HIGH SCHOOL EDITION
Journal of Student Research
Volume 12 Issue 4 (2023)

Machine Learning Approaches to Detect Brain Tumors from Magnetic Resonance Imaging Scans

Si-Kei Chiu[1] and Shreya Parchure[1#]

[1]International Bilingual School at Tainan-Science-Park
[#]Advisor

ABSTRACT

Artificial intelligence (AI) models have significantly transformed various industries, including healthcare, in recent years. Among the many areas benefiting from AI, brain tumor detection has seen remarkable advancements. Accurate brain tumor detection plays a crucial role in the timely diagnosis and treatment of neurological disorders. AI models have made detecting brain tumors more precise and efficient. Our study utilized a comprehensive dataset of brain magnetic resonance imaging (MRI) scans to compare and assess the performance of different baseline AI models. These models included the K-Nearest Neighbors (KNN) Classifier, Logistic Regression (LR), Decision Tree Classifier, and Multi-Layer Perceptron (MLP). Our analysis revealed that the KNN Classifier yielded the highest accuracy at 88.5%, making it the most suitable AI baseline model for brain tumor detection. These findings underscore the potential of AI models in achieving accurate and efficient brain tumor detection, paving the way for further advancements in this technology.

Introduction

Accurate segmentation and classification of brain tumors using magnetic resonance imaging (MRI) remains a challenge. Previous research has identified several factors that prevent precise segmentation and classification: location, size, and shape variation, magnetic field fluctuations, the infiltrative nature of gliomas, and the intricacy of stroke lesion segmentation (Amin, et al. 2022). These challenges are further exacerbated by the fact that brain tumors are often diagnosed at later stages when they are harder to detect. Therefore, there is a pressing need to enhance diagnostic accuracy and enable timely treatments.

Other research has shown promising results in utilizing machine learning (ML) for brain tumor detection. For example, a study demonstrated significant improvements in detecting tumor boundaries, with a consensus model achieving a 27% improvement in detecting enhancing tumor boundaries, a 33% improvement in detecting tumor core boundaries, and a 16% enhancement in detecting whole tumor boundaries (ScienceDaily, 2023). Additionally, Health Analytics research indicated that incorporating deep learning techniques could improve the classification and diagnosis of brain tumors, leading to a 12% increase in the average accuracy of neurologists (Kennedy, 2022). These findings underscore the potential of ML to address the challenges associated with brain tumor detection and classification.

The practical application of solving the research problem is significant for multiple reasons. Firstly, accurate classification of brain health can assist healthcare professionals in determining whether the tumor exists and its precise location and size, facilitating treatment decision-making. Secondly, it can lead to earlier detection of tumors, improving patient outcomes and survival rates. Therefore, by addressing the limitations and barriers in current approaches, this research has the potential to significantly impact patient care and contribute to advancements in medical imaging analysis.

Methodology

Dataset:

The research utilizes a publicly available Kaggle dataset comprising 253 brain MRI scans, including 98 healthy and 155 tumor scans. The initial preprocessing steps involve converting the color images of brain tumors to grayscale to minimize variations in pixel intensities and normalization to transform the data to a standard scale or range.

Software Used:

This research used Python programming language in the Google Colab environment, employing several primary libraries: NumPy, Matplotlib, Open Source Computer Vision Library (Open CV), and scikit-learn.

```
# function to normalize inputted image (standardized data format)
def normalize_one_image(image):
  output = (image - np.min(image)) / (np.max(image) - np.min(image))
  return output

# function to normalize inputted image (standardized data format)
def normalize_one_image(image):
  output = (image - np.min(image)) / (np.max(image) - np.min(image))
  return output

norm_gray_images = []
for i in image_data:
  gray_i = color.rgb2gray(i)
  norm_gray_i = normalize_one_image (gray_i)
  norm_gray_images.append(norm_gray_i)

flat_images = []
for i in norm_gray_images:
  image=i.flatten()
  flat_images.append(image)
```

