ANN Final Report
Deep Learning Based Brain Tumor Detection

Bekir Emirhan Akay (19290286), Mehmet Alpay(20290310)
Fatih Yalçın (20290372), Murat Tatlı (20290366)

January 29, 2024

1 Introduction

The incorporation of artificial intelligence (AI) applications into the healthcare domain stands as a pivotal advancement for our collective future. This research endeavors to significantly contribute to the healthcare sector, with a specific focus on the identification of brain tumors. Annually, a substantial number of individuals worldwide succumb to the detrimental effects of brain tumors and related health conditions.

The process of brain tumor detection entails the identification of tumor presence within the brain, typically achieved through the utilization of medical imaging modalities such as magnetic resonance imaging (MRI). The complexity of detecting brain tumors arises from variations in their location, shape, and size. In response to this challenge, the application of machine learning and deep learning techniques emerges as a promising avenue for solutions.

This study is undertaken to address a significant issue and contribute valuable enhancements to the field of medical diagnostics. Pioneering hardware and technological advancements play a pivotal role in ensuring that research in this domain remains contemporary, introducing fresh perspectives to existing knowledge. By harnessing the latest information, our objective is to present a novel study in this specialized field.

The primary aim of this research is to comprehend the current impediments encountered in the identification of brain tumors, devise solutions to overcome these challenges, and make a substantial contribution to diagnostic processes within the healthcare sector. In this context, our aspiration is to lay the groundwork for the application of artificial intelligence and deep learning methods in the realm of healthcare.

2 Related Work

While tumor images are our focus, methods used for evaluation for other types of brain abnormalities can be employed to handle the task. For example, random forest method was used for injured brain image segmentation. [1] Usage of multiple methods themselves be used in assemble or hierarchically is something studied. The latter was done by Mitra et al. who employed Markov random fields and Bayesian inference at the bottom of with random forest algorithm for lesion detection. [2] It is worth mentioning that in segmentation tasks involving classification in addition to metrics regarding success of classification, Dice Similarity Coefficient is utilized to measure the success pixel-wise. This can be said to be similar to grad-cam visualization which interests our work.

In the last decade, deep learning have become widely popular. Due to input sizes, convolutional neural networks (CNN) are the default choice in image processing. There are various convolutional neural network architectures.

A popular method is transfer learning. In transfer learning models are pretrained. While the model can be applied as is, fine tuning is possible by further training the model with the data at hand. One CNN architecture is U-Net which was developed with computer aided health condition evaluation in mind [3]. A network derived from U-Net was used for detection of brain cancer [4]

It is widespread that before processing, preprocessing is performed. Pereira et al. bias field distortion primarily [5].

Rai et al. [6] presented the LeU-Net model, showcasing a remarkable 98% accuracy rate for cropped images and achieving a commendable 94% accuracy for uncropped images. Noteworthy is the model's exceptional processing efficiency during simulation, with training times of only 244.42 seconds and 252.36 seconds for 100 epochs on uncropped and cropped images, respectively.

Tiwari et al. [7] employed a Convolutional Neural Network (CNN) to tackle the challenge of brain tumor classification. The introduced model has exhibited notable success in categorizing brain images into four distinct classes: the absence of a tumor, indicating that the provided MRI scan does not exhibit any tumor; glioma; meningioma; and pituitary tumor. Remarkably, the model achieved an impressive accuracy rate of 99%.

Siar et al. [8] employed the Softmax Fully Connected layer, achieving a notable 98.67% accuracy in image classification. In comparison, the CNN with the Radial Basis Function (RBF) classifier achieved 97.34%, and the Decision Tree (DT) classifier yielded 94.24%. Beyond mere accuracy, our comprehensive evaluation considered benchmarks such as Sensitivity, Specificity, and Precision, revealing the Softmax classifier as the top performer within the CNN framework during image testing. Our innovative approach combines feature extraction techniques with CNN for brain tumor detection, resulting in an impressive 99.12% accuracy on test data, underscoring its significant contribution to the field.

Ferdous et al. [9] present LCDEiT, a groundbreaking linear-complexity, data-efficient image transformer designed for small dataset training. Utilizing a teacher-student strategy and an external attention mechanism, the model achieves linear computational complexity relative to the number of patches. The teacher model, a custom gated-pooled convolutional neural network, imparts knowledge to the transformer-based student model for precise classification of MRI brain tumors. Remarkably, on benchmark datasets, including Figshare and BraTS-21, the model achieves average classification accuracy and F1-scores of 98.11% and 97.86%, and 93.69% and 93.68%, respectively. This highlights the efficacy of LCDEiT in optimizing performance with limited data resources.

