Sadia Akther

MRI Brain Tumor Classification

Problem Statement:

Brain tumors are among the deadliest conditions, contributing significantly to global disability and mental health issues. Early and accurate detection plays a crucial role in patient survival and treatment planning; however, manual diagnosis from MRI scans is time-consuming, requires specialized expertise, and is prone to human error. To address this challenge, I develop an image processing-based pipeline to build and evaluate deep learning and machine learning models, including Convolutional Neural Networks (CNN), Visual Geometry Group 16 (VGG-16), Support Vector Machines (SVM) and Artificial Neural Network (ANN). These models are trained to classify MRI scans into tumor and no-tumor categories, and their classification performances are compared to determine the most effective approach for improving diagnostic accuracy. This work aims to enhance computer-aided diagnosis systems, supporting radiologists and potentially leading to faster, more reliable brain tumor detection.

Data Wrangling:

Since this project is primarily focused on preprocessing and modeling, the initial step was to prepare the dataset for analysis. I began by uploading the MRI dataset consisting of two classes: tumor and no tumor. Using the glob package, I identified and confirmed the dataset size: 155 images with tumors and 98 images without tumors, giving a total of 253 MRI images.

To enable model training, the images were transformed into arrays and concatenated into a single dataset, while corresponding labels were generated to distinguish between tumor and no-tumor categories. The final dataset was shaped into (253, 224, 224, 3), representing the number of samples, image height, width, and color channels. The labels were shaped into (253,), corresponding to the binary classification task. Separately, the tumor images had a shape of (155, 224, 224, 3), while the no-tumor images had a shape of (98, 224, 224, 3).

Exploratory Data Analysis (EDA):

To begin the analysis, numeric labels were assigned to the dataset: 1 for tumor and 0 for no tumor, since binary classification requires numerical encoding. The class distribution was then examined (Fig. 1), revealing that tumor images accounted for 61.3% of the dataset, while no-tumor images represented 38.7%. This imbalance indicated a potential risk of biased model performance, where the model might favor the majority class. To mitigate this issue, augmentation techniques such as rotation, flipping, and zooming were applied during preprocessing to balance the classes and improve generalization.

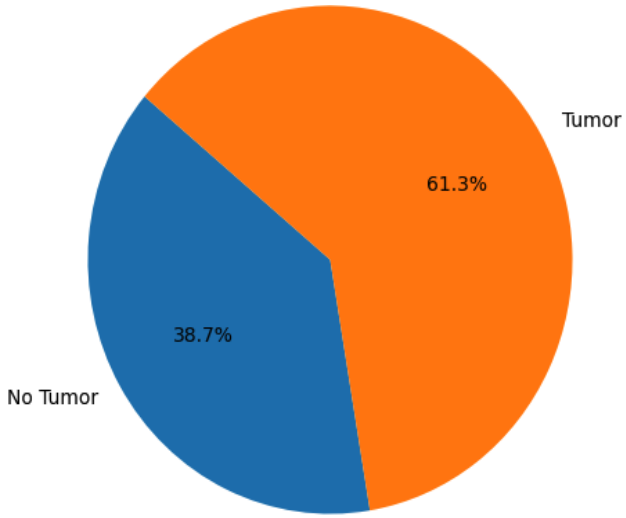

Figure 1: Distribution of Tumor and No Tumor Images

In addition to class counts, the dataset was visually explored by randomly sampling and displaying MRI scans from both categories. A function was implemented to select a specified number of tumor and no-tumor images without replacement and plot them in separate figures (Fig. 2). This step provided a visual quality check of the data, confirming that images were loaded correctly, uniformly resized, and free from obvious errors. Furthermore, these visualizations highlighted subtle differences between tumor and no-tumor images, offering intuitive insight into the classification challenge.

