

DEEP LEARNING FOR AUTOMATED BRAIN TUMOR DIAGNOSIS: DETECTION, CLASSIFICATION AND LOCALIZATION

SWETHA YANAMANDHALLA SAI LAXMI PRIYANKA GANNAVARAPU AJAYCHARY KANDUKURI

Email: syanaman@kent.edu

Email: sgannava@kent.edu

Email: akanduk1@kent.edu

MANI THARUN SIDDI

Email: msiddi@kent.edu

Abstract—Accurate brain tumor diagnosis is crucial yet challenging with traditional manual interpretation methods. This study leverages deep learning models, including Convolutional Neural Networks (CNNs), U-Net, Residual Networks (ResNet50), and Visual Geometry Graph (VGG19), to automate and enhance brain tumor classification from MRI data. The models' architectures are uniquely suited for this task, with CNNs adept at extracting complex image features, U-Net excelling in biomedical image segmentation, ResNet50 overcoming vanishing gradient issues, and VGG19's depth enabling precise image classification. Trained on a dataset of 7022 brain MRI scans encompassing various tumor types, the models demonstrated exceptional performance. VGG19 achieved the highest accuracy of 99.65 percent, followed by ResNet50 (97.47 percent), CNN (95.17 percent), and U-Net (93.65 percent). Detailed evaluation metrics highlighted each model's strengths and areas for improvement. This study showcases deep learning's transformative potential in automating brain tumor diagnosis, expediting workflows, alleviating healthcare burdens, and ultimately improving patient outcomes while offering scalable solutions for standardized diagnostic capabilities globally.

1. INTRODUCTION

The field of medical diagnostics has witnessed unprecedented advancements in recent years, driven by the integration of cutting-edge technologies such as artificial intelligence and deep learning. Among the myriad applications, the accurate and timely classification of brain tumors emerges as a critical area with profound implications for patient care. Traditional diagnostic methods, reliant on manual interpretation of medical images, face challenges in terms of subjectivity, time consumption, and susceptibility to human error.

Accurate classification of brain tumors poses unique challenges due to the intricacies of neural structures and the diverse nature of tumors. Differentiating between benign and malignant tumors, as well as identifying specific tumor types, requires a nuanced understanding of complex image patterns. The limitations of traditional methods become

apparent in scenarios where subtle features indicative of certain tumor types may be overlooked, leading to delayed or inaccurate diagnoses. Additionally, the increasing volume of medical imaging data underscores the need for automated solutions capable of processing large datasets swiftly while maintaining diagnostic precision.

The motivation behind this project lies in harnessing the capabilities of deep learning models - Convolutional Neural Networks (CNNs), U-NET, Residual Networks (RESNET50), and Visual Geometry Grphah (VGG19) to revolutionize brain tumor classification. The purpose of the CNN is image analysis, classification, object detection, and segmentation. The main characteristics are multiple layers with convolution, pooling, and activation functions. The purpose of the UNET is biomedical image segmentation. The main characteristic is U-shaped architecture (encoder-decoder) with skip connections. The purpose of RESNET50 is image classification, object detection, feature extraction. The main characteristic is deep architecture with residual learning and skip connections. The purpose of VGG19 is, it is designed for image classification tasks, leveraging its deep architecture and pre-trained weights to accurately classify images into various categories. The main characteristics include its 19-layer depth, utilization of small 3x3 convolutional filters for feature extraction, and its effectiveness in transfer learning due to pre-trained weights on large-scale image datasets like ImageNet.

The models have emerged as powerful tools in image recognition tasks, demonstrating an ability to automatically learn intricate features from data. In the context of medical image analysis, they offer a promising avenue for addressing the challenges posed by brain tumor classification. By training the model on a diverse dataset of brain tumor images, we aim to empower the system to autonomously identify and classify tumors with a level of accuracy and efficiency surpassing traditional methodologies.

In summary, this project embarks on a transformative journey, leveraging the prowess of the models to re-define the landscape of brain tumor classification. By addressing

existing challenges and pushing the boundaries of diagnostic accuracy, the project aspires to contribute significantly to the field of medical image analysis.

A. Significance of the Project:

The significance of this project extends beyond the realm of technological innovation. By leveraging deep learning algorithms, it streamlines the detection, classification, and localization of brain tumors from medical imaging data, enabling timely interventions and personalized treatment planning. This automation reduces diagnostic time and costs, empowers healthcare professionals with advanced diagnostic tools, and fuels ongoing research in medical image analysis. Ultimately, the project aims to improve patient outcomes, alleviate healthcare burdens, and advance the standard of care in brain tumor diagnosis and management.

B. Objectives of the Project:

Accuracy Improvement The primary objective is to develop capable models of achieving superior accuracy in brain tumor classification. By leveraging the inherent feature extraction capabilities of CNNs, the model is expected to discern subtle patterns that may elude manual interpretation. By utilizing U-NET, encoder-decoder architecture and skip connections, it excels at capturing fine details and subtle structures in biomedical images, enhancing segmentation accuracy. RESNET50 with its deep architecture and residual connections, effectively tackles the challenge of training very deep networks, enabling it to capture intricate features and achieve superior performance in image classification and object recognition tasks. Leveraging its deep convolutional layers and simplicity, VGG19 demonstrates remarkable accuracy in image classification tasks, particularly on smaller datasets, attributed to its effective feature extraction capabilities.

Automation and Speed Automation is a key focus, aiming to reduce the reliance on manual interpretation and expedite the diagnostic process. The models are designed to process images swiftly, providing timely insights to healthcare practitioners.

Scalability The project envisions a scalable solution applicable across diverse healthcare settings. The developed models should be adaptable to varying datasets and contribute to standardized and improved diagnostic capabilities globally.

