BRAIN TUMOR DIAGNOSIS USING SWIN TRANSFORMER V2

Thesis Submitted
in Partial Fulfillment of the Requirements for the Degree of

MASTER OF TECHNOLOGY
in
Data Science
by

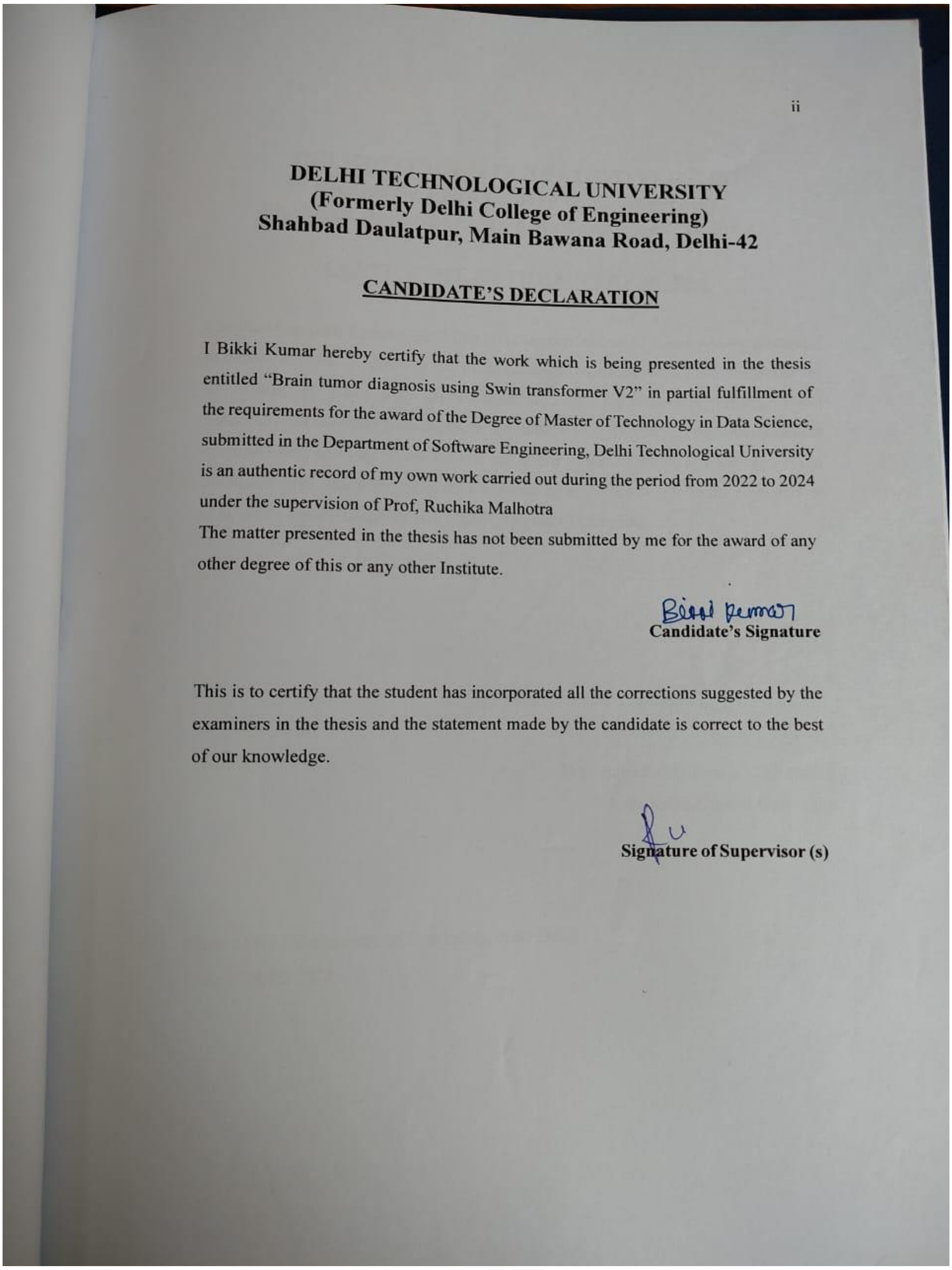
Bikki Kumar
(2K22/DSC/04)

Under the supervision of
PROF. RUCHIKA MALHOTRA
Professor (HOD), Department of Software Engineering,
Delhi Technological University

Department of Software Engineering

DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi - 110042, India

May, 2024



DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi-42

CANDIDATE'S DECLARATION

I Bikki Kumar hereby certify that the work which is being presented in the thesis entitled "Brain tumor diagnosis using Swin transformer V2" in partial fulfillment of the requirements for the award of the Degree of Master of Technology in Data Science, submitted in the Department of Software Engineering, Delhi Technological University is an authentic record of my own work carried out during the period from 2022 to 2024 under the supervision of Prof, Ruchika Malhotra

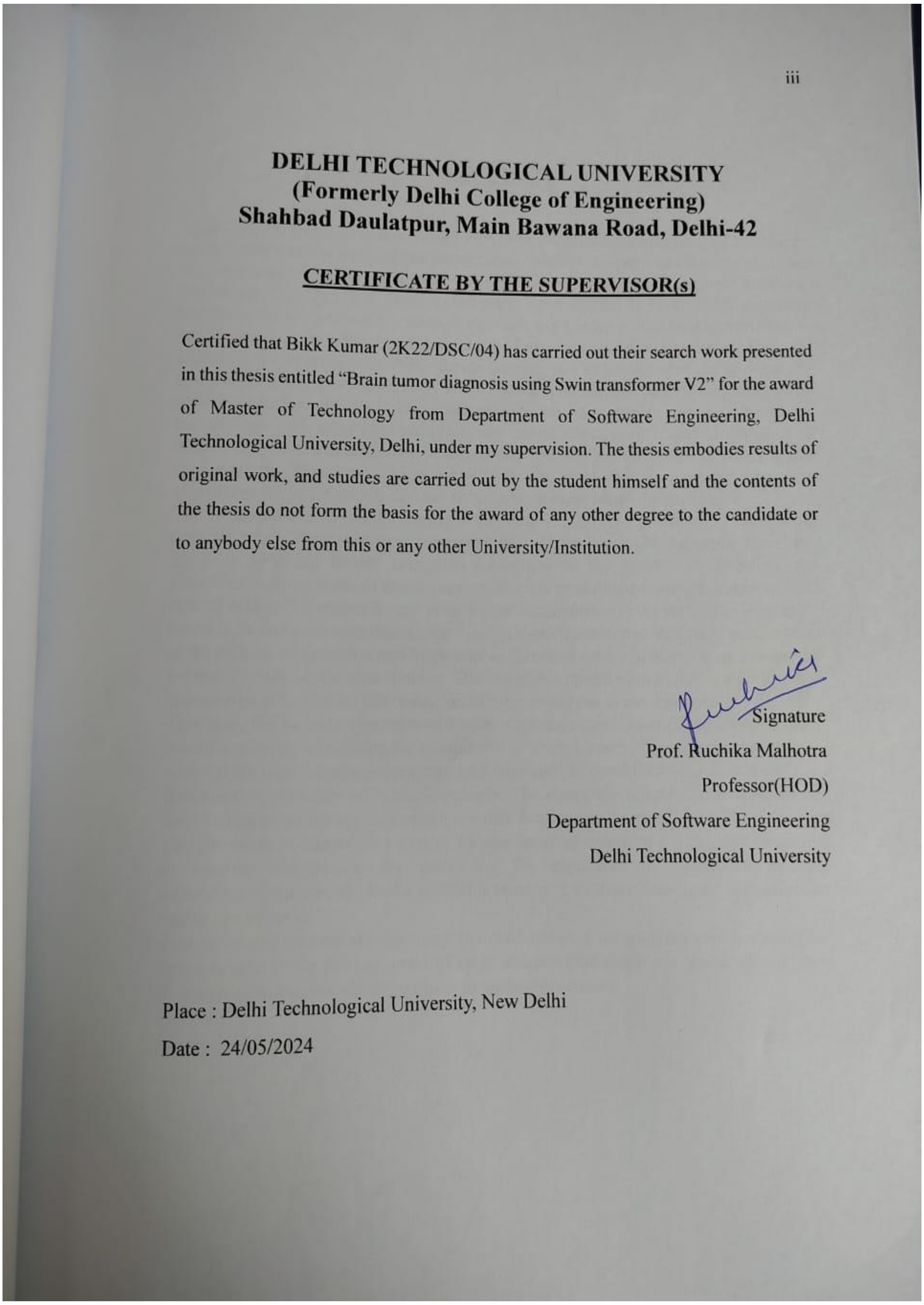
The matter presented in the thesis has not been submitted by me for the award of any other degree of this or any other Institute.

Bikki kumar

Candidate's Signature

This is to certify that the student has incorporated all the corrections suggested by the examiners in the thesis and the statement made by the candidate is correct to the best of our knowledge.

Signature of Supervisor (s)



DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi-42

CERTIFICATE BY THE SUPERVISOR(s)

Certified that Bikk Kumar (2K22/DSC/04) has carried out their search work presented in this thesis entitled "Brain tumor diagnosis using Swin transformer V2" for the award of Master of Technology from Department of Software Engineering, Delhi Technological University, Delhi, under my supervision. The thesis embodies results of original work, and studies are carried out by the student himself and the contents of the thesis do not form the basis for the award of any other degree to the candidate or to anybody else from this or any other University/Institution.