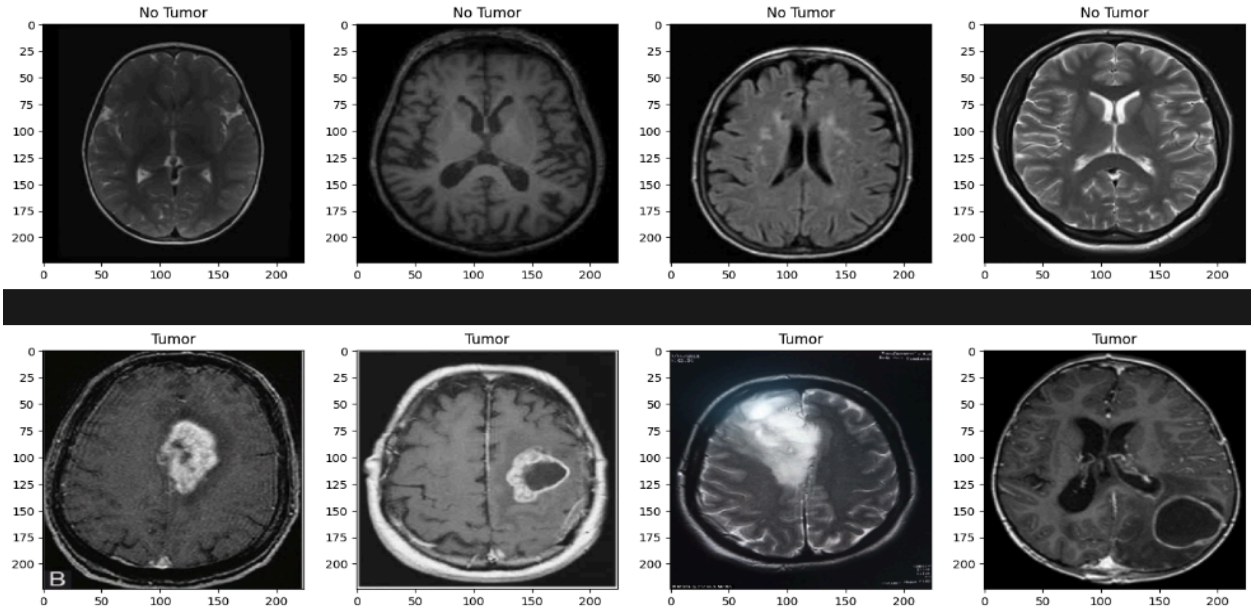

Fig. 2: Tumor and No Tumor Images

To further examine image characteristics, pixel intensity distributions were plotted using histograms. This analysis revealed the brightness and contrast levels across the dataset, highlighting areas where normalization would be beneficial for ensuring consistent input to the models. (Fig.3)

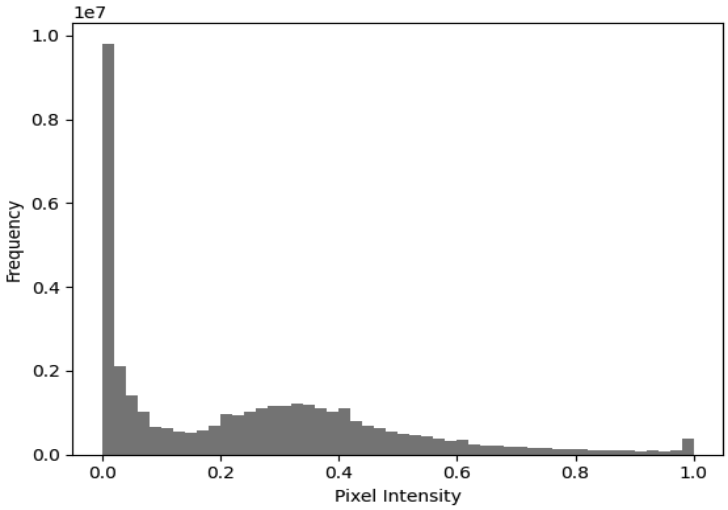

Figure. 3: Pixel Intensity Distribution

Pre-Processing

In the preprocessing phase, **data augmentation** was applied to address the imbalance between tumor and no-tumor images and to enhance the overall dataset quality for training. Augmentation is a technique that generates new training samples by applying transformations such as rotation, shifting, shearing, brightness adjustments, and flipping to existing images. By creating synthetic variations, augmentation increases the effective dataset size and introduces diversity, which helps prevent overfitting. In this project, augmentation was particularly useful for balancing the dataset, since the number of tumor images (61.3%) exceeded the number of no-tumor images (38.7%). Generating additional no-tumor images ensured that both classes were more evenly represented, reducing the risk of bias toward the majority class. Furthermore, augmentation improved model robustness by allowing it to learn from a wider range of possible variations in MRI scans, making it better suited to generalize when classifying unseen data. The new distribution of MRI images after augmentation was much more balanced, as shown in Figure 4. The dataset increased from 253 images (155 tumor, 98 no tumor) to 1,732 images (894 tumor and 838 no tumor). This adjustment resolved the imbalance issue, providing the model with a more reliable dataset for training.

