www.nature.com/articles/s41598-025-99576-7 (accessed Dec. 10, 2025).

[2] N. Shahid, T. Rappon, and W. Berta, "Applications of artificial neural networks in health care organizational decision-making: A scoping review," PloS one, https://pmc.ncbi.nlm.nih.gov/articles/PMC6380578/ (accessed Dec. 10, 2025).

[3] S. P. Singh et al., "3D deep learning on Medical Images: A Review," Sensors (Basel, Switzerland), https://pmc.ncbi.nlm.nih.gov/articles/PMC7570704/ (accessed Dec. 10, 2025