Figure 1. Codes used to grayscale and normalize data set

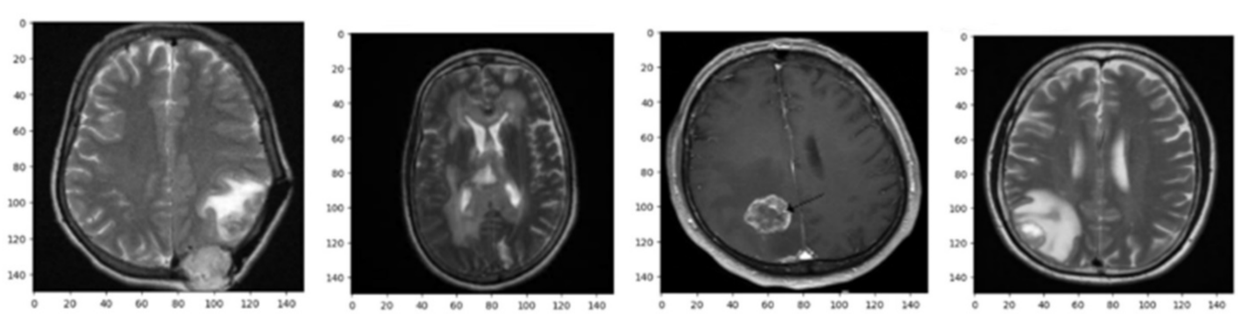


Figure 2. Examples of MRI slices in the dataset showing brain tumors

ML Approaches:

Several ML algorithms will be implemented and evaluated for brain tumor classification. These algorithms include:

KNN Classifier:

classifies samples based on the majority vote of their k nearest neighbors, calculating the distance between a given model and its neighbors to determine its class (Www.Javatpoint.Com., 2023) (top left in images below)

LR Classifier:

models the relationship between the input features and the binary outcome using the sigmoid function by estimating the probabilities of different classes and assigns the sample to the class with the highest probability (José, 2022) (top right in images below)

Decision Tree Classifier:

builds Decision Trees (hierarchical structure of nodes that split the data based on future thresholds to make decisions) by recursively splitting the data and assigning class labels to the leaf nodes (Www.Javatpoint.Com., 2023) (bottom left in images below)

Multi-Layer Perceptron (MLP) Classifier:

type of artificial neural network with multiple layers of nodes, including input, hidden, and output layers that use backpropagation and gradient descent to adjust the weights and biases of the network during training, allowing it to learn complex patterns and make predictions (Mohanty, 2019) (bottom right in images below)

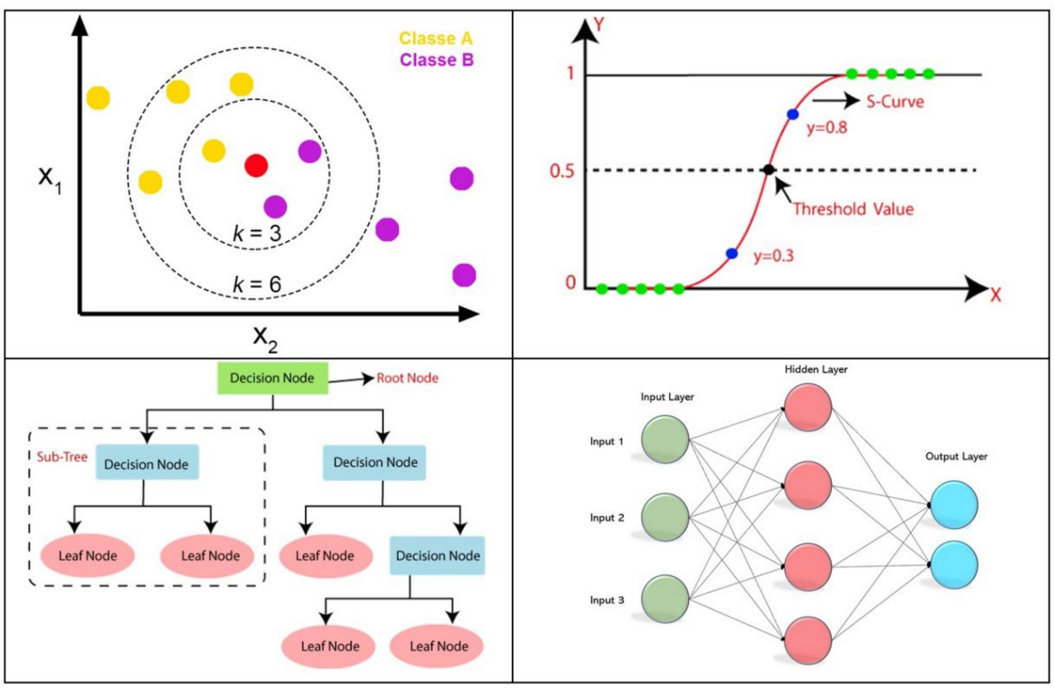


Figure 3. Different Machine Learning approaches

Training and Testing:

The dataset is divided into a training set (80% of the data) and a testing set (20% of the data) to assess the models' performance on unseen samples. After division, we trained the ML models to evaluate their performance on the testing set. The accuracy metric will assess how well the models can classify brain tumors. We employed parameter optimization techniques to optimize the models and improve their performance, such as adjusting hyperparameters (e.g., the number of neighbors for KNN or the learning rate for MLP) and using techniques like grid search or random search. After optimization, we assessed the models to identify the most accurate and efficient algorithm for brain tumor classification.

Results

Upon analyzing the results of our experiments, we found that the KNN Classifier emerged as the most effective AI model for brain tumor classification. By fine-tuning the model with different parameters, we discovered that the KNN Classifier achieved its highest accuracy of 88.5% when utilizing 3, 11, and 15.5 neighbors. Comparing this accuracy to other models, such as LR and MLP, the KNN Classifier consistently outperformed them.

The LR model also exhibited promising results, with an accuracy consistently above 85% after setting an appropriate threshold for the maximum number of iterations, which suggests that the LR model could provide reliable and accurate brain tumor classification by optimizing the parameter to the maximum number of iterations.

On the other hand, the Decision Tree Classifier demonstrated comparatively lower accuracy levels. Its maximum accuracy of 75.5% was achieved when employing optimal maximum depths of 5, 9, and 10. Although the Decision Tree Classifier had lower accuracy than the KNN Classifier and LR, it could still prove helpful in specific scenarios or as part of an ensemble approach for brain tumor classification.

Finally, the MLP model also displayed promising performance with an accuracy of 84.3%. This was achieved by configuring the model with a suitable number of neurons, precisely 10. Despite not having the highest accuracy, it remains a viable option for brain tumor classification tasks.

Discussion

The findings of our study underscore the pivotal role that ML approaches, particularly the KNN Classifier, can play in enhancing brain tumor detection accuracy using MRI scans. The achievement of an 88.5% accuracy by the KNN Classifier is remarkable, suggesting its effectiveness in distinguishing between healthy brain tissues and tumor-afflicted regions. However, while the KNN Classifier demonstrated superiority, it is crucial to consider the broader context and implications of our results.

Notably, the LR, Decision Tree, and MLP models also showcased competitive accuracy rates of 85%, 75.5%, and 84.3%, respectively, after appropriate parameter tuning. These accuracies highlight the potential of these models as viable alternatives for brain tumor classification. The LR model's consistent performance above 85% accuracy suggests its robustness and reliability, mainly when sufficient iterations allow convergence.

The Decision Tree Classifier's slightly lower accuracy than other models should maintain its value. While it might have yet to achieve the highest accuracy, its simplicity and interpretability can make it valuable (Kingsford, et al. 2008) when understanding the decision-making process is crucial. Moreover, the Decision Tree Classifier could contribute to a more comprehensive and accurate classification strategy as part of an ensemble approach.

The promising performance of the MLP model aligns with the increasing interest in deep learning techniques for medical image analysis (Rabbi, et al. 2022). Despite not securing the highest accuracy, the MLP model's ability to capture complex patterns within the data suggests its potential to excel in more intricate brain tumor classifications. This opens doors for future investigations into the effectiveness of more complex neural network architectures.

Nonetheless, as with any research, some limitations warrant discussion. One notable limitation is the reliance on a specific dataset from Kaggle. While this dataset is publicly available and validated, its scope might only partially encompass the diversity of brain tumor cases encountered in the real world. The generalizability of the models to a broader population could be enhanced by incorporating more extensive and diverse datasets, ensuring that the models can cater to a wider range of patients and conditions.