Tabatabaei et al. [10] introduce a two-branch parallel model seamlessly integrating the Transformer Module (TM) with the Self-Attention Unit (SAU) and Convolutional Neural Networks (CNN) for brain tumor classification in MR images. Additionally, they present iResNet, a lightweight and enhanced CNN architecture tailored for discerning tumor features from MR images. Using 20% of the 3064 slices from the Figshare dataset as unseen images, the remaining data is partitioned into 60%, 20%, and 20% for training, validation, and testing, respectively. When compared with other CNN networks like iVGG and iDensNet, iResNet stands out with an impressive accuracy of 99.30%, reinforcing its efficacy in tumor classification.

Dishar et al. [11] have achieved an exceptional accuracy of 99.83% by strategically amalgamating the resilient Xception and NASNETMobile network architectures. This integrated approach effectively addresses the often-overlooked challenge of overfitting in existing networks, ensuring heightened reliability and improved generalization capabilities.

Aulia et al. [12] employ the Clip Limit Adaptive Histogram Equalization (CLAHE) method for image enhancement, specifically in segmenting brain MRI. Subsequently, brain tumor classification on MRI is conducted using the Visual Geometry Group-16 Layer (VGG-16) model. CLAHE is selectively applied, with instances including FLAIR image enhancement in green color and CLAHE application in the Red, Green, Blue (RGB) color space. Experimental results showcase the highest performance metrics, with accuracy, precision, and recall reaching 90.37%, 90.22%, and 87.61%, respectively. Notably, the CLAHE method in the RGB Channel, coupled with the VGG-16 model, demonstrates remarkable efficacy in predicting oligodendroglioma classes in RGB enhancement, boasting precision at 91.08% and recall at an impressive 95.97%.

Younis et al. [13] utilized the Faster CNN approach, employing the VGG 16 architecture as the primary network to generate convolutional feature maps, subsequently classifying these to yield suggestions for tumor regions. Our method effectively identified brain tumors in MR images, surpassing current conventional approaches in testing data. The algorithm demonstrated superior performance with precision values of 96%, 98.15%, and 98.41%, and F1-scores of 91.78%, 92.6%, and 91.29%, respectively. Furthermore, an excellent overall accuracy was achieved, with CNN at 96%, VGG 16 at 98.5%, and the Ensemble Model at 98.14%.

Maqsood et al. [14] conducted experiments on the BraTS 2018 and Figshare datasets using the MobileNetV2 architecture and eXplainable Artificial Intelligence (XAI) techniques. The results demonstrated impressive classification accuracy rates of 97.47% and 98.92%, respectively.

Ullah et al. [15] introduced the TumorDetNet model utilizing leaky ReLU activation. The model achieved remarkable results, accurately identifying brain tumors with a rate of 99.83%. Furthermore, it demonstrated ideal accuracy in distinguishing between benign and malignant brain tumors, reaching 100%, while achieving a high accuracy of 99.27% in classifying meningiomas, pituitary tumors, and gliomas.

Minarno et al. [16] utilized a deep learning model, DenseNet 201, in conjunction with Support Vector Machines (SVM), to classify three types of brain tumors—glioma, meningioma, and pituitary. The study achieved outstanding accuracy results of 99.65% through a training-testing data split of 80% and 20%, oversampled using the SMOTE method. The combination of DenseNet 201 and SVM proved effective in accurately categorizing brain tumors in MRI images, with the highest accuracy observed in this study being 99.65%.

3 Our Approach

We developed 2 main methods to predict brain tumors. These methods are listed in below. Moreover, We added Gaussian's Noise our dataset to challenge how our model behaves against distorted data.

	Glioma	Meningioma	Pituitary	no_tumor
Training	4687	4421	4578	1441
Validation	950	900	920	280
Test	950	900	920	280

Figure 1: Dataset Partition

3.1 Dataset

We find our dataset from Kaggle. We chose Kaggle as a data source because it is a good hub for data engineers. There are lots of datasets for various problems. For our problem we chose Brain Tumors MRI images[24]. It consists of 21227 images from 4 classes of brain tumor type which are Glioma, Meningioma, Pituitary and No tumor. Dataset partitioning is shown in figure-1. Also, we chose imbalanced dataset to avoid false negative as much as possible. Therefore, our model is prone to false positives, however this is not an issue since our model is a helper tool for doctors. If it comes up with positive result, a doctor can always double check.

3.2 Preprocessing

For preprocessing we standardized all of the images via using normalization and resizing images. We choose 224x224 for size of images in the dataset. This size is generally used in literature and similar works. Also, we experiment CLAHE (Contrast Limited Adaptive Histogram Equalization) filter to enhance image quality. The filter basically differentiates contrasted regions of the image thus improves details without disturbing image. The result of filter is increase contrast important detail of the images. The filter uses variant of the histogram equalization. Moreover, we generate data augmentation to get more data. These data augmentation technique are shearing, random zooming and horizontal flip. In figure-3 we showed result of the CLAHE filter and figure-2 we gave some example of data augmentation.