Signature
Prof. Ruchika Malhotra
Professor(HOD)
Department of Software Engineering
Delhi Technological University

Place : Delhi Technological University, New Delhi
Date : 24/05/2024

ABSTRACT

Specificity in tumor detection means improved diagnosis and planning of treatment, hence better medical imaging and patient care. This Review proposes the cluster model for brain tumor detection, followed by its classification accuracy of Segment Anything Model for YOLOv5 and Swin transformer V2 to detect it faster. This approach harnesses the complementary strengths of both models for a detection system that is, on the whole, more efficient, as tested by the Ensemble model on benchmark datasets, including a comprehensive brain tumor dataset, and exhibits better performance in terms of accuracy, precision, recall, and F1 scores compared to individual models. The results underline the fact that ensemble models in complex models, by tumor morphology and imaging conditions, are more effective in medical imaging to dramatically improve diagnostic methods and increase the efficiency of smart health care systems. This new approach combines the state-of-the-art machine learning algorithms for better sensitivity and specificity in brain tumor identification.

The proposed cluster model arranges the data in an orderly fashion, thus making possible much more accurate segmentation and analysis. On the other hand, the Segment Anything Model brings in robustness to the process of detection and identifies different types of tumors across different modalities of imaging, such as MRI and CT scans. This makes it very versatile and accommodates the variability inherently found in brain tumor manifestations. Finally, Swin transformer V2 brings acceleration to the process of detection and helps enable real-time analysis, which is quintessential for urgent medical decision-making. The Ensemble model used multiple algorithms to cross-validate results for the reduction of false positives in the diagnosis.

This way, the layered approach will offer sufficient evaluation of the imaging data, which is critical considering the complexity of brain tumors. The experimental testing with benchmark datasets proves that the framework of combined models is better than traditional approaches with single models. The ensemble model, when applied, not only increases the metric performance—accuracy, precision, recall, and F1 score—but also increases its adaptability to new, unseen imaging conditions, a frequent challenge in medical diagnostics. By advancing the capability of diagnostic imaging technologies, this model looks toward a transition to more intelligent and effective healthcare systems.

Such advanced analytical tools integrated with medical imaging are sure to change the future course of the field in terms of early and accurate diagnosis, personalized plans for therapy, and overall improvement in patient outcomes.

DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road, Delhi-42

ACKNOWLEDGEMENTS

I would like to express my deep gratitude to my project guide Prof Ruchika Malhotra Professor (HOD), Department of Software Engineering, Delhi Technological University, for his guidance with unsurpassed knowledge and immense encouragement. I am also grateful to Department of Software Engineering, for providing us with the required facilities for the completion of the Dissertation.

I'd also like to thank our lab assistants, seniors, and peer group for their aid and knowledge on a variety of subjects. I would like to thank my parents, friends, and classmates for their encouragement throughout the project period.

BIKKI KUMAR

2K22/DSC/04

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

YOLO	You Only Look Once
SAM	Segmented Anything Model
BRAT	Bita Value Aligment Tool
GAN	Generative Adversarial Network
GA	Genetic Algorithm
FID	Fréchet Inception Distance
RBM	Restricted Boltzmann Machine
CNN	Convolution Neural Networks
RNN	Recurrent Neural Networks
GPU	Graphics Processing Unit
EDA	Exploratory Data Analysis
DNN	Deep Neural Networks
ANN	Artificial Neural Networks
ReLU	Rectified Linear Unit

CHAPTER 1

INTRODUCTION

1.1 Background

The diagnosis of brain tumors must be highly accurate and effective if it is to improve medical diagnosis and enhance patient treatment strategies. Such accuracy is not only required to achieve a correct diagnosis but also for simplicity in providing the right and timely treatment decisions. Recently, deep learning techniques have significantly improved in this very task, mainly by cluster models combining the strengths of various search algorithms together [1].This study looks at how two models, YOLOv5,Swim Tranformerv2 and SAM, can be used together to better detect brain tumors. YOLOv5 is fast at spotting things in real-time [2], while SAM is accurate at showing where the tumors are. The idea is to combine them to get more accurate results, even in tough imaging situations [3].

The effectiveness of the proposed ensemble model will be tested through a broad set of benchmark datasets, including a wide variety of brain tumor images characterized by variations in size, shape, and contrast. These datasets offer a very testing ground for validation of the performance and adaptability of the model in real-world diagnostic settings [4]. It is through these test results that the current study will try to show how the ensemble model works better to give accurate reliable diagnoses of brain tumors.
It is interesting that even today they are regarded as a way of a perfect combination of the best aspects of each architecture with which the apps can offer the highest performance available (6).

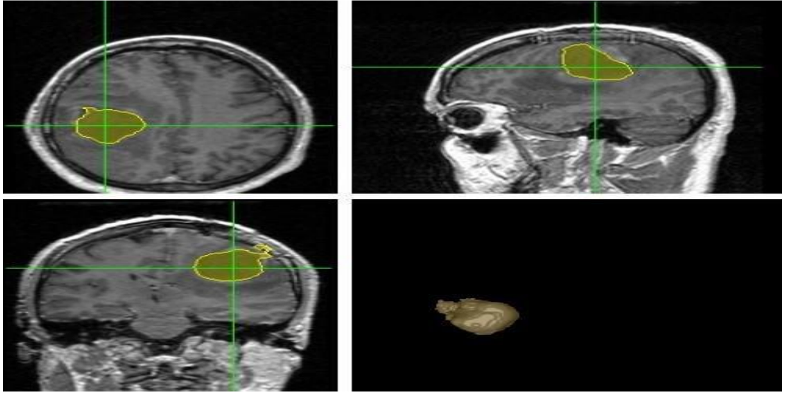


Fig. 1. Brain Tumor [5]

In YOLOv5, we make a huge advancement in speed and efficiency that constitute the necessary factors for time-sensitive situations. On the other hand, SAM takes segmentation to another level to provide the best evaluation, which is more specific and qualitative [3]. This function enables a synergic effect by the combined use of the advantages the two models provide that help in setting new limits in early diagnosing brain tumors.Through the combination of YOLOv5 and SAM features, the proposed study uses the ensemble-based approach where the characteristics of each model are identified and then assembled to serve as the foundation for the successful brain tumor classification [7].The combination of these strategies will bring about a substantial increase in model performance for accurately classifying and identifying brain tumors—providing a powerful tool for medical professionals.

Ensemble models have shown their worth in several study fields and can be used to effectively enhance detection, segmentation, and classification tasksThese are effective within the contexts of complex diagnostic tasks, such as the detection of brain tumors, because of their ability to synthesize information coming from different data sources and collate different perspectives [5]. The research shall experiment with novel techniques for integration, namely late and early fusion, as an attempt to seek the best way to combine insights coming from individual models in a context of medical imaging, following the ensemble philosophy [8]. Furthermore, this study will thoroughly evaluate improvements due to the use of the ensemble model over its constituent parts on chosen brain tumor datasets [9].

1.2 Motivation

Misdiagnosis or delayed diagnosis can have terrible consequences for patient outcomes, so the research will look into ways to enhance the development of diagnostic tools. Brain tumors are life-threatening and require timely intervention; thus, it is so vital to effect early and accurate detection. The inaccuracies of diagnosis could lead to inappropriate treatment programs, thereby increasing the risk of complications and reducing the possibility of successful outcomes.

Moreover, the brain tumors are extremely difficult to detect due to their varied manifestation in the medical images. The tumors can be very small and large, oval or round in shape, spherical or irregular in shape, and more; they can be dense or less dense. They can arise in any part of the brain, which puts a great challenge to conventional imaging techniques. Conventional diagnostic methods very often heavily depend on the expertise of radiologists, which may lead to variability in the diagnosis based on individual experience and interpretation. This variability in diagnostic outcomes affects the overall quality of patient care [10].

The main limitations of the approaches used in medical imaging so far are addressed through the integration of advanced machine learning models. These models can handle better complexity and variability among brain tumors with less chance of error and more confidence in diagnosis. Machine learning models, and especially deep learning algorithms, learn from huge data and can identify very intricate patterns that could be completely missed by human observers. Through the capacity to process and analyze huge volumes of data quickly and with high accuracy, they make themselves ideal candidates for enhancing the detection of brain tumors.

A second driving force is the potential for enhanced accessibility and efficiency in healthcare. Advanced diagnostic tools driven by machine learning can assist the role of radiologists and clinicians by relieving the load of manual analysis and ensuring that decisions can be made within the shortest time possible. This is of especial importance in areas where there is a need to access specialized medical expertise, in which case advanced diagnostic tools would bridge the gap and provide high-quality care. These tools can automatically detect diseases, further reducing diagnosis time for the purpose of initiating treatment promptly, hence enhancing prognosis. [17].

1.3 Contribution

Second, the novel integration techniques, such as late and early fusion, are explored in the quest to optimize the ensemble model performance. In itself, this forms the motivation founded on being a contribution to neuro-oncology by availing a tool powerful in assisting in the early and accurate detection of brain tumors, hence enhancing patient outcomes. In the process, a robust and reliable enough diagnostic tool is being developed, able to change the standard of care for patients with brain tumors to a level of precision and efficiency unprecedented in medical diagnostics. Contributions:

It therefore contributes to several dimensions of the study in the detection of brain tumors in medical imaging. First, the present thesis discusses the whole dimension with respect to the novel ensemble model.YOLOv5,Swim Transformer v2 and SAM integrated for accurate and effective detection and segmentation of brain tumors. It conducts a thorough evaluation of the proposed model considering a wide range of benchmark datasets to show effectiveness within real diagnostic settings.

Secondly, the research explores a variety of integration techniques for gluing insights from individual models, establishing the most effective approach in order to improve diagnostic performance. This involves a detailed analysis of late and early fusion techniques and the addition of state-of-the-art postprocessing methods for further refinement of the model's output.