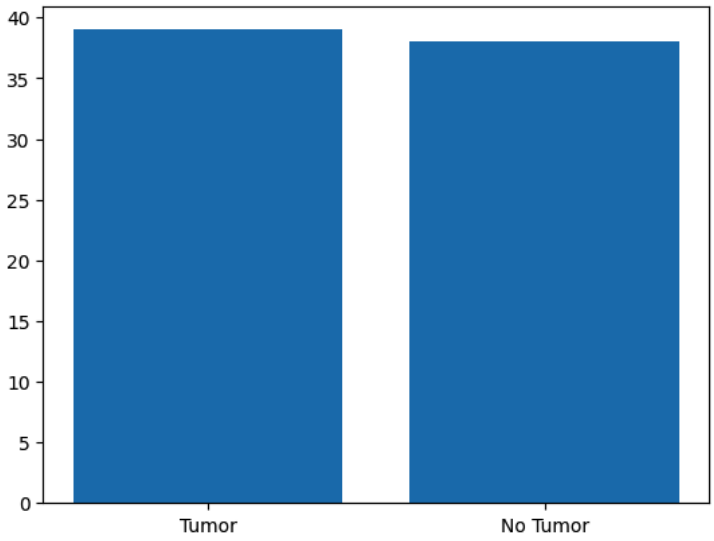

Figure. 4: MRI Image Distribution

Next part of this project is cropping images to improve the model performance and reduce noise in the dataset. This step ensured that the model focused only on the brain region while ignoring irrelevant background and skull areas. The process began by converting the image to grayscale and applying Gaussian blur to reduce noise. Next, thresholding and morphological operations (erosion and dilation) were used to separate the brain from the background. The largest contour was then identified, which corresponded to the outline of the brain, and its extreme points (top, bottom, left, and right) were located. Using these boundary points, the image was cropped to include only the brain area, effectively removing irrelevant background and skull regions. This preprocessing step was essential to ensure that the model concentrated solely on the region of interest, reducing computational complexity, minimizing distractions from non-brain areas, and ultimately improving classification accuracy. To better understand the process, four subplots are generated (Fig. 5), each showing a different stage of the transformation:

1. **Original Image** – This is the raw MRI scan as it appears in the dataset. It contains the entire brain region along with background areas that are not relevant for tumor classification. Displaying the original image provides a baseline to compare the subsequent preprocessing steps.
2. **With Contour** – In this subplot, a contour is drawn around the detected tumor region. The contour is obtained by thresholding and finding the largest connected component in the image. This highlights the tumor boundaries, making it clear where the abnormal region is located within the brain scan.
3. **Extreme Points** – Here, the extreme points of the contour are marked on the image: the leftmost, rightmost, topmost, and bottommost coordinates. These points are used to

define a bounding box around the tumor. By identifying these points, the function determines the exact region that needs to be cropped out from the original image.

4. **Cropped Image** – Finally, the image is cropped using the bounding box defined by the extreme points. This results in a zoomed-in view that focuses only on the tumor region, eliminating unnecessary background. The cropped image is crucial for training deep learning models, as it reduces noise and ensures the model learns features specific to the tumor instead of irrelevant parts of the scan.

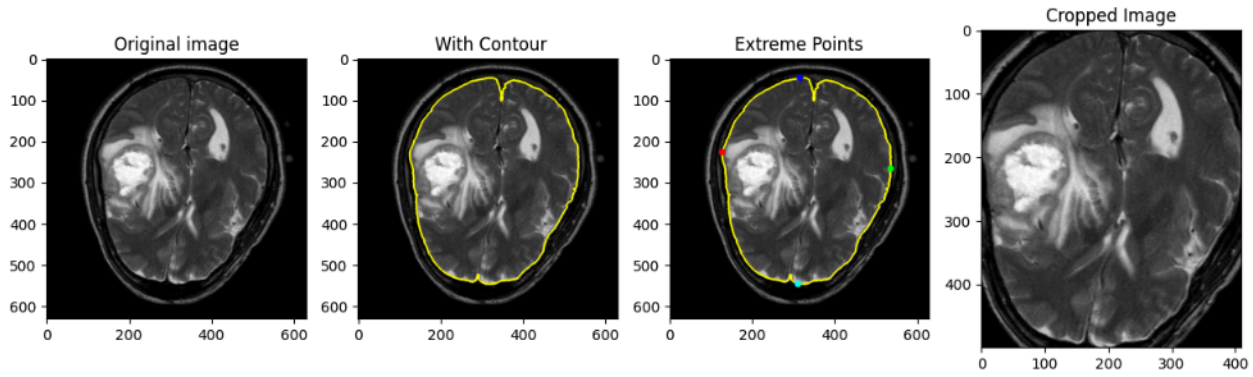

Figure 5: Cropped Images

Following augmentation and cropping, the MRI images were split into three subsets: **80% for training, 10% for testing, and 10% for validation**. This division ensures that the model can learn effectively while being evaluated on unseen data at different stages. The training set is used to fit the model, the validation set helps tune hyperparameters and prevent overfitting, and the test set provides an unbiased evaluation of the final model performance.