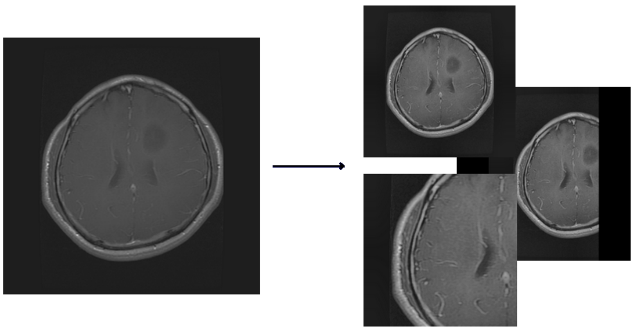


Figure 2: Some Example of Our Data Augmentation

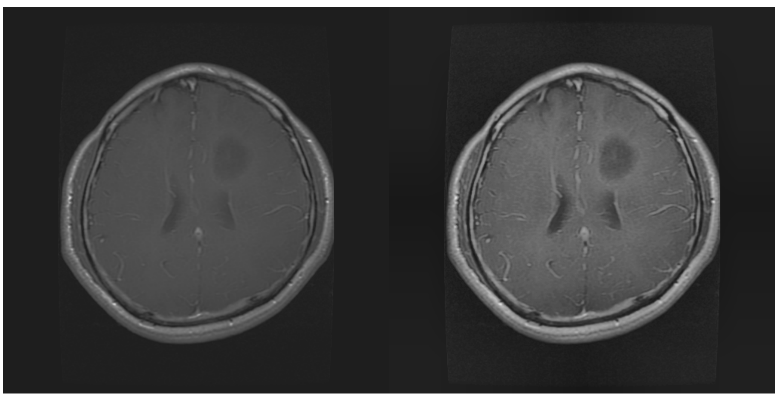


Figure 3: Example CLAHE Filter Result on Our Dataset(Right: Original, Left: Filtered)

3.3 Convolutional Neural Networks

In this section, we experiment some of the well known algorithms that is used in literature. We choose 6 architectures to show our model. In CNN methods, first we preprocess images via using data augmentations and standardization. We wrote a base model which is main architecture of the our model such as MobileNetV2, ResNet152V2. Then, we added Fully Connected (FC) layers to make better generalization and increase our models capacity. We used dropout layer because some of our models was being overfitting and dropout layer is one of the key point of solving overfitting problem. As shown in figure-5, we made our FC layer content with dropout layer. Lastly, we added Softmax layer to get class scores for our model because we have 4 classes and Softmax is performing well for multi-class problems. In figure-4 we showed basic model schema for our CNN based approach. Also, in our model we used two FC layers based on 1024 dense, then dropout layer for overfitting and ReLU activation. After first part of layers we added two extra layers based 512 dense, and dropout layer for overfitting and ReLU activation. Lastly, we chose to add one layer based on 256 dense and dropout layer and Softmax function for 4 class problem. The reasons for choosing CNN algorithm are explained in below.

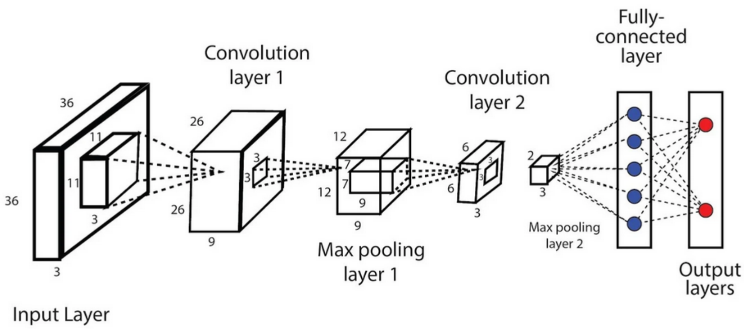


Figure 4: Basic Schema of Convulutional Neural Network [23]

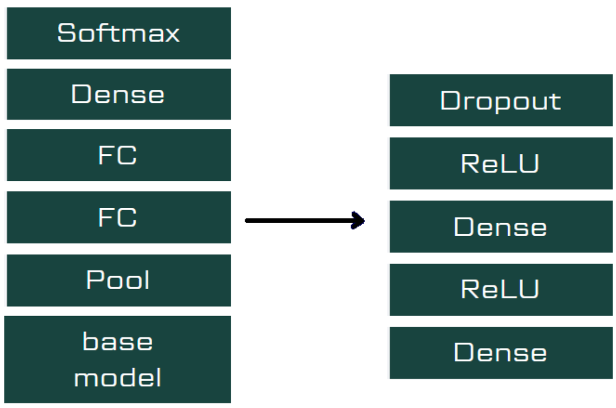


Figure 5: Our Method for Adding to Base Model

Advantages

- CNN architectures preservers spatial location and this makes a good feature extraction over image.
- CNN provides very deep architectures so it can learn its task from very detailed images and problems.
- Our model can even generalize with different type of filters and Gaussian Noise.
- Process of constructing CNN model and increasing capacity of our model are easy to make it.
- We have lots of dataset, and we can train our model with more epochs so results are solid.