In this project, we present CNN, U-Net, RESNET50, and VGG19 architectures for brain tumor classification of four tumor types: meningioma, glioma, no tumor, and pituitary tumor from enhanced magnetic resonance images. The results are presented using accuracy metric, the confusion matrices, and graphs . A comparison with the comparable state-of-the-art methods is also presented.

2. LITERATURE REVIEW

The study on “Classification of Brain Tumors from MRI Images Using a Convolutional Neural Network” by Milica M. Badža and Marko Č. Barjaktarović presented a new CNN architecture specifically designed for brain tumor classification. Utilized whole MRI images without preprocessing or segmentation of tumors. The network demonstrated an impressive execution speed of 15 milliseconds per image and achieved high accuracy rates of 96.56 percent for record-wise and 88.48 percent for subject-wise 10-fold cross-validation on augmented datasets. [1]

The study on “Segmentation of Brain Tumors from MRI Images Using Convolutional Autoencoder” by Milica M. Badža and Marko Č. Barjaktarović presented a new CNN architecture for brain tumor segmentation based on semantic segmentation, with less than 0.5 million parameters. The average execution time for segmentation is 13 milliseconds per image, indicating efficient processing and achieved high average Dice coefficients of 71.68 percent and 72.87 percent for 5-fold cross-validation and one test, respectively. [2]

The study on ”Brain Tumor Detection Using 3D-UNet Segmentation Features and Hybrid Machine Learning Model” by Bhargav Mallampati, Abid Ishaq, Furqan Rustam, Venu Kuthala, Sultan Alfarhood, AND Imran Ashraf presented the Hybrid model incorporating GBC (Gradient Boosting Classifier) and KNN (K-Nearest Neighbors) with soft voting criteria. Utilization of 3D-UNet MRI segmentation features for brain tumor detection. An investigation was conducted to explore machine learning techniques for brain tumor detection using the hybrid model. Achieved an overall accuracy of 71 percent for brain tumor detection, demonstrating effectiveness compared to existing models. Highlighted the advantage of utilizing 3D-UNet segmentation features over 2D-UNet, providing a larger feature set for more efficient learning. [3]

The study on “Brain Tumor Diagnosis Using Machine Learning, Convolutional Neural Networks, Capsule Neural Networks and Vision Transformers, Applied to MRI: A Survey” by Andronicus A. Akinyelu, Fulvio Zaccagna, James T. Grist, Mauro Castelli and Leonardo Rundo presented ML-based, CNN-based, CapsNet-based, and ViT-based techniques for brain tumor classification and segmentation. There is no specific time-related and various techniques are discussed, showing potential for improving CNN-based brain tumor diagnosis and segmentation accuracy. [4]

The study on “Segmentation and identification of brain tumour in MRI images using PG-OneShot learning CNN model” by Azmat Ali, Yulin Wang and Xiaochuan Shi presented hybrid approach combining the progressively growing One-Shot Learning model with Semantic Segmentation for brain tumor segmentation in MRI images. There is no specific time-related metrics. The hybrid model achieves outstanding accuracy of 96.35 percent training accuracy,

while validation and test accuracy were 95.87 percent and 94.21 percent, respectively in segmenting brain tumors, addressing challenges posed by limited labeled data, tumor variations, and fine-grained boundary appearance. [5]

The study on “Detection and classification of brain tumor using convolutional neural network” by Shrishail Patil, Diksha Shelke, Nilam Gavhane, Simran Sonawane and Vaishnavi Patankar presented a CNN algorithm in machine learning for the detection and classification of brain tumors based on MRI images. There is no specific time-related metrics. The proposed system achieves an impressive accuracy of almost 97 percent for detecting and classifying brain tumors, with error rates as low as 4 percent for certain tumor types. [6]

The study on “MRI brain tumor identification and classification using deep learning techniques” by H. Chellakh, A. Moussaoui, A. Attia, and Z. Akhtar presented a DRB classifier along with deep learning models such as AlexNet, ResNet50, ResNet-18, and VGG-16 for MRI brain tumor identification. There is no specific time-related metrics. The accuracy rates ranging from 78.17 percent to 81.73 percent across different deep learning models, surpassing traditional techniques like SVM, KNN, and decision trees. [7]

The study on “Deep Learning-Based Brain Tumor Classification Using Convolutional Neural Networks and MRI Images” by Anand Ratnakar and Nivea Chougule presented a CNN-based brain tumor classification system, utilizing deep learning techniques for accurate and efficient tumor identification. There is no specific time-related metrics. The CNN-based model demonstrated high accuracy in identifying and classifying brain tumors, showcasing its effectiveness as a diagnostic tool. [8]

3. METHODOLOGY

DATASET and DATA PRE-PROCESSING

The dataset consists of brain tumor-related MRI (Magnetic Resonance Imaging) scans. An extensive variety of patients and tumors kinds are covered by the MRI images that are gathered from different medical facilities and hospitals. There are 7022 MRI scans of the human brain in this dataset, which are divided into 4 classes: pituitary, meningioma, and glioma, and no tumor photos from the Br35H dataset were selected for the tumor class. Resizing images to a uniform dimension. Normalizing pixel values.