The splitting process was carried out separately for both tumor (Yes) and no tumor (No) classes to maintain balanced distributions across all subsets. The procedure followed these steps:

1. **Training Set (80%)** – The first 80% of images from each class (yes and no) were copied into the respective training folders. There are 1385 images belonging to 2 classes (yes and no). This provides the model with the majority of the data for learning patterns and features.
2. **Testing Set (10%)** – Next, 10% of the images were selected and copied into the testing folders. There are 172 images belonging to 2 classes (yes and no). These images remain completely unseen during training and validation, making them suitable for the final performance evaluation.
3. **Validation Set (10%)** – Another 10% of images were copied into the validation folders. There are 172 images belonging to 2 classes (yes and no). This set is used during training to tune the model and check how well it generalizes before reaching the test stage.

To confirm the dataset was properly divided, the number of images in each folder was programmatically verified. A bar chart (Figure 6) was created to visualize the distribution of tumor and no-tumor images across subsets. The training set contained 715 tumor images and 670 no-tumor images, while both the validation and testing sets included 89 tumor and 83 no-tumor images each. These results confirmed that the dataset was well balanced across all subsets, ensuring fair learning and reliable evaluation of the model.

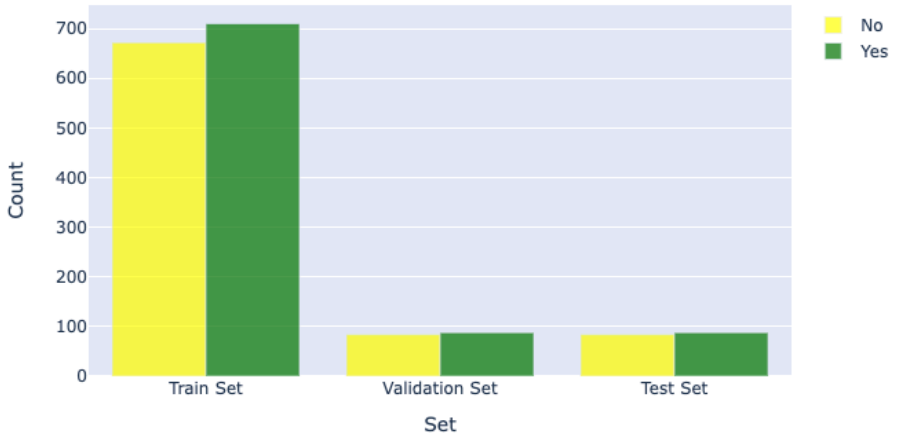

Figure 6: Distribution of tumor and no tumor images across training, validation, and test sets

Modeling

CNN:

For this part of the project, a Convolutional Neural Network (CNN) was chosen because CNNs are particularly effective for image classification tasks such as MRI brain tumor detection. CNNs automatically learn hierarchical features from the data: the early layers capture simple patterns like edges and textures, while deeper layers identify complex structures, such as tumor shapes and boundaries. This ability to learn directly from pixel data makes CNNs more powerful than traditional methods that rely on manual feature extraction. In medical imaging, where subtle variations in tissue structure are important, CNNs are especially beneficial for detecting tumors with high precision.

The performance of the CNN in this project confirmed its effectiveness. The model achieved a training accuracy of 93.75% with a loss of 0.1726, while maintaining a validation accuracy of 91.87% with a validation loss of 0.1879 (Fig.7 & 8), showing that the network generalized well without significant overfitting. On the completely unseen test set, the CNN

reached an impressive accuracy of 91.87% with a test loss of 0.1879, demonstrating its reliability in classifying MRI images as either tumor or no-tumor.