Disadvantages

- CNN architectures are very big so learning process is very low especially for our problem because of the resolution.
- It is difficult to detect that if our model capacity is proper for dataset or not because sometimes validation results was higher than training results.

Our experiments mainly conducted on CNN architectures. Therefore, we focus on CNN architectures while choosing model. Our criteria for choosing model is mainly performance/timing trade-off. We looked up ImageNet performance table and parameter number table.[17] We paid attention on accuracy in ImageNet as high as possible and parameter number as low as possible because of limited hardware issues. Moreover, we created batched data loader to avoid overflow RAM.

3.4 Transformers

Transformer models are quite useful models and relatively newly found method for classification. Normally, they are used by famous natural processing models such as BERT, GPT. [18] We chose Swin Transformer[22] model to experiment on our dataset. It is quite useful and relatively has less parameter than other known transformers model. Unlike Convolutional Neural Networks (CNNs) that typically focus on local features, Transformers can model long-range dependencies in the input data. This is particularly beneficial for images where the relationship between distant pixels can be crucial for understanding the content.[19] Transformers have shown remarkable success in multi-modal medical image classification. They can effectively fuse information from different modalities, which is crucial in medical imaging where different types of scans (e.g., MRI, CT, PET) provide complementary information[20]. Studies have shown that Transformer models, when pre-trained with a sufficient amount of data, are at least as robust as their CNN counterparts on a broad range of perturbations. They are also robust to the removal of almost any single layer[21].
We write Swin Transformer model with similar to CNN FC layers with dropout value of 0.5. Moreover, we used Batch Normalization to increase test result quality. We set dropout is very high because our Swin Transformer model prone to overfitting so we want to overcome this problem with these solution. The basic schema of the Swin Transformer model is shown in figure-6.

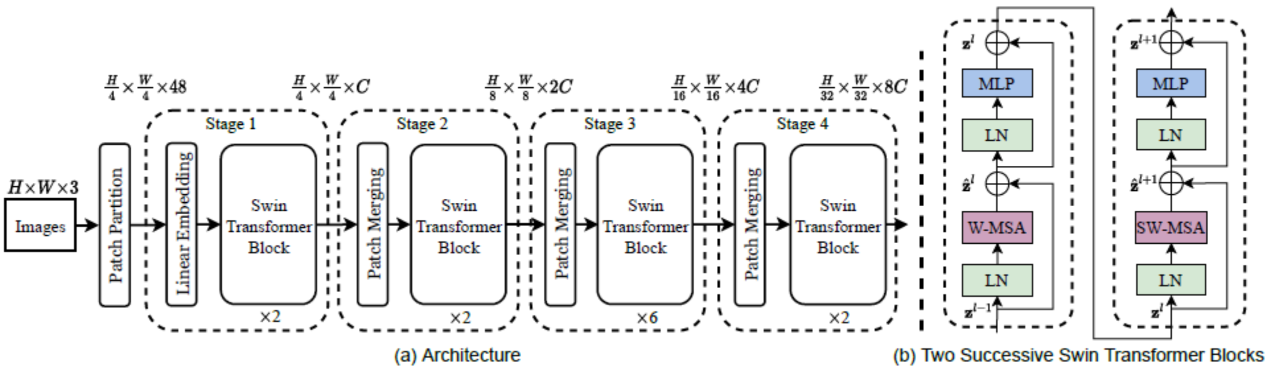


Figure 6: Swin Transformer Architecture Schema

Advantages

- They can effectively fuse information from different modalities, which is crucial in medical imaging where different types of scans provide complementary information, transformers advance multi-modal medical image classification.
- They are also robust to the removal of almost any single layer so modifying transformer model very easy.
- Transformers can model long-range dependencies in the input data. This is particularly beneficial for images where the relationship between distant pixels can be crucial for understanding the content.

Disadvantages

- They require huge amount of data to train efficiently.
- Their architecture is very complex so training process requires too much computational resource.

For preprocessing we standardized all of the images via using normalization and resizing images. We choose 224x224 for size of images in the dataset. This size is generally used in literature and similar works.

3.5 Experiment with Gaussian Noised Data

To make our problem more challenging, we add Gaussian Noise as 0.2 probability all of the images in the dataset. We tried 2 different model for this approach. 1st approach is trying only ResNet152V2 and 2nd approach is trying ResNet152V2 with CLAHE filter. Results are showed that CLAHE filtered model much more successful than unfiltered one. Even filtered model is near the most successful model in our work. The results is shown in Experimental Results section.