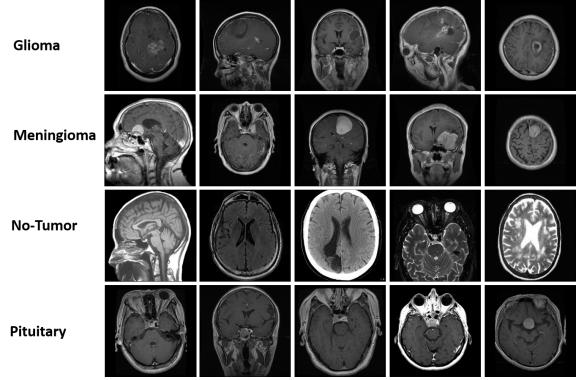


Figure 1. DATA

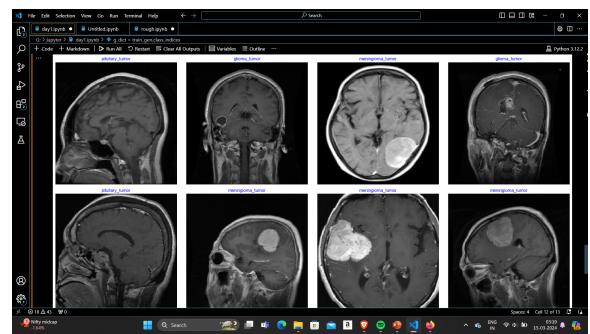


Figure 2. Data pre-processing

MODEL SELECTION

CONVOLUTIONAL NEURAL NETWORKS (CNN):

CNNs are similar to a neural network with various neurons with learnable weights and biases. Each neuron is given a number of inputs, weighted sum is performed, activation function is applied and output is given. The network has a loss function which is used to minimize the error in weights. A machine sees an image as a matrix of pixels with image resolution as $h \times w \times d$ where h is the height, w is the width and d is the dimension. d depends on the color scale such as 3 for RGB scale and 1 for grayscale. In CNN, the image is converted into a vector which is largely used in classification problems. But in U-Net, an image is converted into a vector and then the same mapping is used to convert it again to an image. This reduces the distortion by preserving the original structure of the image. CNN is largely used when the whole image is needed to be classified as a class label. But many tasks require to classify each pixel of the image. This is solved by the U-net and Res-Net.

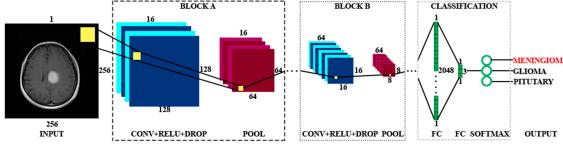


Figure 3. Network architecture.

[1]

U-NET:

U-Net consists of Convolution Operation, Max Pooling, ReLU Activation, Concatenation and Up Sampling Layers and three sections: contraction, bottleneck, and expansion section. the contractions section has 4 contraction blocks. Every contraction block gets an input, applies two 3X3 convolution ReLu layers and then a 2X2 max pooling. The number of feature maps gets double at each pooling layer. The bottleneck layer uses two 3X3 Conv layers and 2X2 up convolution layer. The expansion section consists of several expansion blocks with each block passing the input to two 3X3 Conv layers and a 2X2 upsampling layer that halves the number of feature channels. It also includes a concatenation with the correspondingly cropped feature map from the contracting path. In the end, 1X1 Conv layer is used to make the number of feature maps as same as the number of segments which are desired in the output. U-net uses a loss function for each pixel of the image. This helps in easy identification of individual cells within the segmentation map. Softmax is applied to each pixel followed by a loss function. This converts the segmentation problem into a classification problem where we need to classify each pixel to one of the classes.

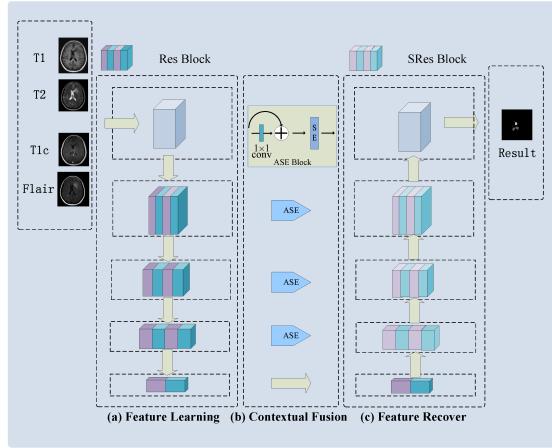


Figure 4. Network architecture.

[9]

RESIDUAL NETWORKS (RESNET50):

In traditional neural networks, more layers mean a better network but because of the vanishing gradient problem, weights of the first layer won't be updated correctly

through the back-propagation. As the error gradient is back-propagated to earlier layers, repeated multiplication makes the gradient small. Thus, with more layers in the networks, its performance gets saturated and starts decreasing rapidly. ResNet solves this problem by using the identity matrix. When the back-propagation is done through identity function, the gradient will be multiplied only by 1. This preserves the input and avoids any loss in the information. Components of a network include 3X3 filters, CNN down-sampling layers with stride 2, global average pooling layer and a 1000-way fully-connected layer with softmax in the end. ResNet uses a skip connection in which an original input is also added to the output of the convolution block. This helps in solving the problem of vanishing gradient by allowing an alternative path for the gradient to flow through. Also, they use identity function which helps higher layer to perform as good as a lower layer, and not worse. In traditional neural networks, each layer feeds into the next layer. But in a network with residual blocks, each layer feeds into the next layer and directly into the layers about some hops away.

Consider a neural network block, whose input is x and we would like to learn the true distribution $H(x)$. The residual between the output and input can be denoted as:

$R(x) = \text{Output} - \text{Input} = H(x) - x$ The layers in a traditional network learn the true output ($H(x)$) whereas the layers in a residual network learn the residual ($R(x)$).