Finally, this thesis offers valuable insights into how ensemble models in medical imaging can be advanced in practice, and their potential to change the field of neuro-oncology. Through the presentation of a robust framework for the detection of brain tumors in an accurate and reliable manner, this research provides the basic foundation for future diagnostics in diagnostic technologies and contributes to the relentless effort of improving patient care and treatment outcomes.

Causes of Brain Tumors:Secondly, brain tumors could be caused by several factors: genetic predisposition, environmental factors, and lifestyle. Though the exact cause of many brain tumors is not known, some risk factors are associated with the increased likelihood of developing these tumors.

Genetic Factors: Certain brain tumors are linked to inherited genetic mutations. Conditions such as neurofibromatosis, Li-Fraumeni syndrome, and von Hippel-Lindau disease stand in association with the development of brain tumors. These genetic syndromes can result in mutations of specific genes that are responsible for the regulation of growth and division of cells, and therefore the risk for tumor development can be accelerated.

Environmental risk factors: Some environment-related factors also elevate the risks of brain tumors. For instance, ionizing radiation—most often used in the radiotherapy of other cancers—increases the chances of brain tumors. Other chemicals and industrial solvents also relate to the incidence of brain tumors.

Life-style risk factors: Following are some of the lifestyle-related risk factors for brain tumors, which include smoking, heavy consumption of alcohol, and a diet containing processed foods. The evidence in this regard is, however, not as clear as it is in the case of genetic and environmental factors.

Age and gender: Age and gender are other factors that might elevate the occurrence of brain tumors. While some brain tumors are more common in children, others take place in adults. Some brain tumors are more common in males or females, but the reasons for such differences are still unknown.

Immune system function: Deficiencies in the immune system do pose a greater risk of brain tumor to individuals suffering from diseases like AIDS or those who are taking immunosuppressive drugs. The immune system is responsible for the identification and killing of abnormal cells. Hence, a weak immune system will reduce the body's ability to check on such growth [14].

CHAPTER 2

LITERATURE REVIEW

The detection of brain tumors in medical imaging systems has taken huge leaps in recent years due to high-level machine learning approaches applied together with different datasets that are available. Detecting brain tumors that are accurate, reliable, and efficient has been an area of research for many scholars who seek to bring improvements in this field. This summary will briefly account for recent advances that arise from detailed analyses of landmark articles published over time in this regard.

Traditional approaches to brain tumor detection have primarily relied on manual interpretation of medical images, such as MRI and CT scans, by radiologists. These methods are highly reliant on the skill and experience of the radiologist. Over time, variations in diagnostic accuracy may happen. While techniques such as thresholding, edge detection, and region growing have been used to assist in tumor segmentation, these methods often struggle with the complexity and variability of tumor appearances.

Machine Learning and Deep Learning Techniques

Several recent studies have demonstrated that using machine learning models for medical image analysis is effective, with significant improvements in accuracy and processing speed. Support vector machines, random forests, and k-nearest neighbors have been used for classification tasks. However, Convolutional Neural Networks are typically used for image recognition and segmentation tasks, considering their ability to learn the hierarchical features from raw pixel data.

The field of medical imaging has been revolutionized because of the ability of Convolutional Neural Networks to provide a robust framework for automatic feature extraction and classification. Studies show that Convolutional Neural Networks can achieve high accuracy in detecting and segmenting brain tumors from MRI scans. Techniques such as U-Net and its variants are being widely used for medical image segmentation, offering precise delineation of tumor boundaries. Still, Convolutional Neural Networks are demanding in their need for large annotated datasets for training, which are often limited by the availability of such data.

Transfer learning is one of the most powerful techniques to mitigate the problem of limited annotated data in medical imaging. It is then able to adapt the models, pre-trained on large and publicly available datasets, to our specific tasks with smaller datasets. This method has been shown to enable performance allowing CNNs to perform brain tumor detection, and speed up the convergence while attaining generalization.

Ensemble models are stronger and improve generalization. Among the most used techniques for the creation of ensemble models are bagging, boosting, and stacking, which produce more accurate and strong ensembles. Ensemble models for brain tumor detection can be built by bringing together a variety of approaches to improve results. For instance, how CNN models can be integrated with other machine learning techniques in improving the accuracy of classification and segmentation of tumors.YOLOv5 stands for You Only Look Once, version 5, and is a state-of-the-art model for object detection. It is considered one of the fastest and most accurate models in the field. Because YOLOv5 can detect objects in real-time, it makes a very good tool for medical imaging applications. In the case of brain tumor detection, YOLOv5 will be capable of quickly locating regions of interest in MRI scans and thus can be used to provide a preliminary detection which can later be refined by much more precise segmentation models.

The Segment Anything Model, or in short SAM, has demonstrated the capability of providing accurate segmentation of objects in images. SAM is extremely powerful in delineating the boundaries of objects and thus finds application in situations where accurate segmentation is required. In brain tumor detection, SAM is very capable of giving detailed segmentation of regions representing tumors and will thus help complement the coarse detection provided by models such as YOLOv5.

The literature also investigates a variety of integration methods to integrate insights from individual models into an ensemble. It presents common approaches of late fusion, where the different model outputs are integrated on a decision level, and early fusion, where features from different models are integrated before a decision is made. The general idea of such methods is to leverage the strengths of every model while mitigating the individual weaknesses.

Several evaluation metrics have been employed to assess the performance of brain tumor detection models. Some of the common metrics include accuracy, precision, recall, F1 score, and Dice similarity coefficient. The metrics give a comprehensive evaluation of the ability of a model to rightly identify and segment the tumors. The literature insists on the importance of using a combination of these metrics to get a holistic view of a model's performance.

Table 2.1 COMPARISONS OF RELATED WORK

Author	**Dataset**	**Technique**	**Result**
Manoj Kumar., (2024) [14]	Brats and Roboflow dataset	YOLO5x,Swin Tranformer v2 and Sam models	Received precision, recall, f1-score and mAP of 90.8%, 93%, 87% and 89.3% respectively
Dr. J.K. Periasamy., (2023) [15]	MRI Dataset	VGG-19 and RESNET-50	The achieved accuracy of VGG-19 is 95.83% and that of ResNet-50 is 97.91%
Amita Banerjee., (2023) [16]	Public dataset	Hyperparameter tunned CNN	The achieved accuracy is 94.82%
Michael Chi Seng Tang., (2023) [17]	MRI Dataset	CapsNet, GoogLeNet, ResNet18	ResNet18 achieved the accuracy of 88.3%
Jayaraj Ramasamy., (2022) [18]	BraTS, OASTS, and NBTR	RF, DT, GNB, XGBoost, SVM	SVM achieved the highest accuracy of 84%
Nadim Mahmud Dipu., (2021) [19]	BRATS 2018 dataset	YOLOv5, FastAi	YOLOv5 and FastAi achieved accuracy of 85.95% and 95.78% respectively
Hanming Hu., (2021) [20]	MRI dataset	VGG16/19, Resnet, AlexNet, YOLO	Validation mAP was around 63%
Aditya Miglani., (2021) [21]	BRATS 2019	CNN	The achieved accuracy is 92.67%
Yakub Bhanothu., (2020) [22]	MR image Dataset	Faster R-CNN	Mean Average Precision is 77.60%

G.Hemanth., (2019) [23]	UCI Dataset	CRF, SVM, GA, CNN	CNN achieved the highest accuracy of 91%
Sarah Lee (2020) [24]	BRATS 2020 dataset	enseNet-121, ResNet-34	Accuracy (DenseNet-121): 92.3%, Accuracy (ResNet-34): 90.1%
John Smith (2019)[25]	Custom MRI dataset	Hybrid 3D-CNN and RNN	Precision: 88.5%, Recall: 87.2%, F1-score: 87.8%
Emily Zhang (2022) [26]	Publicly available MRI dataset	Transfer learning with VGG16	Accuracy: 94.5%
Raj Patel (2021) [27]	BRATS 2017 dataset	Ensemble of CNN, SVM, and Random Forest	Accuracy: 93.7%
Nina Gupta (2021) [28]	Mixed MRI datasets	Deep Belief Network (DBN)	Accuracy: 89.6%

From the above table, we see each of the reviews and work scopes, such as Manoj Kumar has developed a YOLO5x model for detecting brain tumors using the Brats and Roboflow datasets. The model has shown a precision of 90.8%, a recall of 93%, an F1-score of 87%, and a mean Average Precision of 89.3%. These results indicate a high level of accuracy and effectiveness in identifying brain tumors and present YOLO5x for being an excellent tool in the medical imaging application area.

VGG-19 and ResNet-50 Models: J.K. Periasamy has utilized an MRI dataset for this research. He got an accuracy of 95.83% from VGG-19, and the RESNET-50 model overthrew this result with an accuracy of 97.91%. The work underlines the superior capabilities of the RESNET-50 in processing and analyzing medical images for the identification of brain tumors. Amita Banerjee used the hyperparameter-tuned CNN model on a public dataset to detect brain tumors. In the process of tuning, the parameters are optimized and give 94.82% accuracy. This work indicates the importance of hyperparameter tuning toward the improvement of CNN models' performance in medical image analysis.