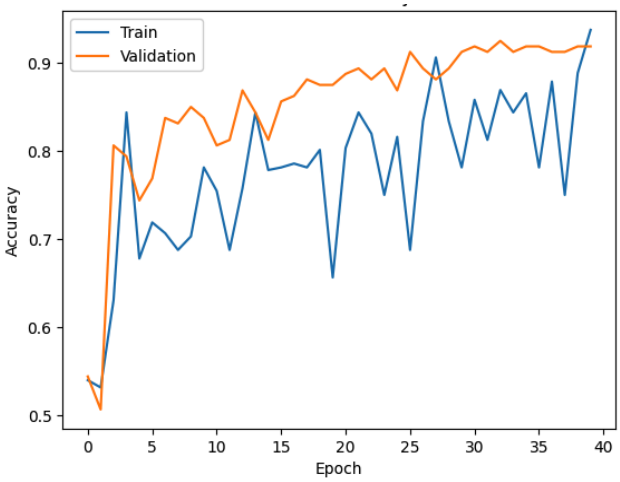

Figure 7: Model Accuracy

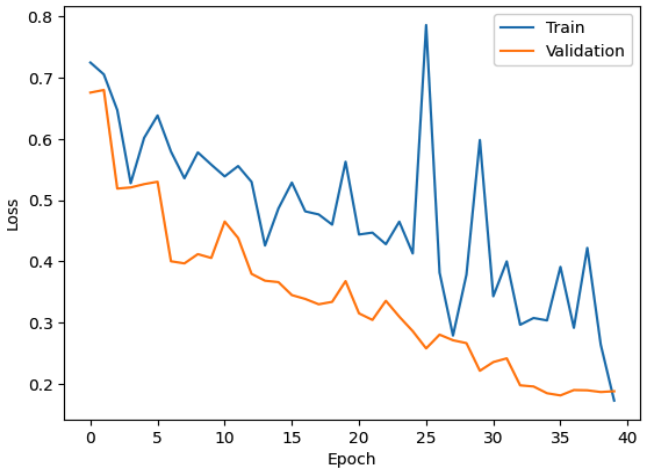

Figure 8: Model Loss

After training the Convolutional Neural Network (CNN), the model weights were saved and reloaded to evaluate performance on both the validation and test sets. The CNN achieved a validation accuracy of 92.44% with a loss of 0.1768 and a test accuracy of 92.44% with the same loss of 0.1768, indicating consistent generalization across unseen data. These results show that the network was able to learn discriminative features from MRI images without significant overfitting.

To further assess performance, additional metrics were computed using the test set. The CNN achieved a precision of 88.78%, meaning that when the model predicted a tumor, it was correct nearly 9 out of 10 times. The recall was 97.75%, demonstrating the model's strong ability to correctly identify patients with tumors. The F1-score of 93.05% balances precision and recall, showing overall robust classification performance. The confusion matrix (Fig.9) illustrates the classification results: out of all no tumor images, 72 were correctly classified and 11 were misclassified as tumors, while for the tumor class, 87 were correctly identified with only 2 misclassified as no tumor. These results indicate that the CNN is highly effective in detecting tumors, especially in minimizing false negatives, which is crucial in medical applications where missing a tumor case could have severe consequences.

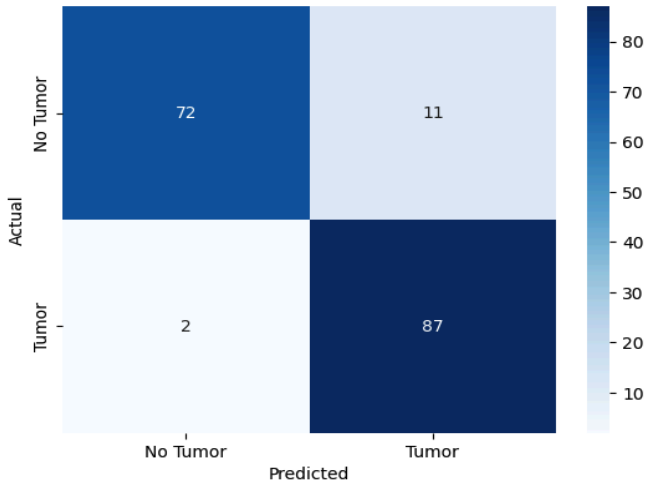

Figure 9: Confusion Matrix CNN

VGG16:

The Visual Geometry Group (VGG16) model was used in this project because it is a proven convolutional neural network (CNN) architecture known for its simplicity and effectiveness in image classification tasks. VGG16 consists of 16 layers of convolutional and fully connected layers that progressively extract deeper features from images, starting from basic edges and textures to more complex shapes and patterns. In this project, the pre-trained VGG16 model was applied through transfer learning, where the original convolutional layers were frozen to retain their powerful feature extraction capabilities, while new fully connected layers were added and trained specifically on the MRI brain tumor dataset. This approach allowed the model to leverage general image features learned from large-scale datasets like ImageNet and adapt them to the specialized task of distinguishing between tumor and non-tumor MRI images.

The performance of the VGG16 model in this project demonstrated its ability to extract meaningful features from MRI images. Throughout training, the model reached 91% accuracy with loss of 41% and the validation accuracy at 87% with loss of 33% (Fig. 10 & 11). These results highlight the effectiveness of transfer learning, as the pretrained VGG16 layers provided robust feature representations that enhanced classification performance on the MRI dataset.