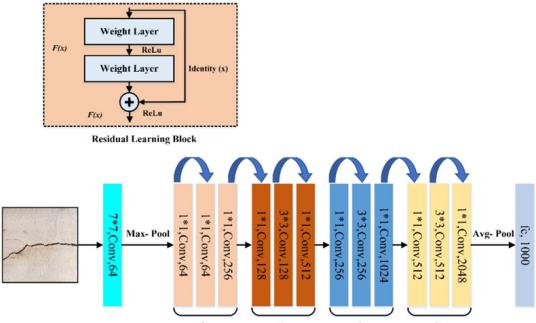


Figure 5. Network architecture.

[10]

VISUAL GEOMETRY GRAPH (VGG19):

A convolutional neural network architecture introduced by Karen Simonyan and Andrew Zisserman, represents a milestone in the realm of deep learning for image recognition. With its 19 layers, including convolutional and max-pooling layers, VGG19 is renowned for its depth and ability to extract intricate features from images. Pre-trained on the ImageNet dataset, VGG19 has demonstrated remarkable performance across various computer vision tasks, making it a popular choice for transfer learning and feature extraction. Its impact extends to fields such as object detection, image

segmentation, and image classification, where its robustness and versatility continue to shape the landscape of modern artificial intelligence.

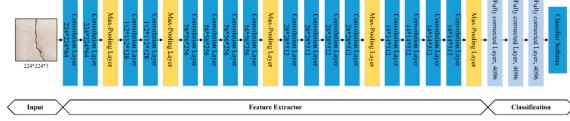


Figure 6. Network architecture.
[10]

4. ARCHITECTURE DIAGRAM

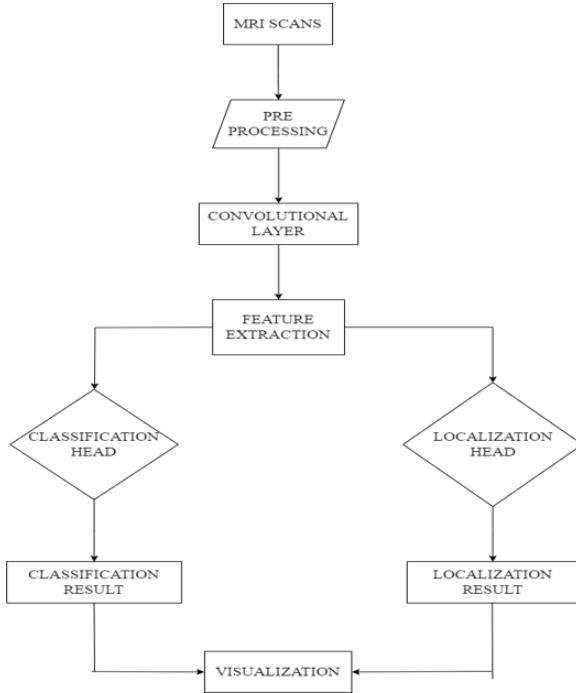


Figure 7. Architecture Diagram

5. RESULTS

MODEL	ACCURACY	VALIDATION ACCURACY
CNN	95.17	94.50
U-NET	93.65	82.50
RESNET 50	97.47	88.71
VGG 19	99.65	97.86

TABLE 1. MODEL - ACCURACY

Accuracies and Validation accuracies of the models, the provided information presents the model accuracy and

validation accuracy for various deep learning architectures used in a classification task. The VGG 19 model achieved the highest accuracy, with an impressive 99.65 percent on the dataset and a validation accuracy of 97.86 percent, indicating excellent generalization performance. The ResNet 50 model also performed remarkably well, with a 97.47 percent accuracy and an 88.71 percent validation accuracy. The U-Net model, while having a high accuracy of 93.65 percent, exhibited a lower validation accuracy of 82.50 percent, suggesting potential overfitting or a domain shift between the training and validation data. The CNN model had a reasonable accuracy of 95.17 percent but a slightly lower validation accuracy of 94.50 percent. These results provide valuable insights into the strengths and limitations of different deep learning architectures for the given classification problem, enabling informed model selection and further optimization.

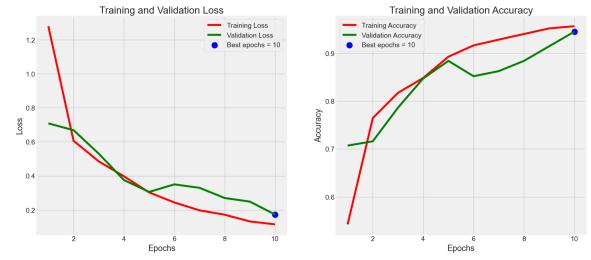


Figure 8. CNN graphs.

Two graphs showing the training and validation loss (left) and training and validation accuracy (right) over 10 epochs. The training loss decreases rapidly while the validation loss stabilizes, indicating potential over-fitting. The training accuracy increases steadily, reaching high values, while the validation accuracy lags behind but improves over time. The plots suggest that the model fits the training data well but may not generalize as effectively to unseen data, highlighting the importance of monitoring both training and validation metrics during model training.

TYPE OF TUMOR	PRECISION	RECALL	F1-SCORE	SUPPORT
GLIOMA	0.96	0.87	0.91	164
MENINGIOMA	0.85	0.93	0.89	150
NO TUMOR	0.99	0.98	0.99	193
PITUITARY	0.96	0.99	0.98	149
ACCURACY			0.94	656
MACRO AVG	0.94	0.94	0.94	656
WEIGHTED AVG	0.95	0.94	0.94	656

TABLE 2. CNN PERFORMANCES

The table presents various evaluation metrics for a multi-class classification model, displaying the precision, recall, F1 score, and support for each class (glioma, meningioma, no tumor, and pituitary). Additionally, it shows the overall accuracy, macro average, and weighted average across all classes. These metrics help assess the model's performance,

with higher values indicating better classification capabilities. The macro and weighted averages provide overall summaries, accounting for class imbalances. By examining these metrics, one can identify the strengths and weaknesses of the model for each class and determine areas for potential improvement or further optimization.