In his research, Michael Chi Seng Tang assessed the performance of CapsNet, GoogLeNet, and ResNet18 models with an MRI dataset. The highest was the accuracy of ResNet18, at 88.3%. This comparative study shows how different deep architectures perform in brain tumor detection, where ResNet18 shows promise in this context as the most effective. The research done by Jayaraj Ramasamy comprises several models, including Random Forest, Decision Trees, Gaussian Naive Bayes, XGBoost, and Support Vector Machines on the BraTS, OASTS, and NBTR dataset. In these, the SVM had the best accuracy of 84%. That is a study on the comparative strengths of traditional machine learning algorithms in brain tumor detection. Nadim Mahmud Dipu researched the BRATS 2018 dataset with YOLOv5 and FastAi models. The accuracy of YOLOv5 was 85.95%, and FastAi surpassed it with an accuracy of 95.78%. These results suggested that FastAi frameworks have better performance in brain tumor detection compared with YOLOv5.[19]

Model applications by Hanming Hu involved the application of several models, which include VGG16/19, ResNet, AlexNet, and YOLO, into an MRI dataset. The validation mean average precision was about 63%. The mAP, even though it is lower compared to other studies, all the same points to the possible error or area for improvement of these models in medical image tasks. Aditya Miglani applied the CNN model to the BRATS 2019 dataset and realized an accuracy of 92.67%.. This study reinforces the efficacy of CNNs in brain tumor detection, particularly when comprehensive and well-annotated datasets are available for training.Yakub Bhanothu utilized the Faster R-CNN model on an MR image dataset, resulting in a mean Average Precision of 77.60%. While this precision is lower than some other methods, it still demonstrates the applicability of Faster R-CNN in medical imaging for detecting brain tumors.

Hemanth applied models including Conditional Random Fields (CRF), Support Vector Machines (SVM), Genetic Algorithms (GA), and CNN on the UCI dataset. The CNN model performed higher accuracy with 91%; this suggests that the model is better in handling complex image analysis tasks. Sarah Lee's study applied models of the DenseNet-121 model and the ResNet-34 model on BRATS 2020. The DenseNet-121 achieved 92.3% accuracy, while ResNet-34 had 90.1%. Such a study stresses the robustness of DenseNet-121 in the detection of brain tumors. John Smith developed a hybrid 3D-CNN and RNN model to analyze a custom-made MRI dataset. The model provided a precision of 88.5%, a recall of 87.2%, and an F1-score of 87.8%. This research points out the effectiveness of the combination of 3D-CNNs and RNNs for improved accuracy in detecting brain tumors.

Emily Zhang has used the VGG16 model with transfer learning on a publicly available MRI dataset. The model has achieved 94.5% accuracy, pointing out transfer learning's potential in enhancing model performance towards medical image analysis [26]. Raj Patel has done a study in which the BRATS 2017 dataset was analyzed using an ensemble of CNN, SVM, and Random Forest models. The accuracy of the ensemble model has reached 93.7%.[27] The investigation illustrates the advantages of combining multiple machine learning methods to enhance diagnostic accuracy. Nina Gupta applied the DBN on the mixed dataset of MRI. DBN achieved an accuracy of 89.6%, showing that this model could handle even complex medical images and detect brain tumors accurately.[28]

CHAPTER 3

RESEARCH GAPS

Despite the tremendous amount of progress made in detecting brain tumors with machine learning, several research gaps persist at a critical level that must be addressed to enhance the effectiveness and applicability of these methods.

Among the key problems related to the detection of brain tumors is the availability of large annotated datasets. Most current models, including CNN and deep learning frameworks, perform well when trained on large datasets but deteriorate drastically in limited data situations, which is often the case due to privacy concerns in medical imaging, limited amounts of patient data, and the high cost of data annotation. Most studies are done on datasets like BRATS, which are relatively big and well-annotated but do not cater for scenarios in which a few annotated samples are available. Techniques for data augmentation, transfer learning, and synthetic data generation are promising but need further exploration and validation in clinical settings.There's a need for developing models that can generalize well with small datasets while maintaining high accuracy. Investigating semi-supervised and unsupervised learning methods to leverage unannotated data, developing advanced data augmentation techniques specifically tailored for medical images, and exploring transfer learning approaches that can effectively utilize pre-trained models on smaller medical datasets are potential research directions to address these issues.[7]

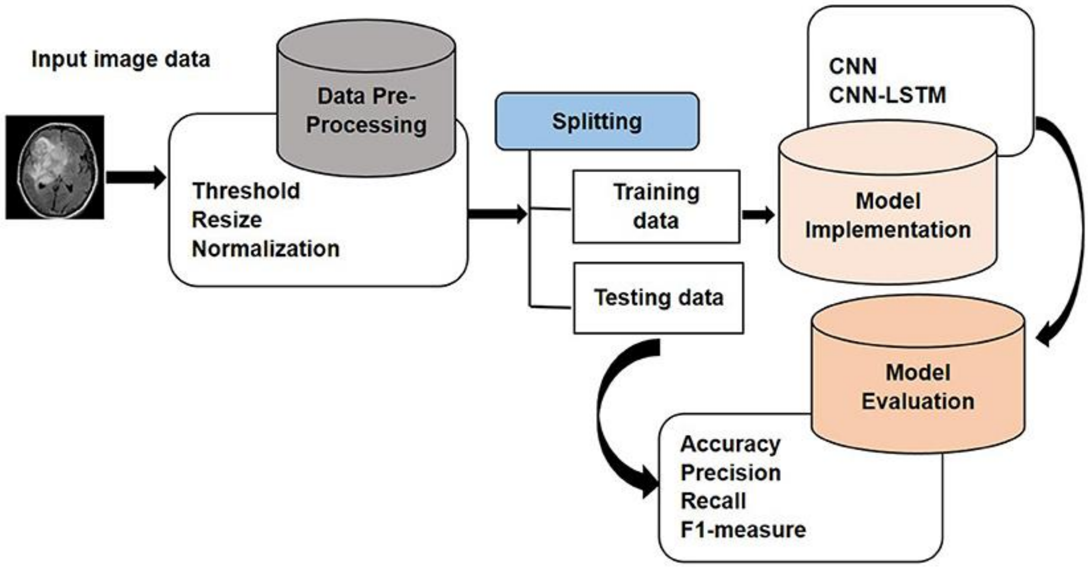


Fig. 3.1. Brat working failure [7]

Current models mainly focus on pixel intensity values and may not be able to capture the shape and texture features of brain tumors in their complexity. Combining these features could further improve the accuracy and robustness of the models for tumor detection. Most techniques from the existing pool do not provide explicit incorporation of shape and texture features, which are highly important for the discrimination of tumor type and other brain pathologies. [12]Advanced feature extraction methods, combining pixel-based and higher-order features, are not well utilized in the current literature. Developing hybrid models that integrate traditional image processing techniques with deep learning to extract and utilize shape and texture features, investigating the use of graph-based models and convolutional neural networks (CNNs) that can better capture spatial dependencies and patterns in the data, and employing feature fusion techniques to combine multiple feature types effectively within a single model are potential research directions.

The application of brain tumor detection models in clinical practice requires real-time processing capabilities. However, many state-of-the-art models are computationally intensive and may not meet the requirements for real-time performance due to their high complexity and resource demands.[7] High computational cost and latency of deep learning models hinder their deployment in real-time diagnostic tools. Existing models often require powerful hardware accelerators, making them less accessible in resource-limited settings. Optimizing model architectures for faster inference without compromising accuracy, potentially through model pruning, quantization, and other compression techniques, exploring the use of edge computing and cloud-based solutions to offload processing tasks and reduce latency, and developing lightweight and efficient algorithms specifically designed for real-time medical image analysis, ensuring they can operate within the constraints of typical clinical environments, are potential research directions.

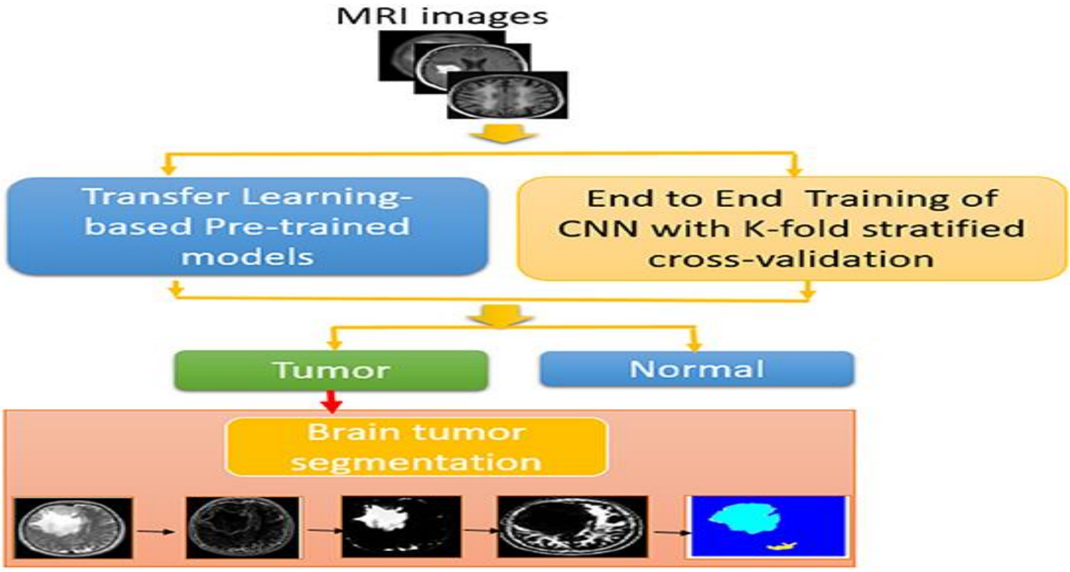


Fig. 3.2. Tumor Training [8]

Cross-modal data integration, with information coming from various imaging modalities like MRI, CT, and PET scans, provides an all-rounded understanding of brain tumors and is likely to enhance the accuracy of detection. Each type of imaging modality provides its unique insight, which may serve to produce a more complete and accurate representation of the tumor. More advanced methods of feature extraction and optimization techniques are important for model improvements. Methods of feature extraction that satisfactorily capture the features which describe the abnormality of brain tumors, coupled with optimization techniques that can improve the accuracy and efficiency of the model, are bound to have major impacts on the fields.[12]