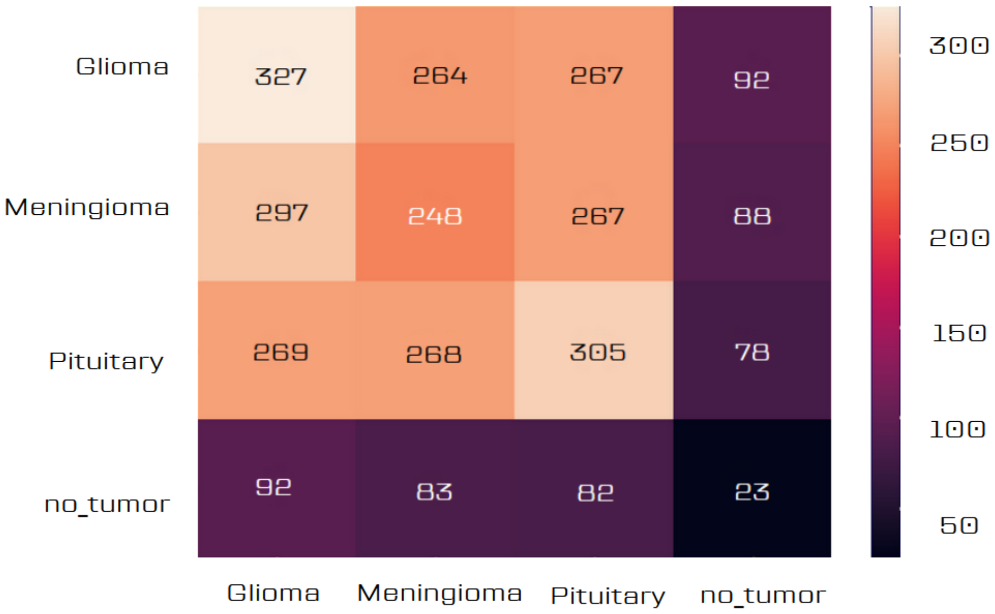


Figure 8: Confusion Matrix of ResNet152V2 model

3.6 Data Visualization

Data visualization is very crucial part of the any machine and deep learning tasks. With data visualization, we can analyze our results to develop more advanced and accurate models. For this purpose, we generated Gradient-weighted Class Activation Mapping (Grad-CAM) image and Confusion Matrix.

3.6.1 Grad-CAM

Grad-CAM is a technique that helps to explain the predictions of a deep neural network model for image classification tasks. It uses the gradients of the classification score with respect to the final convolutional feature map to identify the parts of an input image that most impact the classification. In figure-7 we give Grad-CAM image example.

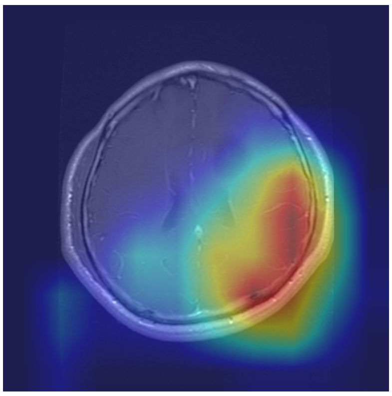


Figure 7: Example of Grad-CAM image in Our Dataset

3.6.2 Confusion Matrix

A confusion matrix serves as a concise tabular representation of a machine learning model's performance on test data. Comprising a 2x2 matrix, it reveals true positives, false negatives, false positives, and true negatives. This matrix facilitates the computation of metrics like accuracy, precision, recall, and F1 score for a comprehensive evaluation. Notably, we observed a lower accuracy in tumor class detection, indicating a need to address data imbalances for improved model performance. The results of the confusion matrix is shown in figure-8.

4 Experimental Results

In our proposed methodologies, consistent hyperparameters were employed to ensure controlled and comparable experiments across various CNN architectures. These hyperparameters include utilizing the SGD optimizer with a learning rate of 10e-̂3 and setting the number of epochs to 5 for transformers we set epoch number as 20. While this epoch count might seem low for typical image classification tasks, we leverage transfer learning to mitigate the necessity for a higher epoch count. The model weights are initialized using pre-trained weights from ImageNet, enhancing the model's initial learning capabilities. For each CNN architecture, we applied data augmentation techniques, including shearing to cut random parts of the image, random rotation, and flips. These augmentations play a crucial role in enhancing the model's performance by introducing diverse data to the training process. The original dataset comprises approximately 22k images, and through augmentation, the dataset is nearly doubled. This increase in data volume contributes to the robustness of our models. Furthermore, to mitigate the risk of overfitting, L2 regularization was incorporated into our training process.

Additionally, two new models, Swin Transformer and ResNet152V2WGNoise (ResNet152V2 with Gaussian Noised Images), have been introduced to the experimental setup. These models undergo the same training conditions, and their results are presented below alongside the previously evaluated CNN architectures.

When assessing the performance of Swin Transformer and ResNet152V2WGNoise models in comparison to the other architectures, it is noteworthy that these models exhibit distinct trends. Swin Transformer demonstrates robust results, showcasing competitive performance in terms of accuracy and loss metrics. On the other hand, ResNet152V2WGNoise, while maintaining commendable performance, may display variations in specific metrics.