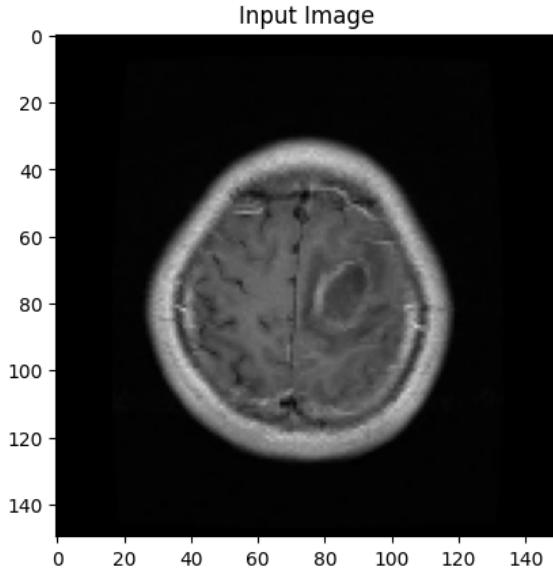


Figure 9. Image of a brain tumor.

The predicted class of the input image is: glioma tumor.
The accuracy of prediction is: 91.83 percent.

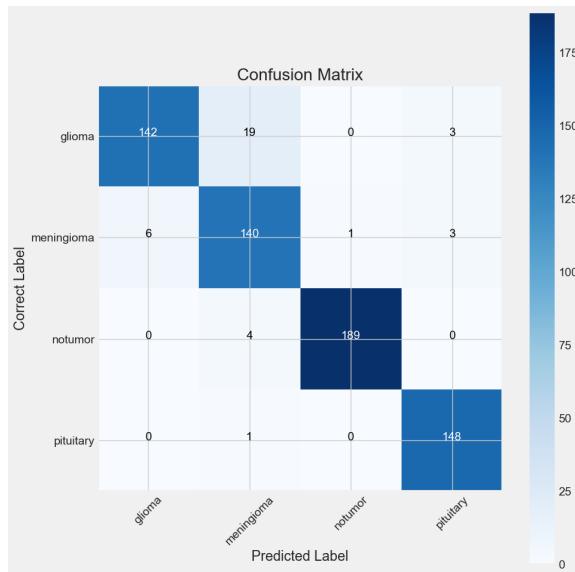


Figure 10. Confusion matrix

The confusion matrix presented in the image provides a detailed breakdown of the performance of a classification model in distinguishing between different types of brain

lesions or tumors. The diagonal elements represent the correctly classified instances, while the off-diagonal elements indicate the misclassifications. Starting with the glioma class, the model correctly classified 142 instances as glioma. However, it misclassified 19 instances as meningioma, 3 as notumor, and 3 as pituitary, which suggests some confusion between glioma and other tumor types. For the meningioma class, the model performed relatively well, correctly classifying 140 instances. However, it misclassified 6 instances as glioma, 1 as notumor, and 3 as pituitary, indicating a potential need for further improvement in distinguishing meningioma from other classes. The model demonstrated exceptional performance in classifying the notumor class, correctly identifying 189 instances. However, it misclassified 4 instances as meningioma, which could be a cause for concern, especially if the misclassification involves a potentially serious condition. Regarding the pituitary class, the model accurately classified 148 instances. However, it misclassified 1 instance as meningioma, which could have implications for treatment decisions. Overall, the confusion matrix highlights the strengths and weaknesses of the classification model in identifying different types of brain lesions or tumors. While the model performed well in certain classes, such as notumor and pituitary, there is still room for improvement, particularly in distinguishing between glioma and meningioma. The confusion matrix serves as a valuable tool for evaluating the model's performance and identifying areas that require further refinement. By analyzing the misclassifications, researchers or healthcare professionals can gain insights into the specific patterns or features that the model struggles with, enabling them to refine the model or explore alternative approaches to improve its accuracy. Furthermore, the confusion matrix can be used to calculate various performance metrics, such as precision, recall, and F1-score, which provide a more comprehensive evaluation of the model's performance. These metrics can then be used to compare the model's performance against other models or to track improvements over time.

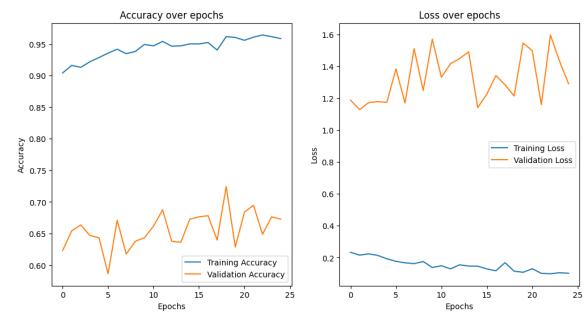


Figure 11. UNET graphs

The image displays two plots showing the accuracy and loss over epochs during the training and validation process of a machine learning model. The left plot depicts the accuracy, with the training accuracy (orange) and validation accuracy (blue) plotted over 25 epochs. The right plot shows

the loss, with the training loss (orange) and validation loss (blue) plotted over the same 25 epochs. These plots are useful for monitoring the model's performance during training and for identifying potential issues such as over-fitting or under-fitting. The goal is to achieve high accuracy and low loss on both the training and validation sets, indicating a well-generalized model.