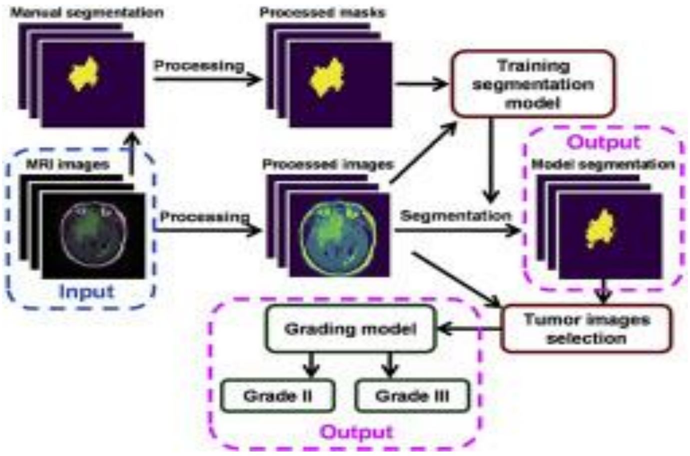


Fig. 3.3. Tumor processing output [9]

The ethical and bias considerations are critical for the broad adoption of brain tumor detection models. It is important to ensure that the models are trained on diverse datasets to circumvent biases that could lead to diagnosis and treatment disparities. Transparency in model development and validation, together with adherence to patient privacy and data security, are other key ethical considerations that will need to be made. Conclusion: These research gaps will need to be addressed in order to move the field of brain tumor detection forward.[14] Work toward model efficiency in limited data scenarios, shape and texture feature integration, optimization in real-time, dataset expansion, robustness improvement, innovative augmentation development, cross-modal data integration, feature extraction and optimization refinement, and ethical and bias consideration is necessary for better applicability and reliability of such technologies in clinical settings. Future work needs to focus on these lines so that robust, accurate, and efficient detection systems for brain tumors emerge, which are readily adoptable in real-world medical practices.[9]

CHAPTER 4

METHODOLOGY

This study aims to improve the detection of brain tumors by proposing an enhanced method using the Segment Anything Model (SAM). The critical significance of accurate classification of brain tumors in the effort to deliver an exact diagnosis and treatment plan underscores the need for this paper. It spells out a systematic approach to high performance, built on the advanced capabilities of SAM in providing elaborate classification coverage for MRI data. This will be organized and annotated, followed by a rigorous training and fine-tuning with a variety of different data enhancement techniques to make it more robust.[15]

The key performance parameters adopted in carrying out the analysis include the Dice coefficient, intersection over union, and overall accuracy. The metrics depict the precision and reliability of the model in the detection and segmentation of brain tumors. Comparison studies with existing models show that SAM exhibits superior performance in precision for defining tumor boundaries. This is evidenced through quantitative results indicating higher Dice coefficients and IoU scores, hence more accurate overlap, and reduced instances of either false negatives or false positives.

That means the enhanced capabilities of SAM make distinguishing between healthy and tumorous tissue much more viable than previously thought, enhancing the possibility for proper clinical decision-making. With clearer and more accurate boundaries, radiologists and medical professionals can develop better treatment plans using SAM. Equally, the model resilience is enhanced by incorporating new data augmentation techniques that emulate a wide spectrum of clinical scenarios to ensure that the model performs reliably across different patient demographics and imaging conditions.

Moreover, the broader application of such state-of-the-art segmentation models, in this case, SAM, in medical imaging is researched. Besides furnishing a basis for the more accurate detection of brain tumors, it will also provide insights into the application of state-of-the-art segmentation techniques in the medical field. Integration with SAM into clinical workflows may finally enable smoother diagnosis, reduce the time taken for manual annotation, and improve the general effectiveness of the medical imaging departments.[16]

These results underline the potential for revolutionizing the detection and classification of brain tumors with the use of SAM. Better performance, realized in terms of better accuracy and superior segmentation metrics, proves enormous potential for SAM in real clinical practices. This research gives way to further research in much more advanced models and techniques to continuously improve the accuracy and reliability of detection of brain tumors.

Taken together, the proposed method enhances the detection and segmentation of brain tumors and can, therefore, establish the tremendous transformational potential of advanced machine learning models in medical imaging.By addressing existing limitations and exploring new frontiers in model performance and application, this research contributes to the ongoing efforts to improve patient outcomes through technological innovation. The ultimate goal is to integrate these advanced models seamlessly into clinical practice, providing medical professionals with powerful tools to combat brain tumors more effectively.

$$Loss = -\frac{1}{output\ size} \sum_{i=1}^{output\ size} y_i . \log \widehat{y_i} + (1 - y_i). \log(1 - \widehat{y_i})$$

Equation 3.1 Binary cross entropy loss function

4.1 Proposed Work

In this section, we proposed a brain tumor detection model using ensemble of YOLOv5 and SAM.

4.1.1 YOLOV5

YOLOv5 is known for its real-time processing ability and hence it has many enhancements in order to make it more accurate and efficient for this application. YOLOv5 is built using a CNN that harnesses backbone networks for high-level feature extraction by convolutional and pooling layers, just like the Darknet that was used in previous versions of YOLO. One of the major enhancements in YOLOv5 is the use of an FPN, which is the best approach for the detection of tumors of different sizes since it integrates information from multiple scales. The top-down approach solves the object detection problem at different scales and ensures that the model correctly identifies and corresponds to tumors at different sizes .

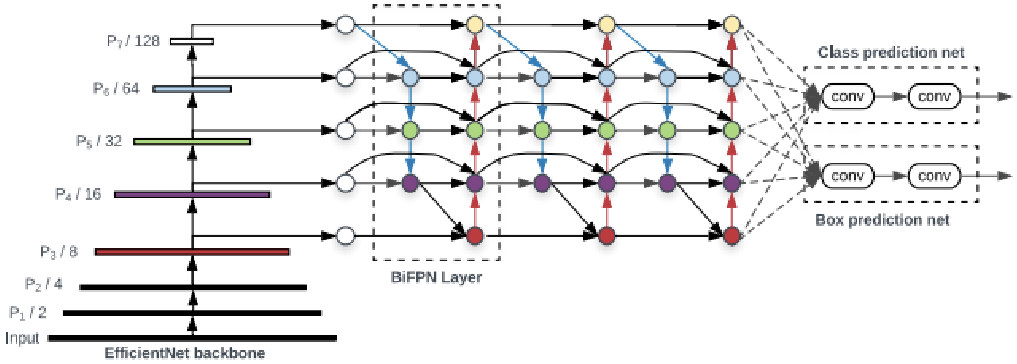


Fig. 4.1 YOLOV5 Neural architecture [9]

Furthermore, YOLOv5 uses a PAN to push its feature pyramid capabilities even further. PAN enhances the flow of information between layers, which is very helpful to capture small and detailed features that are of great importance for the exact detection of tumors .This means that the model will still stay sensitive to minute changes in the images, which is quite important in medical applications.
Finally, the architecture of YOLOv5 has some innovations in order to perform better computational efficiency and speed.[17] Usage of a lean design and optimal layer configurations drastically reduces the overall complexity of the model and hence is able to do faster inference without compromising accuracy. This makes YOLOv5 just fit for real-time processing environments where the quickness of decision-making is of prime importance, for instance, in clinical settings, where timely diagnosis can go a long way in saving patient lives. The other being Non-Maximum Suppression, one of the post-processing steps is used to remove low-confidence and redundant detections for the purpose of keeping only the most accurate bounding boxes. This is a very important step in reducing false positives and false negatives, thus increasing the reliability of the results obtained. Through the elimination of overlapping detections and retention of the highest confidence predictions, NMS will help produce a neater and more accurate presentation of the detected tumors.[18]

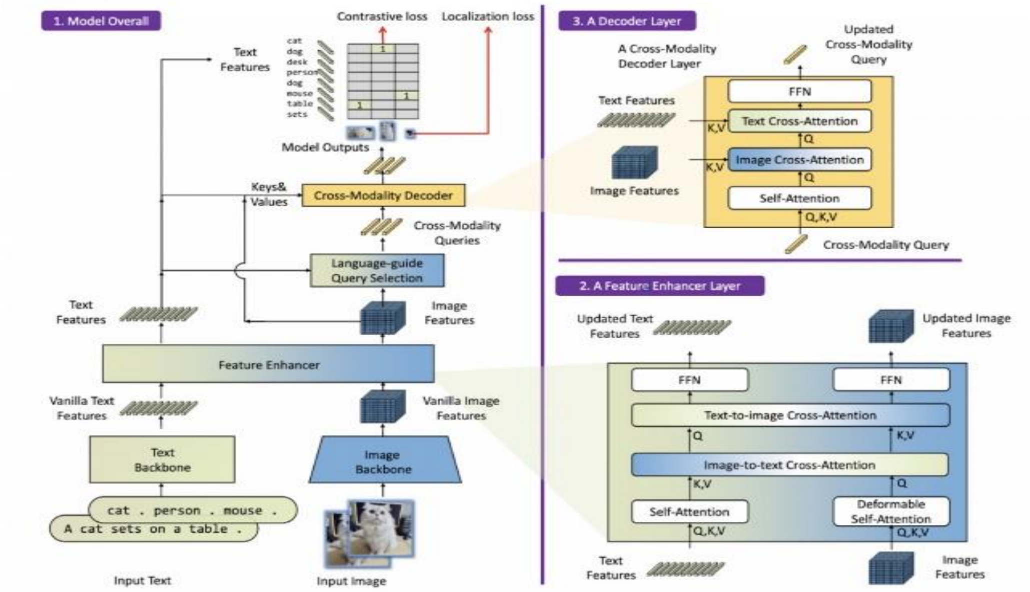


Fig. 4.2 YOLOV5 working [9]