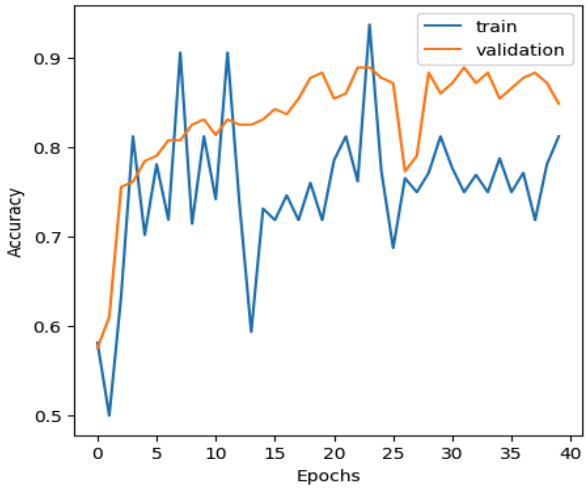

Figure 10: VGG16 Accuracy

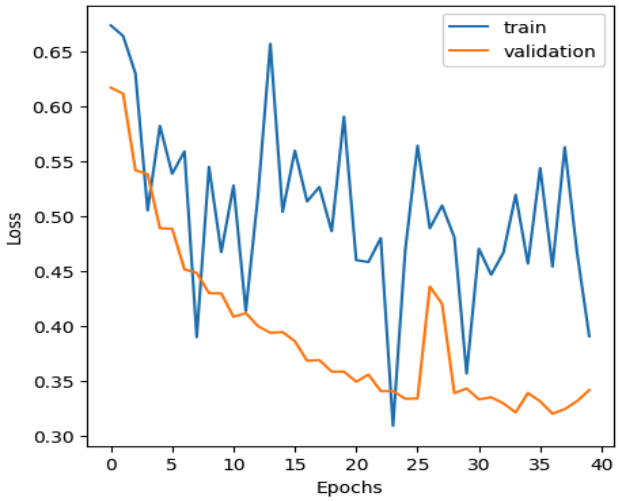

Figure 11: VGG16 Loss

Finally, the evaluation results show that the VGG16-based CNN achieved strong classification performance. On the test dataset, the model reached an accuracy of 84.88%, with precision of 79.44%, recall of 95.51%, and an F1-score of 86.73%. The high recall indicates that the model is particularly effective at detecting tumor cases, minimizing the risk of false negatives, which is critical in a medical setting. The confusion matrix (fig. 12) reveals that out of 172 test images, the model correctly classified 61 no tumor cases and 85 tumor cases, but misclassified 22 no-tumor images as tumors and 4 tumor images as no tumor. Although the VGG16 model performed slightly lower than the CNN model, however; results demonstrated that it is highly effective for medical imaging, achieving reliable accuracy and minimizing false negatives, which is especially critical in healthcare applications.

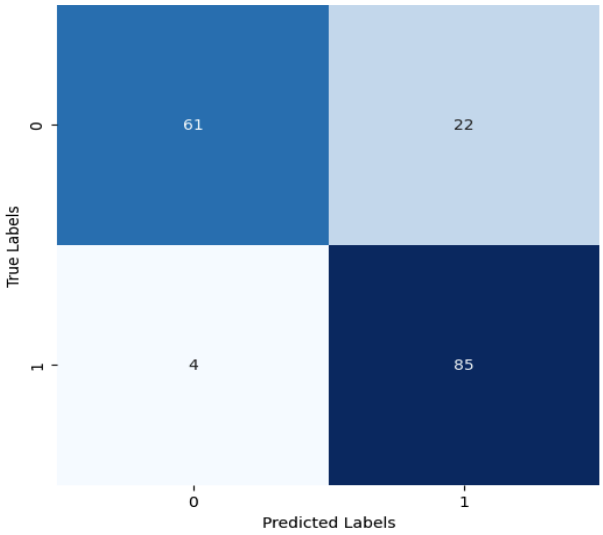

Figure 12: Confusion Matrix VGG16

SVM:

To complement the CNN and VGG16 models, a Support Vector Machine (SVM) classifier was also implemented using traditional feature descriptors. Before training the SVM, the extracted features were standardized, ensuring that all features contributed equally by rescaling them to have zero mean and unit variance. Using Principal Component Analysis (PCA) for dimensionality reduction (95% variance explained), the SVM model was tuned with GridSearchCV. The best configuration was found to be C=10, gamma=scale, and kernel= rbf. This indicates that a non-linear decision boundary was necessary to separate tumor and non-tumor classes effectively, and that a higher penalty on misclassification improved performance without leading to overfitting.