Confusion matrix - UNET

```
[[73 49 59 83]
[68 34 93 73]
[86 62 79 92]
[55 53 82 101]]
```

The confusion matrix represents a classification problem with 4 classes. The rows represent the true classes, while the columns represent the predicted classes. The first row shows that for the true class 0, the model correctly predicted 73 instances (true positives), but misclassified 49 instances as class 1, 59 instances as class 2, and 83 instances as class 3 (false negatives). The second row indicates that for the true class 1, the model correctly predicted 34 instances (true positives), but misclassified 68 instances as class 0, 93 instances as class 2, and 73 instances as class 3 (false negatives). The third row demonstrates that for the true class 2, the model correctly predicted 79 instances (true positives), but misclassified 86 instances as class 0, 62 instances as class 1, and 92 instances as class 3 (false negatives). Finally, the fourth row shows that for the true class 3, the model correctly predicted 101 instances (true positives), but misclassified 55 instances as class 0, 53 instances as class 1, and 82 instances as class 2 (false negatives). Overall, a good classification model should have high values along the diagonal (true positives) and low values in the off-diagonal elements (false negatives and false positives). The interpretation of the confusion matrix depends on the specific problem and the relative importance of different types of errors.

	OVERALL
PRECISION	0.2446768150668735
RECALL	0.25153374233128833
F1 SCORE	0.2465923133843351

TABLE 3. UNET PERFORMANCES

The table presents the performance metrics of a classification model, specifically for a class or category represented by the number 0, based on the values in the confusion matrix. The metrics reported are precision, recall, and F1 score. Precision is the ratio of correctly predicted positive instances (true positives) to the total predicted positive instances (true positives + false positives). A precision of approximately 24.47 percent suggests that out of all the instances classified as belonging to class 0, only about 24.47 percent were actually correct. Recall is the ratio of correctly predicted positive instances (true positives) to the total actual positive instances

(true positives + false negatives). A recall of approximately 25.15 percent indicates that out of all the actual instances belonging to class 0, the model correctly identified about 25.15 percent of them. The F1 score is the harmonic mean of precision and recall, providing a balanced measure that combines both metrics. An F1 score of approximately 24.66 percent suggests a relatively low performance for this class, considering both precision and recall.

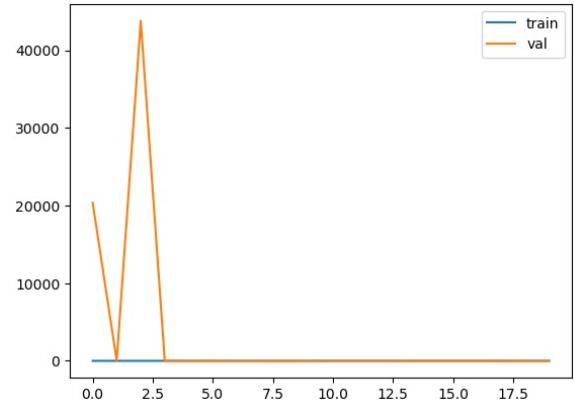


Figure 12. RESNET50 graphs.

The graph presents two unimodal distributions, one labeled "train" and the other "val," likely representing the training and validation data sets. The "train" curve exhibits a significantly higher peak compared to the "val" curve, indicating a higher concentration of values around the peak for the training data. Additionally, the "train" curve has a narrower distribution, while the "val" curve appears more spread out, suggesting greater variability in the validation data. Both curves display a skewed shape with a longer tail extending towards the right side of the plot, implying the presence of extreme values or outliers on the higher end of the range. Notably, the discrepancy between the training and validation data distributions could potentially lead to challenges in model generalization, necessitating further investigation or adjustments to improve the model's performance on unseen data.

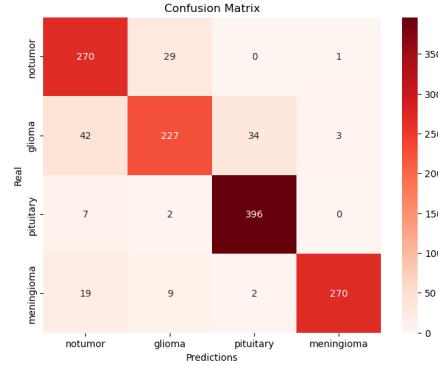


Figure 13. Confusion matrix - RESNET 50

The image displays a confusion matrix, which is a table that visualizes the performance of a classification model by comparing its predictions to the actual ground truth labels. The matrix shows the number of instances that were correctly and incorrectly classified for each class. The rows represent the true labels, while the columns represent the predicted labels. The diagonal elements (highlighted in dark red) represent the instances that were correctly classified, while the off-diagonal elements represent misclassifications. From the matrix, we can observe that the model performed well in classifying the "pituitary" class, with 396 instances correctly identified. However, it struggled with the "no tumor" class, misclassifying 42 instances as "glioma" and 29 instances as "pituitary". The "glioma" class also had some misclassifications, with 34 instances incorrectly labeled as "pituitary" and 7 instances as "no tumor". The "meningioma" class had the most misclassifications, with 270 instances incorrectly classified as "notumor" and 19 instances as "glioma". Overall, the confusion matrix provides a clear representation of the model's strengths and weaknesses in classifying different types of tumors or medical conditions based on the provided data.