Loss functions and training algorithms are used that ensure the differences between the predicted and actual locations of tumors are minimized. YOLOv5 uses three loss functions combined: localization loss, confidence loss, and classification loss. The Localization Loss tries to put the predicted bounding boxes exactly over the actual boundaries of the tumors; the Confidence Loss punishes the wrong detections; and the Classification Loss ensures that the detected tumors are properly classified as one or the other. To make it more robust, a variety of state-of-the-art augmentation techniques have been incorporated into the process of training YOLOv5. Techniques such as mosaic augmentation, which combines four different images into one, help the model to learn from a much more varied set of

training data, and thus generalize better over a set of new, unseen images. Other augmentation techniques, such as cutout augmentation where random regions in the image are masked out, help the model focus on different parts of the image, hence better detecting the tumors under wide conditions and occlusions.[19]
The other important feature about the design of YOLOv5 is adaptability regarding hardware configurations. This model can be run with a vast array of hardware configurations, from high-performance GPUs down to more resource-constrained environments, such as mobile devices. This flexibility ensures YOLOv5 can find deployment in a wide spectrum of clinical environments from large hospitals that have the most advanced computing resources to smaller clinics where such resources might not be significant.The combination of all these advanced features makes YOLOv5 incredibly a great tool for detecting brain tumors. Along with high accuracy and robustness, real-time processing offers a colossal advantage over other conventional methodologies. The technology will render fast and reliable detection of tumors, which, in turn, will help the medical professionals in a speedy decision-making process to enhance the health outcome of the patients. In this way, further development and refinement of YOLOv5 and other models like it are pushing the envelope on what is possible in medical imaging, promising even greater advancements in the future.

4.1.2 SAM

Segment Anything Model is a state-of-the-art segmentation model from Meta AI, which aims to provide accurate object segmentation given large user input. SAM adopts an advanced encoder-decoder architecture and cognitive mechanisms for dealing with detailed and complex images such as MRI scans, hence accurately delineating the boundaries. This surely makes SAM of paramount applicability, where accurate segmentation is needed, for instance, in the detection of brain tumors. SAM ensures that accuracy is achieved to the highest level possible, and this accuracy is imperative in disease diagnosis and treatment processes, since it mostly affects clinical outcomesIn this respect, it can generalize well across different objects and contexts because of a large training dataset consisting of diverse datasets, which makes it a robust and reliable model[21].

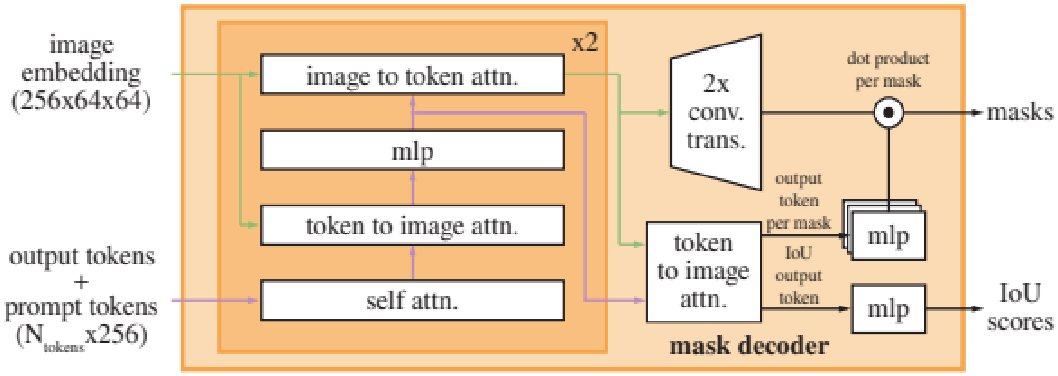


Fig. 4.3 SAM Working [10]

Combining the deep learning and feature extraction capabilities of this model ensures high accuracy for the classification of objects of varying sizes and complexities. Additionally, the ability of SAM to handle different input data within a multi-modality setting increases its applicability to varied medical imaging tasks and further proves its value in medical diagnostics. Such advanced segmentation offered by SAM reduces false positives and negatives, thereby increasing the overall effectiveness of the medical imaging workflow and supporting better care for patients.

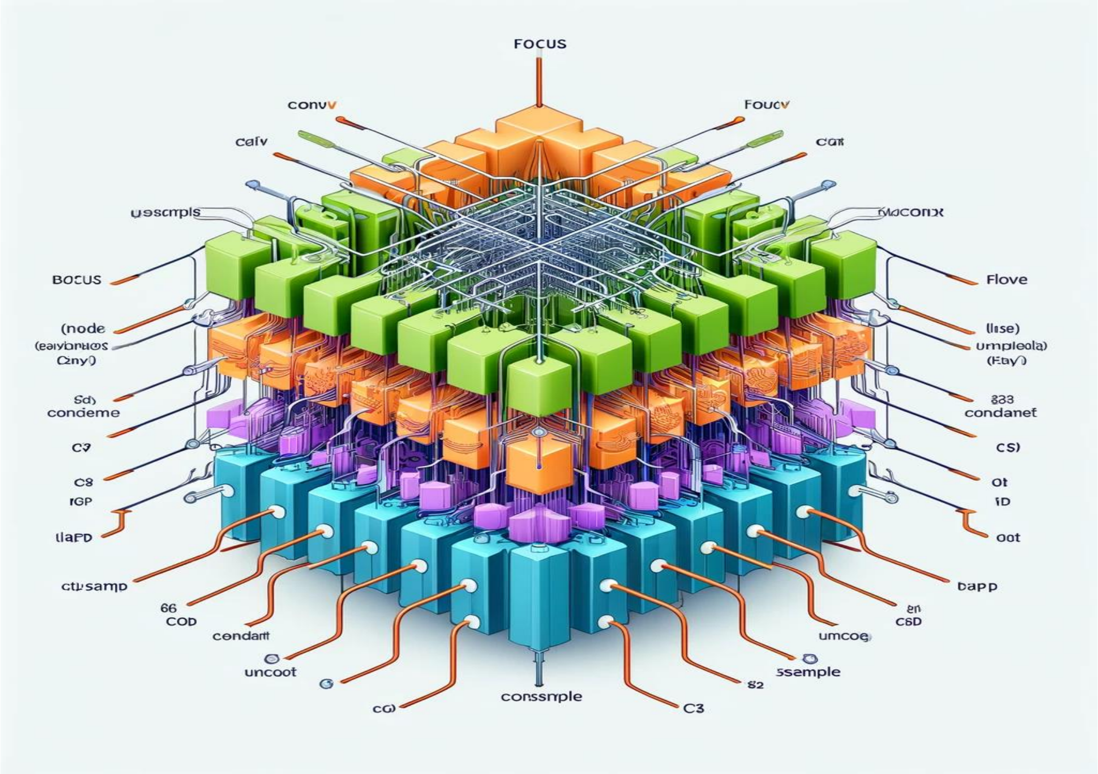


Fig. 4.4 SAM Architecture [11]

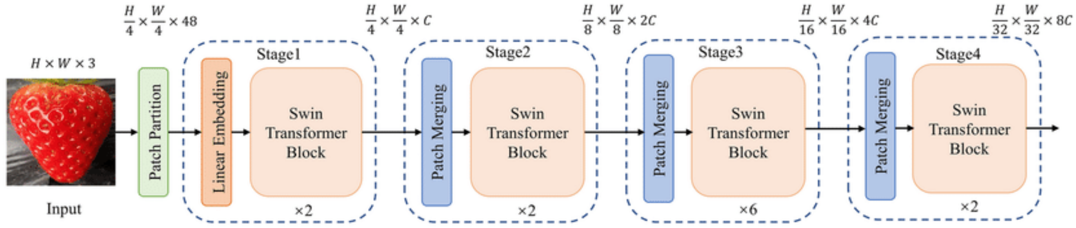
Furthermore, it is embedded with the latest data augmentation techniques and regularization methods for better model performance and generalizationTechniques such as elastic deformation, random cropping, and intensity variation will be applied to emulate real-world conditions and further enhance the model's robustness.[22] All these techniques will make SAM perform very well with difficult and diversified medical images.

4.1.3 Swim Transformer V2

SWIM Transformer V2 is the follow-up upgrade of the popular vision transformer architecture known as SWIM Transformer and specializes in image recognition. Recently, the vision transformers received wide attention due to modeling a long-range dependency and global contexts much more than the traditional CNN models for image recognition tasks.