In direct comparison with the existing models, Swin Transformer tends to yield superior results in certain aspects, demonstrating its efficacy in the image classification task. ResNet152V2WGNoise, while delivering satisfactory outcomes, might show nuances in performance when contrasted with the established architectures.Moreover we tried this task with ResNet152V2WGNoiseC(Resnet152V2 with Gaussian Noise and CLAHE image filter) which as more superior results to ResNet152V2WGNoise model because of CLAHE filter efficiency.

These models and their outcomes are presented below, juxtaposed with the previously evaluated CNN architectures.

Model Name	Train Loss	Train Acc	Train Precision	Train Recall	Train F1 Score
SwinTransformer	0.0480	0.9820	0.9850	0.9804	0.9766
ResNet152V2WGNoise	0.1362	0.9417	0.9545	0.9310	0.9154
ResNet152V2WGNoiseC	0.0643	0.9758	0.9824	0.9691	0.9668
MobileNetV2	0.0805	0.9717	0.9763	0.9687	0.9595
DenseNet121	0.0572	0.9807	0.9838	0.9783	0.9736
Xception	0.1716	0.9478	0.9580	0.9383	0.9285
ResNet50V2	0.0916	0.9682	0.9730	0.9636	0.9557
ResNet101V2	0.0531	0.9799	0.9832	0.9775	0.9736
ResNet152V2	0.0465	0.9827	0.9858	0.9805	0.9757

The name ResNet152V2WGNoise means that the ResNet152V2 model is trained by adding Gaussian Noise. ResNet152V2WChale means the ResNet152V2 model is trained using a Chale Filter and adding Noise.

In the table shown, ResNet152V2 again gives the best results for educational purposes. However, now that SwinTransformer and ResNet152V2 are performing head-to-head, it has become more difficult to distinguish which is the better model. Apart from this, DenseNet121 and ResNet101V2 also have good performance. Verification scores are shown below.

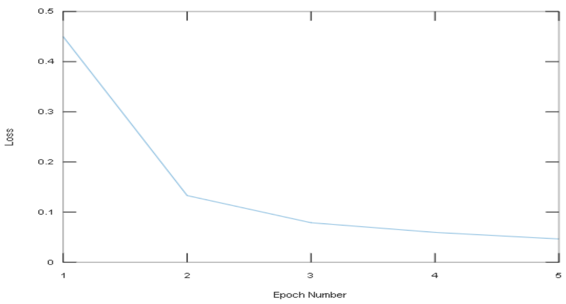
Model Name	Val Loss	Val Acc	Val Precision	Val Recall	Val F1 Score
SwinTransformer	0.3546	0.9111	0.9125	0.9098	0.8378
ResNet152V2WGNoise	0.2526	0.9200	0.9259	0.9134	0.8773
ResNet152V2WGNoiseC	0.1498	0.9508	0.9591	0.9449	0.9299
MobileNetV2	0.3617	0.8843	0.8915	0.8810	0.8807
DenseNet121	0.0859	0.9734	0.9734	0.9728	0.9636
Xception	0.2807	0.9184	0.9312	0.8967	0.8936
ResNet50V2	0.1973	0.9374	0.9424	0.9334	0.9220
ResNet101V2	0.1390	0.9590	0.9599	0.9587	0.9440
ResNet152V2	0.0905	0.9711	0.9724	0.9708	0.9604

In the provided table, DenseNet121 yields the most favorable outcomes for the validation set. Nevertheless, ResNet152V2 demonstrates results that closely rival those of DenseNet121. We contend that among the models considered, namely Resnet152V2WChale, DenseNet121, ResNet152V2, and ResNet101V2, these emerge as the top-performing models. The subsequent test results for models are presented below.

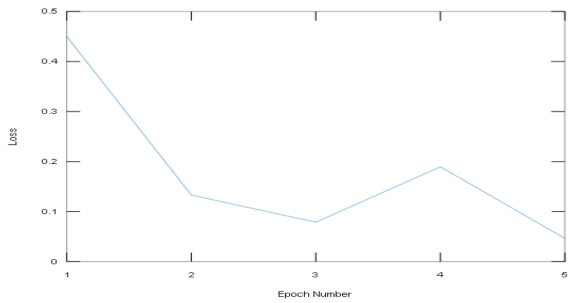
Model Name	Test Loss	Test Acc	Test Precision	Test Recall	Test F1 Score
SwinTransformer	0.4619	0.8954	0.8976	0.8938	0.8145
Resnet152V2WGNoise	0.3225	0.8987	0.9039	0.8915	0.8381
Resnet152V2WGNoiseC	0.2124	0.9439	0.9474	0.9387	0.9122
MobileNetV2	0.4298	0.8561	0.8739	0.8433	0.8173
DenseNet121	0.3907	0.8895	0.8895	0.8836	0.8054
Xception	0.3734	0.8830	0.8918	0.8725	0.8100
ResNet50V2	0.2227	0.9285	0.9325	0.9239	0.9065
ResNet101V2	0.2316	0.9439	0.9448	0.9433	0.9250
ResNet152V2	0.1020	0.9718	0.9731	0.9708	0.9561

When we look at the newly added models, ResNet152V2WChale and ResNet152V2 using Gaussian Noise seem to show better results than many models, but Swin Transformer could not give the expected result. When we look at the test results, the best model with our fine-tuning methods is again ResNet152V2. Moreover, we detected that no tumor class detection relatively less accurate than other classes because of data imbalances.