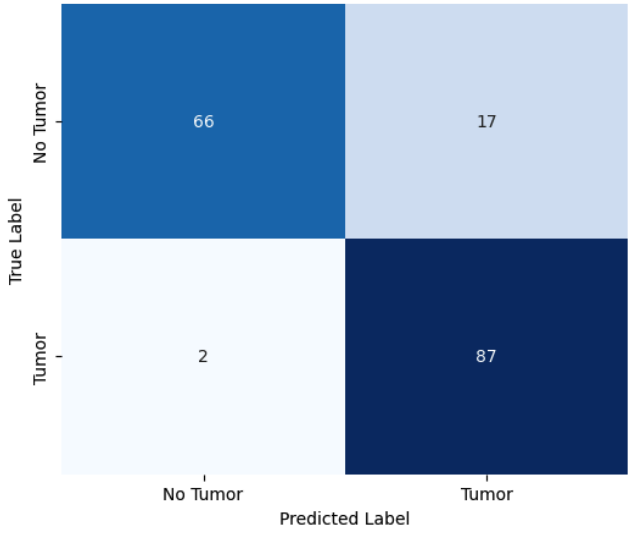

Figure 13: Confusion Matrix SVM

The SVM model achieved 89% accuracy with a strong balance between precision and recall across classes. While the model shows a very high recall of 0.98 for tumor detection, ensuring most tumor cases are identified, its slightly lower precision of 0.84 indicates the presence of some false positives. The F1-score of 0.89 reflects balanced and effective classification performance. Furthermore, the confusion matrix illustrates 66 true negatives, 87 true positives, and relatively few misclassifications with 17 false positives and 2 false negatives (Fig. 13). Additionally, the Receiver Operating Characteristic (ROC) curve yielded an Area Under the Curve (AUC) of 0.91, confirming the SVM's excellent discriminatory power, meaning the model is highly effective at distinguishing between tumor and no-tumor MRI scans (Fig. 14).

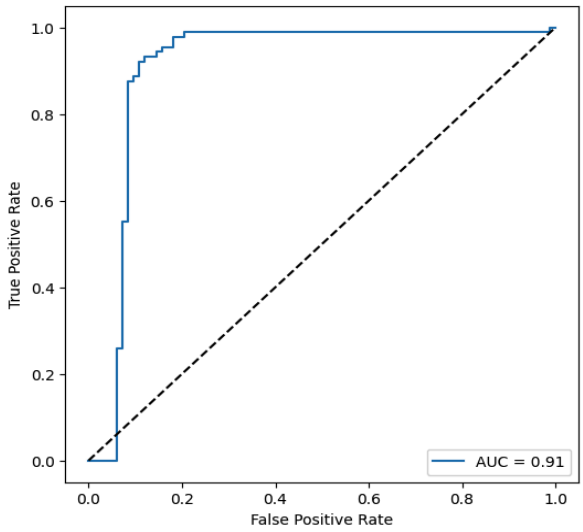

Figure 14: SVM ROC Curve

ANN:

The final part of this project model was An Artificial Neural Network (ANN), which was built using the VGG16 convolutional base for feature extraction, followed by fully connected layers for classification. The CNN feature maps from VGG16 were flattened into a one-dimensional vector and passed through dense layers. After training for 10 epochs using the training and validation generators, the ANN achieved a training accuracy of 85.49% with a loss of 0.3063, and a validation accuracy of 95.35% with a validation loss of 0.1662 (Fig.15 & 16), demonstrating strong generalization and effective learning on unseen data.