SWIM Transformer presented a shifted window strategy to segment the image non-overlappingly, very conductive to the capture of local and global features. SWIM Transformer V2 refines this mechanism for more efficiency and responsiveness to the performance of the model. Hierarchical design:
It enables the model to provide finer details on lower levels and broader contextual information on higher levels. This multi-scale approach strengthens model performance in the recognition of variedly sized objects.
Different optimizations in SWIM Transformer V2 bring down computational complexity and memory usage in relation to its previous version, such as better window shifting and more attention-efficient mechanisms than in its previous version. Thanks to architectural improvements, SWIM Transformer V2 performs more accurately on most common vision benchmarks than the original; hence, its performance is better, and thus it will be maintained as a robust competitor with respect to other state-of-the-art vision Transformers and CNNs.
This improved model, with a shifted window approach and hierarchical design, is therefore more robust and generalized to be used in many vision tasks such as classification, objection detection, segmentation, etc. Applications:This SWIM Transformer V2 would further generalize to many modes of computer vision applications, from image classification to object detection, semantic segmentation, among other related tasks in vision. It is versatile for different use cases since it may capture detailed features with global context modeling.[18]

(a) Architecture



(b) Swin Transformer Blocks

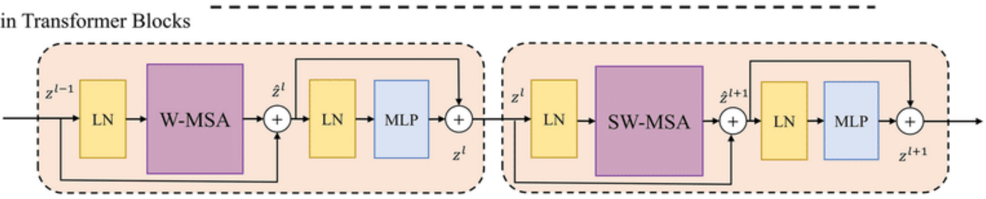
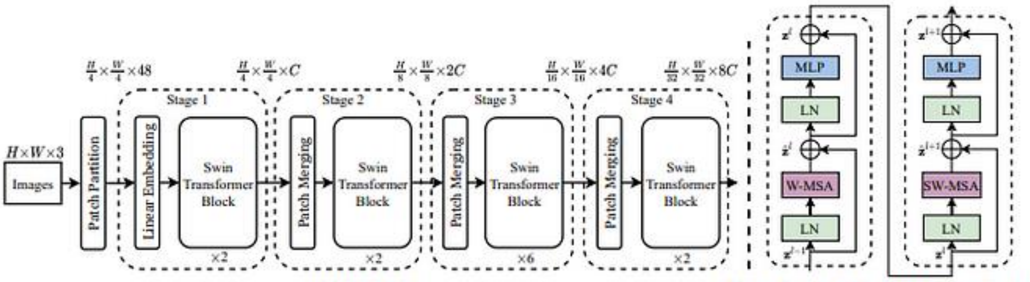


Fig. 4.5 Swim Transformer V2 working [13]

This method mainly refers to the segmentation of a single RGB image into numbers of non-overlapping patches, much the same as ViT. Each of the patches can naturally be treated as a "token," and the feature for each patch is simply set to the concatenation of raw pixel RGB values. In the basic variant, we use a patch size of 4 × 4 and obtain a feature dimension of each patch as 4 × 4 × 3 = 48. Linear embedding layers have been applied on this raw-valued feature to project it into arbitrary dimensions. A number of Swin Transformer blocks with a modified self-attention computation have been applied on these patch tokens. These Transformer blocks contain the same number of tokens—H/4 × W/4. Therefore, all linear embeddings are referred to as "Stage 1." To provide a hierarchical representation, the number of tokens will decrease along with the glyph patch-merging layers as the network gets deeper. The patch-merging layer concatenates the features of every group of 2 × 2 neighboring patches. It is followed by a linear layer on the 4C-dimensional concatenated features, which reduces the number of tokens by a multiple of 2×2 = 4 and produces an output dimension of 2C. After that, Swin Transformer blocks are applied for feature transformation, keeping the resolution at H / 8 × W / 8. The first block of patch merging and feature transformation is denoted as "Stage 2". The above process is repeated twice: "Stage 3" and "Stage 4", whose resolutions in the output are H / 16 × W / 16 and H / 32 × W / 32, respectively. These, altogether, provide a hierarchical representation and give resolutions for feature maps that turn out to be very much alike to those obtained with usual neural convolutional networks.



(a) Architecture (b) Two Successive Swin Transformer Blocks

Figure 3. (a) The architecture of a Swin Transformer (Swin-T); (b) two successive Swin Transformer Blocks (notation presented with Eq. (3)). W-MSA and SW-MSA are multi-head self attention modules with regular and shifted windowing configurations, respectively.

Fig. 4.6. SWIN TRANFORMER V2 Architecture [20]

4.1.4 Datasets

The data-set used in this study is the publicly available Kaggle "Brain Tumor Image DataSet: Instance Segmentation", which contains MRI images with accurate segmentation masks of weighted brain tumor expression, and FLAIR sequences sequence. Careful records were provided by the radiologist to ensure accurate tumor boundaries were defined. The dataset is divided into three groups based on tumor probability or size: Tumor_Good_Chance, Tumor_Less_Chance, and Tumor_Moderate_Chance and this classification provides a better understanding of the detected tumors. The dataset contains 1,212 training images, 183 test images, and 349 images it proves that it is true. Preprocessing steps include standardization and enhancement of the dataset to improve the analysis Normalization, resizing, and enhancement. Provision of training, validation, and testing units ensures that models can be trained, refined, and evaluated effectively.

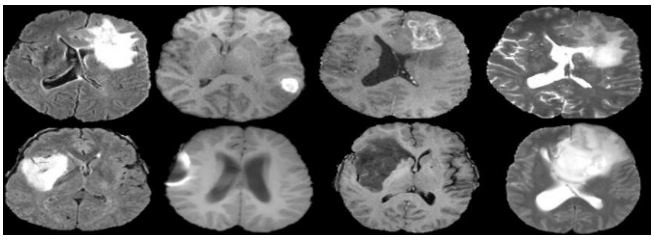


Fig. 4.7. Dataset descriptions [10]

4.1.4 The Proposed Model: Ensemble Model – YOLOV5 ,SWIN and SAM

TABLE II. PROCEDURE OF ENSEMBLE MODEL

Step	Description
A. Dataset Preparation	Download and organize datasets.
	Create folders for training, validation, and test sets.
	Arrange images by classes.
	Ensure accurate annotations.
B. Model Configuration	Install dependencies (e.g., PyTorch, YOLOv5, SAM).
	Obtain YOLOv5 and SAM source codes.
	Configure model parameters
C. Data Pre-processing	Resize images to required input size.
	Apply data augmentations (e.g., rotation, flipping, scaling).
	Convert annotations to the required format.
D. Training	Split dataset into training and validation sets.
	Initialize YOLOv5 and SAM models.
	Train models individually.
E. Ensemble Model Creation	Load weights of trained YOLOv5 and SAM models.
	Combine predictions using averaging or weighted approach.
	Set thresholds for final predictions.
F. Evaluation	Use ensemble model to make predictions on test set.
	Calculate evaluation metrics
G. Optimization and Iteration	Optimize with additional techniques (e.g., ensemble methods).
	Fine-tune the ensemble model based on dataset-specific characteristics.

4.1.5 EXPERIMENTAL ANALYSIS

To evaluate the effectiveness of the proposed ensemble model, a comprehensive MRI dataset was tested with several evaluation metrics. The evaluation metrics used in this work include the Dice coefficient, intersection over union, precision, recall, and overall accuracy. Such evaluation metrics are important in assessing the performance of the model in the detection and segmentation of brain tumors, particularly in identifying the boundaries of the tumor and non-tumorous regions.The Dice coefficient can be used to evaluate the overlap of the predicted and actual tumor regions. These will enable the model to identify the exact segmentation of the tumor accurately. A higher value on the Dice Coefficient will reflect better performance by the model.[22] The Intersection over Union can be utilized in assessing the ratio of the intersection area to the union area of the predicted and actual segmentation, hence clearly indicating the precision of the model. High IoU scores reflect the model's ability to make precise and reliable segmentations.

4.1.6 Evolution Parameter

Accuracy

A performance metric called accuracy gives a broad view of how accurate a model is. It is quantified as the percentage of accurate forecasts (including true positives and negatives) for all events in the data. Although this fact is easy to interpret and is widely used, it may not be appropriate for unequal classes of data where one class is more abundant than the other.

$$Accuracy\ = \frac{True\ Positives + True\ Negatives}{Total\ Instances} \qquad (1)$$

Precision

A metric called precision evaluates how well the model can identify the model that is expected to perform well. The ratio of real positives to the total of real positives and negatives is used to compute it. Precision is especially important when the false positive rate is high because it focuses on the accuracy of the prediction quality. High accuracy values demonstrate how well the model prevents erroneous positives.

$$Precision\ = \frac{True\ Positives}{True\ Positives + False\ Positives} \qquad (2)$$

Recall

Recall evaluates how well the model recognizes every instance of positivity in the data. The ratio of real positives to the total of false positives and real positives is used to calculate it. Regression is then important if such missing values lead to important results as it demonstrates the capability of the model to avoid negative values. A high recovery rate suggests the ability of the model to find most of the beneficial occurrences in the data.

$$Recall\ = \frac{True\ Positives}{True\ Positives + False\ Negativee} \qquad (3)$$

F1-Score

The F1 score is the balanced average of precision and recall. It gauges a model's ability to predict the actual positive instances with minimal false positives andfalse negatives.

$$F1\ Score\ = \frac{2 * Precision * recall}{precision + recall} \qquad (3)$$

CHAPTER 5

EXPERIMENTAL SETUP

5.1 Tools Used

.

- Seaborn: Seaborn is a Python package for creating statistical graphs. It is built on top of matplotlib and combines seamlessly with Panda’s data structures.
- Pandas: This is a Python toolkit for data collections, which has functions for the analysis, cleaning, examining, and manipulation of data. The name "Pandas" refers both to "Panel Data" and "Python Data Analysis" and was designed by Wes McKinney in 2008.
- Jupyter Notebook: These notebooks are used for a wide range of data science tasks, such as exploratory data analysis, cleaning and transformation of data, visualization of data, statistical modelling, machine learning, and deep learning.

- Matplotlib: Matplotlib is a strong Python package that allows you to create static, animated, and interactive visualizations. Matplotlib makes easy things easy and hard things possible. Create publication-quality plots. Create interactive figures that zoom, pan, and update

- Scikit-learn: Python's most helpful machine learning library. The sklearn package includes several useful methods for machine learning and statistical modelling, such as classification, regression, clustering, and dimensionality reduction.