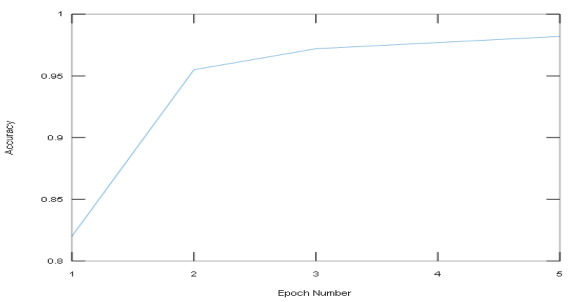
We also drew loss and accuracy graphs for ResNet152V2 model in figure below.



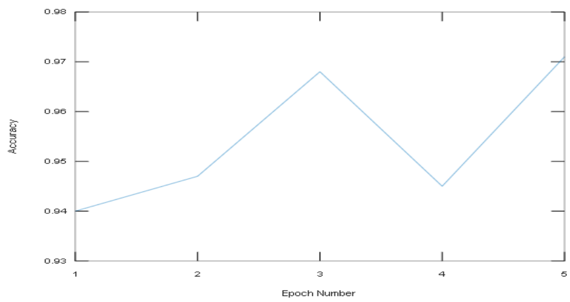
(a) Train Loss



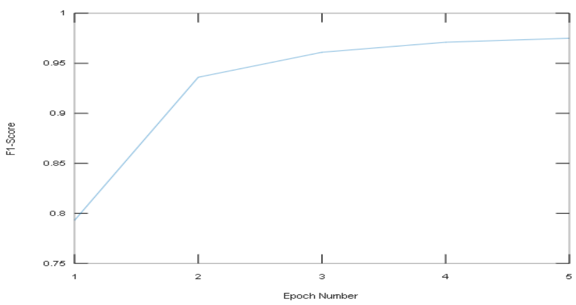
(b) Validation Loss



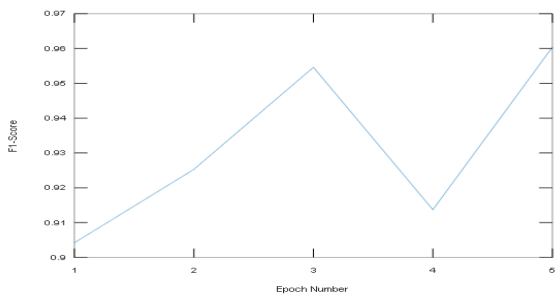
(c) Train Accuracy



(d) Validation Accuracy



(e) Train F1 Score



(f) Validation F1 Score

5 Conclusion

In our study, we observed that especially using deep learning methods and transformer models, can yield valuable insights. The presence of a large dataset and the inherent ease of the problem for the chosen models play a crucial role in achieving these results.

We also incorporated filter examples, such as the CLAHE filter, to explore different suggested methods and acquaint our model with various images. This was aimed at enhancing the overall performance of the model and assessing its potential for adaptation to a broader range of applications. Moreover, with the image filter Gaussian noise challenge can be easily solved. This filter increases accuracy greatly. We obtained that Resnet152V2 surpass all of the models that we tried. In the literature, we are in average for newest researches in some cases, our work even surpass them. However, in addition to a sufficient accuracy for literature for achieving success, we introduced noise to increase the difficulty of the dataset and simulate situations closer to real-world scenarios. This is the process through which we optimize our model, making it more demanding to better adapt to real-world applications.

In conclusion, our study highlights the successful use of deep learning and transformer models in the field of brain tumor detection. Our results are consistent with existing literature and demonstrate that the combination of various methods and filtering techniques can further improve model performance. This study will provide insights for future research and contribute to the advancement of technological developments in brain tumor detection.

5.1 Future Work

We plan to try more advanced transformer techniques such as LCDEIT model to increase general efficiency of our model. Also, we can try other activation and loss functions to increase gradient of our model. Secondly we want to increase accuracy via using more advanced data preprocessing techniques such as image filters, augmentation techniques. Also, we want to implement GAN models for synthetic data generation. These GAN models are very useful such problems to increase data.