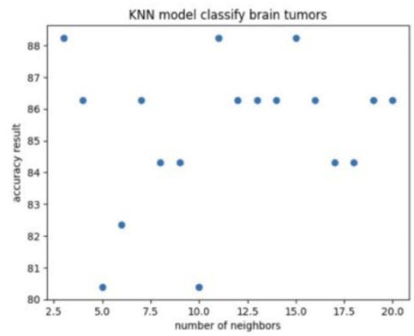
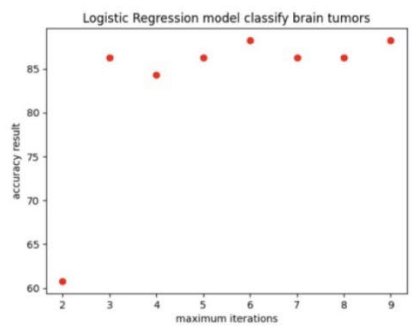
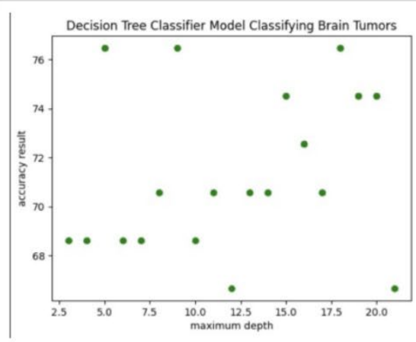
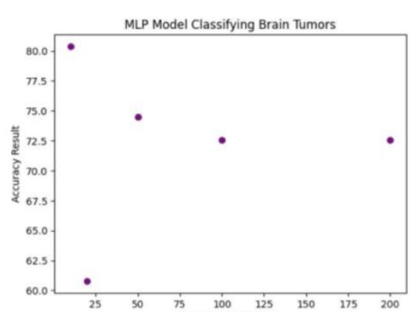
Findings		***Diagram***
K-Nearest Neighbors (KNN)	Parameter: k = 3, 11, and 15.5 Accuracy: 88.5%	
Logistic Regression (LR)	Parameter: max iter > 3 Accuracy: 85%	
Decision Tree	Parameter: max depth = 5, 9, and 10 Accuracy: 75.5%	
Multi-Layer Perceptron (MLP)	Parameter: neuron # = 10 Accuracy: 84.3%	

Figure 4. Summary of Results

Conclusion

In conclusion, our study contributes to the growing body of knowledge demonstrating the potential of ML models in brain tumor detection from MRI scans. The KNN Classifier's remarkable accuracy showcases its immediate practicality in assisting healthcare professionals in diagnosing brain tumors accurately. Nevertheless, the competitive performances of the other models, along with their unique strengths, should be considered. Our study not only reinforces the significance of AI in medical imaging but also points toward future research directions, such as exploring model performance across various brain tumor types, embracing more diverse datasets, and harnessing ensemble methods for further accuracy improvements. Ultimately, these efforts can result in more reliable diagnostic tools that positively impact patient care and contribute to the advancement of medical imaging analysis.

References

AI enables large-scale brain tumor study, without sharing patient data. (n.d.). ScienceDaily. Retrieved July 17, 2023, from https://www.sciencedaily.com/releases/2022/12/221205104205.htm

Amin, J., Sharif, M., Haldorai, A., Yasmin, M., & Nayak, R. S. (2022). Brain tumor detection and classification using machine learning: A comprehensive survey. Complex & Intelligent Systems, 8(4), 3161–3183. https://doi.org/10.1007/s40747-021-00563-y

Decision tree algorithm in machine learning—Javatpoint. (n.d.). Www.Javatpoint.Com. Retrieved July 17, 2023, from https://www.javatpoint.com/machine-learning-decision-tree-classification-algorithm

José, I. (2021). KNN (K-nearest neighbors) #1. Medium. https://towardsdatascience.com/knn-k-nearest-neighbors-1-a4707b24bd1d

Kennedy, S. (2022). Deep learning may improve identification of brain tumors. HealthITAnalytics. https://healthitanalytics.com/news/deep-learning-may-improve-identification-of-brain-tumors

Kingsford, C., & Salzberg, S. L. (2008). What are decision trees? Nat Biotechnol, 26(9), 1011–1013. https://doi.org/10.1038/nbt0908-1011

Logistic regression in machine learning—Javatpoint. (n.d.). Www.Javatpoint.Com. Retrieved July 17, 2023, from https://www.javatpoint.com/logistic-regression-in-machine-learning

Mohanty, A. (2019). Multi layer perceptron (Mlp) models on real world banking data. Medium. https://becominghuman.ai/multi-layer-perceptron-mlp-models-on-real-world-banking-data-f6dd3d7e998f

Rabbi, F., Dabbagh, S. R., Angin, P., Yetisen, A. K., & Tasoglu, S. (2022). Deep Learning-Enabled Technologies for Bioimage Analysis. Micromachines (Basel), 13(2), 260. https://doi.org/10.3390/mi13020260