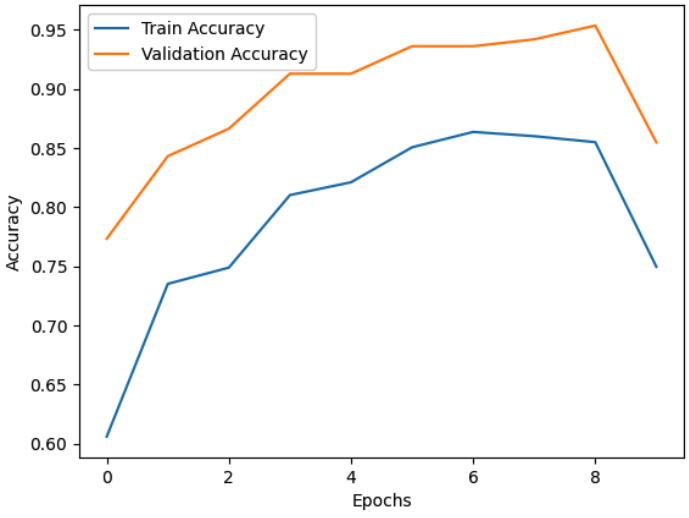

Figure 15: ANN Model Accuracy

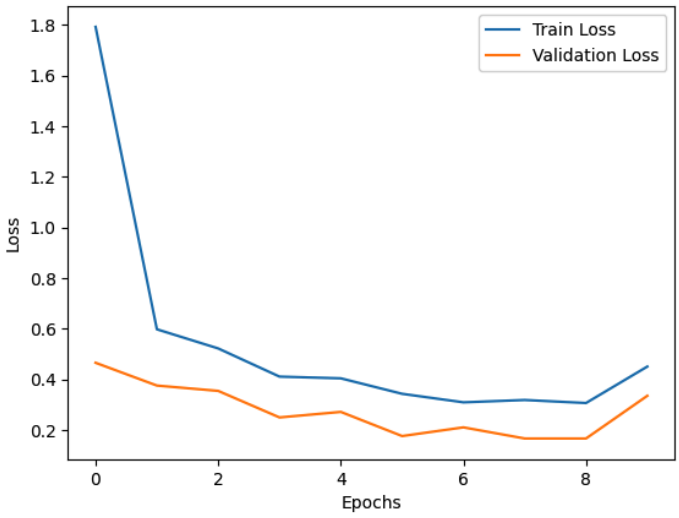

Figure 16: ANN Model Loss

The final model, ANN, achieved an accuracy of 85.47%, demonstrating strong overall performance in classifying tumor and no-tumor MRI scans. The confusion matrix indicated that 83 no tumor images were correctly classified, while 25 tumor images were misclassified as no tumor, and 64 tumor images were correctly identified (Fig. 17). Furthermore classification report highlights precision as 0.84, recall as 0.86 and F1-score as 0.85, showing the model effectively identifies tumors but misses some cases.

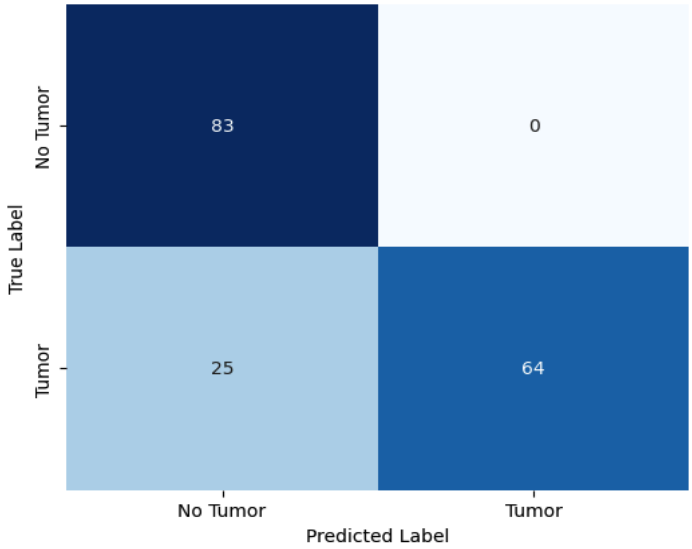

Figure 17: ANN Confusion Matrix

Final Takeaways:

	Accuracy	Precision	Recall	F1-Score
CNN	**92%**	**89%**	**98%**	**93%**
VGG16	85%	79%	96%	87%
SVM	89%	84%	**98%**	89%
ANN	85%	88%	86%	85%

Figure 18: Classification Performance

Among the models, the CNN achieved the highest overall accuracy at 92%, demonstrating its superior ability to correctly classify both tumor and no-tumor MRI images. The SVM followed closely with an accuracy of 89%, while VGG16 and ANN both achieved 85% accuracy. In terms of precision, the ANN slightly outperformed the CNN and SVM for the no-tumor class with 88%, indicating a lower rate of false positives. VGG16 had the lowest precision at 79%, suggesting it produced more false positives than the other models. For recall,

the CNN and SVM both achieved an impressive 98%, showing that these models were highly effective in detecting tumor cases and minimizing false negatives. VGG16 also performed well with 96% recall, while ANN had the lowest recall at 86%, indicating it missed a higher proportion of tumor cases. The F1-score, which balances precision and recall, further highlights CNN's strong performance with a score of 93%, followed by SVM at 89%, VGG16 at 87%, and ANN at 85%. Overall, these results suggest that while all models are capable of tumor detection, the CNN provides the most reliable performance across all metrics, making it the most suitable choice for this MRI classification task.

Future Research:

A potential extension of this project is to develop a web-based application that allows users to upload MRI brain images and receive real-time predictions regarding the presence of tumors. This platform could leverage the trained CNN or hybrid models to provide accurate, automated assessments, making advanced diagnostic tools more accessible to clinicians, researchers, and even patients. By integrating the model into a user-friendly interface, the system would streamline MRI analysis, reduce manual workload, and facilitate early detection of brain tumors.