References

[1] Rao, A., Ledig, C., Newcombe, V., Menon, D., & Rueckert, D. (2014). Contusion segmentation from subjects with traumatic brain injury: A random forest framework. In 2014 IEEE 11th international symposium on biomedical imaging (ISBI) (pp. 333–336). IEEE.

[2] Mitra, J. et al. (2014). Lesion segmentation from multimodal MRI using random forest following ischemic stroke. NeuroImage, 98, 324–335

[3] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18 (pp. 234-241). Springer International Publishing.

[4] Dong, H., Yang, G., Liu, F., Mo, Y., & Guo, Y. (2017). Automatic brain tumor detection and segmentation using U-Net based fully convolutional networks. In Annual conference on medical image understanding and analysis (pp. 506–517). Springer.

[5] Pereira, S., Pinto, A., Alves, V., & Silva, C. A. (2016). Brain tumor segmentation using convolutional neural networks in MRI images. IEEE Transactions on Medical Imaging, 35(5), 1240–1251.

[6] Rai, H. M., & Chatterjee, K. (2021). 2D MRI image analysis and brain tumor detection using deep learning CNN model LeU-Net. Multimedia Tools and Applications, 80, 36111-36141.

[7] Tiwari, P., Pant, B., Elarabawy, M. M., Abd-Elnaby, M., Mohd, N., Dhiman, G., & Sharma, S. (2022). Cnn based multiclass brain tumor detection using medical imaging. Computational Intelligence and Neuroscience, 2022.

[8] Siar, M., & Teshnehlab, M. (2019, October). Brain tumor detection using deep neural network and machine learning algorithm. In 2019 9th international conference on computer and knowledge engineering (ICCKE) (pp. 363-368). IEEE

[9] Ferdous, G. J., Sathi, K. A., Hossain, M. A., Hoque, M. M., & Dewan, M. A. A. (2023). LCDEiT: A Linear Complexity Data-Efficient Image Transformer for MRI Brain Tumor Classification. IEEE Access, 11, 20337-20350.

[10] Tabatabaei, S., Rezaee, K., & Zhu, M. (2023). Attention transformer mechanism and fusion-based deep learning architecture for MRI brain tumor classification system. Biomedical Signal Processing and Control, 86, 105119.

[11] Dishar, H. K., & Muhammed, L. A. (2023). Detection Brain Tumor Disease Using a Combination of Xception and NASNetMobile. International Journal of Advances in Soft Computing & Its Applications, 15(3).

[12] Aulia, S., & Rahmat, D. (2022). Brain Tumor Identification Based on VGG-16 Architecture and CLAHE Method. JOIV: International Journal on Informatics Visualization, 6(1), 96-102.

[13] Younis, A., Qiang, L., Nyatega, C. O., Adamu, M. J., & Kawuwa, H. B. (2022). Brain tumor analysis using deep learning and VGG-16 ensembling learning approaches. Applied Sciences, 12(14), 7282.

[14] Maqsood, S., Damaševičius, R., & Maskeliūnas, R. (2022). Multi-modal brain tumor detection using deep neural network and multiclass SVM. Medicina, 58(8), 1090.

[15] Ullah, N., Javed, A., Alhazmi, A., Hasnain, S. M., Tahir, A., & Ashraf, R. (2023). TumorDetNet: A unified deep learning model for brain tumor detection and classification. Plos one, 18(9), e0291200.

[16] Minarno, A. E., Kantomo, I. S., Sumadi, F. D. S., Nugroho, H. A., & Ibrahim, Z. (2022). Classification of Brain Tumors on MRI Images Using DenseNet and Support Vector Machine. JOIV: International Journal on Informatics Visualization, 6(2), 404-410.

[17] Papers with Code. (n.d.). Image Classification on ImageNet. Retrieved January 5, 2024, from https://paperswithcode.com/sota/image-classification-on-imagenet

[18] Maurício, J., Domingues, I., & Bernardino, J. (2023). Comparing Vision Transformers and Convolutional Neural Networks for Image Classification: A Literature Review. Applied Sciences, 13(9), 5521.

[19] Turing. (2021). The Ultimate Guide to Transformer Deep Learning. Retrieved from https://www.turing.com/kb/brief-introduction-to-transformers-and-their-power

[20] Dai, Y., Gao, Y., & Liu, F. (2021). Transmed: Transformers advance multi-modal medical image classification. Diagnostics, 11(8), 1384.

[21] Bhojanapalli, S., Chakrabarti, A., Glasner, D., Li, D., Unterthiner, T., & Veit, A. (2021). Understanding robustness of transformers for image classification. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 10231-10241).

[22] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., ... & Guo, B. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 10012-10022).

[23] Devi, K. J., Thongam, K., & Manipur, I. A Survey of Automatic Speaker Recognition System Using Artificial Neural Networks.

[24] shadyeldakrory. (2023). shadyfinal3 [Data set]. Kaggle. Retrieved from https://www.kaggle.com/datasets/shadyeldakrory/shadyfinal3