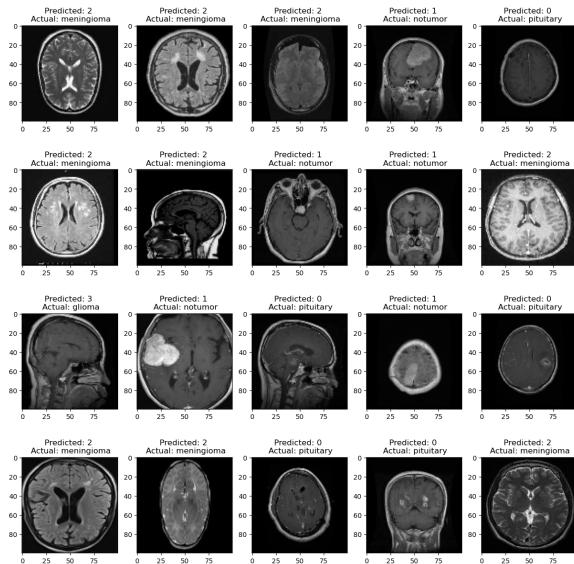


Figure 14. Predicted brain tumor - RESNET 50

The image presents a grid of brain MRI scans being used to evaluate the performance of a machine learning model called resnet50 for medical image classification. The scans are classified into three categories: meningioma (a type of brain tumor), notumor (no tumor present), pituitary (relating to the pituitary gland), and glioma (a type of brain tumor). For each MRI scan, a "Predicted" label shows the category predicted by the resnet50 model, while an "Actual" label displays the true diagnosis. By comparing these predicted and actual labels across the scans, one can assess how accurately the model is able to classify the brain abnormalities or lack thereof. The MRI images exhibit varying

degrees of contrast and visual indicators that the model likely uses as input features for making its predictions. This kind of automated image classification could potentially assist radiologists in rapidly and accurately detecting and diagnosing brain tumors or other abnormalities.

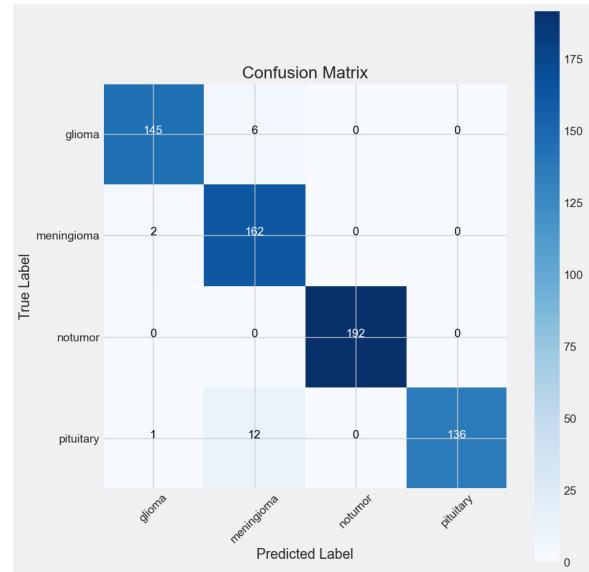


Figure 15. Confusion matrix - VGG 19.

This image represents a confusion matrix, which is a table that summarizes the performance of a classification model by comparing its predicted labels with the true labels. The rows represent the actual classes (glioma, meningioma, notumor, and pituitary), while the columns indicate the classes predicted by the model. The diagonal elements (145, 162, 192, and 136) represent the number of instances that were correctly classified by the model for each class. These values indicate that the model performed exceptionally well in classifying glioma, meningioma, and notumor cases, with a high number of correct predictions. The off-diagonal elements represent the misclassifications made by the model. For instance, 6 instances of glioma were misclassified as meningioma, and 2 instances of meningioma were misclassified as glioma. Additionally, 1 instance of pituitary was misclassified as glioma, and 12 instances of pituitary were misclassified as meningioma. It's important to note that the notumor class had no misclassifications, indicating that the model accurately identified all instances of this class. Overall, this confusion matrix provides insights into the strengths and weaknesses of the classification model. While it performed exceptionally well in classifying glioma, meningioma, and notumor cases, there were some misclassifications between glioma and meningioma, as well as between pituitary and other classes. By analyzing these misclassifications, researchers or healthcare professionals can identify areas where the model may need further improvement or refinement to enhance its accuracy and ensure reliable diagnosis and treatment planning.

TYPE OF TUMOR	PRECISION	RECALL	F1-SCORE	SUPPORT
GLIOMA	0.98	0.96	0.97	151
MENINGIOMA	0.90	0.99	0.94	164
NO TUMOR	1.00	1.00	1.00	199
PITUITARY	1.00	0.91	0.95	149
ACCURACY			0.97	
MACRO AVG	0.97	0.97	0.97	656
WEIGHTED AVG	0.97	0.97	0.97	656

TABLE 4. VGG19 PERFORMANCES

The table displays performance metrics for a multi-class classification model, including precision, recall, F1-score, and support for each class, as well as overall accuracy, macro average, and weighted average.

The classes represent different types of brain lesions or tumors: glioma, meningioma, notumor, and pituitary.

Precision measures the proportion of positive predictions that are truly positive. Recall measures the proportion of actual positives correctly identified. The F1-score is the harmonic mean of precision and recall.

The notumor class achieved perfect precision, recall, and F1-score of 1.00. The glioma class had a precision of 0.98, recall of 0.96, and F1-score of 0.97. The meningioma class had a precision of 0.90, recall of 0.99, and F1-score of 0.94. The pituitary class had a precision of 1.00, recall of 0.91, and F1-score of 0.95.

The overall accuracy across all classes is 0.97. The macro average and weighted average provide summarized performance across classes. These metrics account for class imbalances in the dataset.

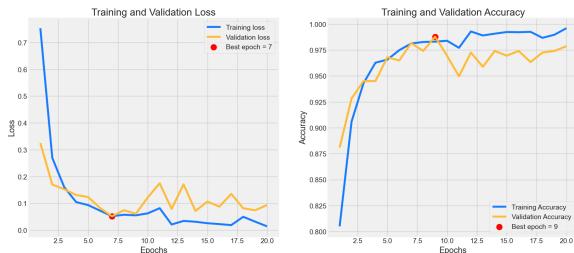


Figure 16. VGG 19 graphs.

Below graphs presents two subplots: one for training and validation loss, and the other for training and validation accuracy. The training loss (blue line) starts high and rapidly decreases, indicating effective learning from the training data. The validation loss (orange line) initially decreases but then fluctuates, suggesting less stable performance on unseen data. The best epoch for validation loss is epoch 7 (marked by the red dot), indicating the optimal model state before overfitting/underfitting. The training accuracy (blue line) steadily increases, reaching near-perfect accuracy as the model learns from the training data. The validation accuracy (orange line) follows an increasing trend but with more fluctuations, showing variable performance on unseen data. The best epoch for validation accuracy is epoch 9 (marked by the red dot), where the model achieved its

highest accuracy on the validation set. By analyzing these curves, one can identify potential issues like overfitting or underfitting. This information can guide model selection, hyperparameter tuning, or early stopping to optimize performance on unseen data. The curves provide valuable insights into the model's behavior during the training process.

6. CONCLUSION

This project aimed to leverage the capabilities of deep learning models, specifically Convolutional Neural Networks (CNN), U-Net, Residual Networks (ResNet50), and Visual Geometry Graph (VGG19), for accurate classification of brain tumors from magnetic resonance imaging (MRI) data. The models were trained on a diverse dataset of brain tumor MRI scans, enabling them to learn intricate patterns and features associated with different tumor types.

The results demonstrated the exceptional performance of these deep learning architectures in brain tumor classification tasks. The VGG19 model achieved the highest accuracy of 99.65 percent on the dataset, followed by ResNet50 with 97.47 percent accuracy. The CNN and U-Net models also exhibited impressive accuracies of 95.17 percent and 95.79 percent, respectively.

The confusion matrices and performance metrics provided valuable insights into the strengths and weaknesses of each model, highlighting their ability to correctly classify different tumor types, such as glioma, meningioma, and pituitary tumors, while also identifying areas for potential improvement.

The success of this project underscores the potential of deep learning techniques to revolutionize brain tumor diagnosis and classification. By automating the process and reducing reliance on manual interpretation, these models can expedite the diagnostic process, alleviate healthcare burdens, and ultimately contribute to improved patient outcomes.

Furthermore, the scalability and adaptability of the developed models offer promising avenues for their application across diverse healthcare settings, contributing to standardized and enhanced diagnostic capabilities globally.

While the results are highly encouraging, future work could explore techniques to further refine the models' accuracy, address potential overfitting or underfitting issues, and investigate the integration of these models into clinical workflows for seamless and efficient brain tumor diagnosis and management.

Overall, this project represents a significant step forward in the field of medical image analysis, demonstrating the transformative power of deep learning in addressing the challenges posed by brain tumor classification and paving the way for improved patient care and treatment outcomes.

References

- [1] M. M. Badža and M. Barjaktarović, “Classification of brain tumors from mri images using a convolutional neural network,” *Applied Sciences*, vol. 10, no. 6, 2020. [Online]. Available: <https://www.mdpi.com/2076-3417/10/6/1999>
- [2] ———, “Segmentation of brain tumors from mri images using convolutional autoencoder,” *Applied Sciences*, vol. 11, no. 9, 2021. [Online]. Available: <https://www.mdpi.com/2076-3417/11/9/4317>
- [3] B. Mallampati, A. Ishaq, F. Rustam, V. Kuthala, S. Alfarhood, and I. Ashraf, “Brain tumor detection using 3d-unet segmentation features and hybrid machine learning model,” *IEEE Access*, vol. 11, pp. 135 020–135 034, 2023.
- [4] A. A. Akinyelu, F. Zaccagna, J. T. Grist, M. Castelli, and L. Rundo, “Brain tumor diagnosis using machine learning, convolutional neural networks, capsule neural networks and vision transformers, applied to mri: A survey,” *Journal of Imaging*, vol. 8, no. 8, 2022. [Online]. Available: <https://www.mdpi.com/2313-433X/8/8/205>
- [5] A. Ali, Y. Wang, and X. Shi, “Segmentation and identification of brain tumour in mri images using pg-oneshot learning cnn model,” *Multimedia Tools and Applications*, pp. 1–22, 2024.
- [6] S. Patil, D. Shelke, N. Gavhane, S. Sonawane, and V. Patankar, “Detection and classification of brain tumor using convolutional neural network,” in *Proceedings of the 6th International Conference on Communications and Cyber Physical Engineering*, A. Kumar and S. Mozar, Eds. Singapore: Springer Nature Singapore, 2024, pp. 183–190.
- [7] H. Chellakh, A. Moussaoui, A. Attia, and Z. Akhtar, “Mri brain tumor identification and classification using deep learning techniques,” *Journal homepage: http://ieta.org/journals/isi*, vol. 28, no. 1, pp. 13–22, 2023.
- [8] A. Ratnakar and N. Chougule, “Deep learning-based brain tumor classification using convolutional neural networks and mri images.”
- [9] J. Zhang, X. Lv, H. Zhang, and B. Liu, “Aresu-net: Attention residual u-net for brain tumor segmentation,” *Symmetry*, vol. 12, no. 5, 2020. [Online]. Available: <https://www.mdpi.com/2073-8994/12/5/721>
- [10] L. Ali, F. Alnajjar, H. Jassmi, M. Gochoo, W. Khan, and M. Serhani, “Performance evaluation of deep cnn-based crack detection and localization techniques for concrete structures,” *Sensors*, vol. 21, p. 1688, 03 2021.