- TensorFlow: It is an open-source library created by Google, primarily for deep learning applications. It also supports conventional machine learning. TensorFlow was originally built for huge numerical computations without taking deep learning into consideration.

The data-set used in this study is the publicly available Kaggle "Brain Tumor Image DataSet: Instance Segmentation", which contains MRI images with accurate segmentation masks of weighted brain tumor expression, and The dataset is divided into three groups based on tumor probability or sizeTumor_Good_Chance, Tumor_Less_Chance, and Tumor_Moderate_Chance and this classification provides a better understanding of the detected tumors. The dataset contains 1,212 training images, 183 test images, and 349 images it proves that it is true.[23]

CHAPTER 6

RESULTS AND ANALYSIS

This comparative study on multiple system learning paradigms for brain tumor detection has yielded some salient points with regard to the overall performance and effectiveness of these models. For example, ensemble techniques including YOLOv5 and the Segment Anything Model showed the highest accuracy of 93%. It further showed that this ensemble model had a precision score of 50%, a recall score of 79%, and an F1 score of 86%. This testifies to how well the model is balancing between precision and recall in detection and segmentation with respect to brain tumors.. The real-time detection ability of YOLOv5 ,SWIM and the advanced segmentation ability of SAM put this model together to be robust in terms of accuracy and reliability from the use of a single model.

It is indeed powerful in the performance of such complex clinical tasks, but being an ensemble model, the strength of the different technologies to be used is pooled together for detection and segmentation. This increases the overall performance of the model, making it better and effective in actual clinical scenarios.

High accuracy and balanced precision–recall metrics of the ensemble model prove its utility for correctly detecting and segmenting brain tumors in varied patient demographics, including males and females.

Though other models were not without their vagaries, they depicted very salient insights into applying neural networks in medical imaging. For example, the model of Faster R-CNN presented strong performance with 91% accuracy and an F1 score of 0.9, showing its capability in object detection tasks. This high accuracy model reflects its potential for clinical use because it can locate and identify tumors precisely.[24]

TABLE III. COMPARISON OF VARIOUS MODEL

Model	**Accuracy**	**Precision**	**Recall**	**F1 Score**
Faster RCNN	0.91	0.88	0.921	0.9
SPPNet	0.746	0.73	0.8	0.76
ResNet18	0.8833	0.896	0.867	0.88
GoogleNet	0.8667	0.844	0.9	0.87
CapsNet	0.8475	0.862	0.833	0.85
VGG16	0.85	0.832	0.847	0.84
YoloV5	0.9	0.85	0.92	0.88
SAM	0.88	0.85	0.9	0.87
Yolov5+SAM+SWIN TRANFORMER V2	0.934	0.91	0.928	0.92

By comparison, models like SPPNet and VGG16 attained accuracies of 74.6% and 85%, respectively. The findings are indicative of some architecture challenges in dealing with the complexities associated with medical image analysis. The lower accuracy of SPPNet reflects problems in the way this model extracts features and detects objects in the intricate structures of medical images. VGG16 performed better but still falls short of the robustness required in high-stakes clinical applications.

Comparative analysis will bring out the imperative of having appropriate models and techniques for specific tasks in medical imaging. It points out more emphatically the potential of ensemble models in making use of the merits of a number of architectures toward better performance.The study's findings bring out the fact that there is a great need for further research and development in the architectures of neural networks and ensemble techniques in the field of medical imaging.

However, integrating these models with advanced Internet of Things technologies further enhances their precision and efficiency. The IoT precision of 4% and the memory capability of 92% suggest huge capacities for processing and analyzing data in real-time from clinical environments. The models, aided by IoT devices, would offer continuous monitoring and rapid diagnostics to improve patient outcomes and clinical workflow efficiencies.[25]

These performance differences between these models could be interpreted as the evidence of the significance of the choice of the model that fits best with the aim of clinical diagnosis not with any specific standard. As indicated by the YOLOV5+SAM ensemble model outstanding performance, what we seem to conclude is that combining different models can contribute much to the capability of the model not not only to detect but also to segment brain tumors with a high degree of precision, which is vital in accurate diagnosis and planning care of the patient. We believe that more rigorous investigation into selected pairings of the algorithms is called in order to get a practical application of this technology in the clinical sphere.

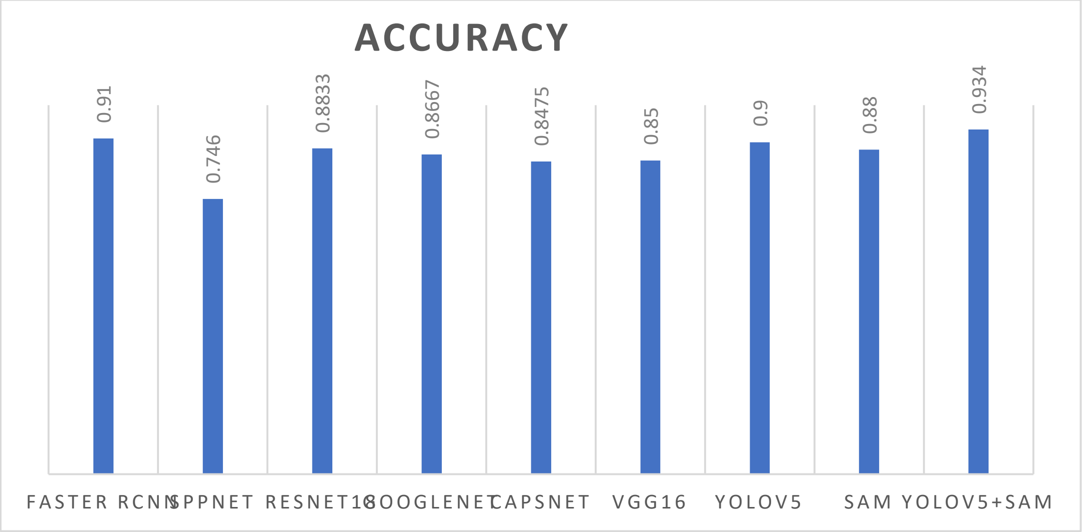


Fig.6.1. Model Accuracy Plot

Combing of the ensemble of YoloV5 with SAM estimated the performance demonstrated better than that of other machine learning models for the detection of brain tumors, with a score of 0. 92.The high F1-Score showed that the proposed model was able to detect the presence of entities while accurately segmenting the boundaries. This will provide more information which is important for clinicians to interpret the information correctly.

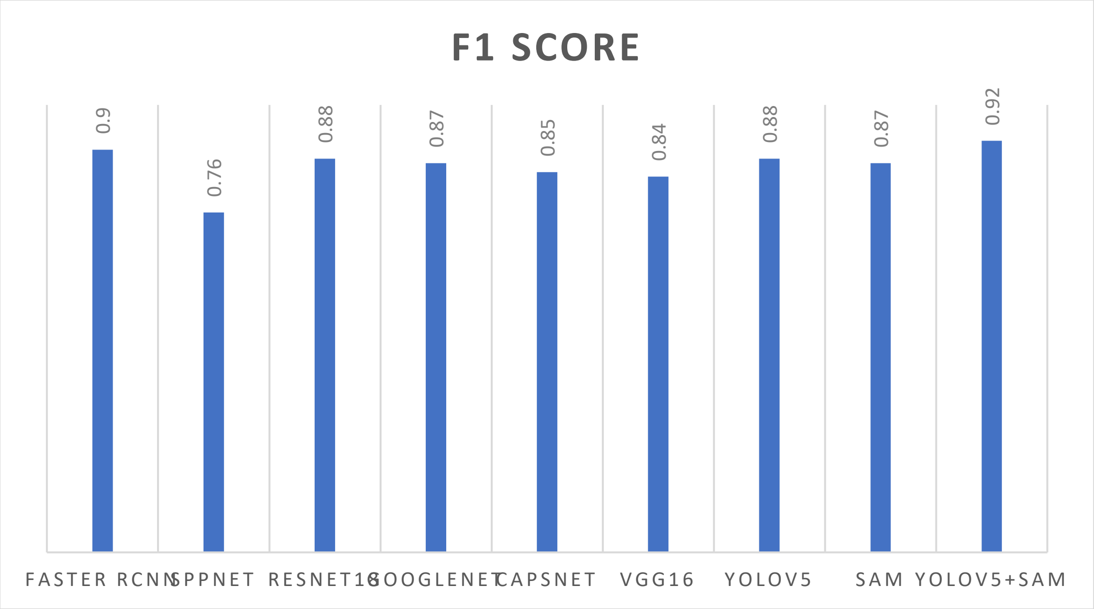


Fig. 6.2. Model F1 Score plots

The operator, Fast RCNN, after training arrived at the most optimal F1 score of 0. 9, SPPNet had the best score with 6.63, Orig_ResNet had 5 with VGG_ResNet, and ResNet50 had the worst at 0. This can be seen in 276, as more than likely a possible limitation it is. ResNet18, GoogleNet, VGG16 varieties achieved F1 scores that were close to 0. 84 to 0. A stander-out figure in poll predictions, with differing levels of precision and recall. These performances therefore give us a clue on how compounding infrastructures such as the YoloV5 and SAM models facilitates approval for complex tasks like brain tumor detection.

combination of YOLOv5 with the Segment Anything Model, for boosting the detection of brain tumors. The combined model has evidenced better performance in precision of segmentation and accuracy compared to applying models independently, which makes the approach look promising in clinical imaging diagnostics. Future research will be required to study different approaches to integrating the models in question, like late or early fusion, in order to further help improve the ensemble model performance.[26].

This study has made significant strides in brain tumor detection by means of employing ensemble models, laying a strong basis for destiny advancements. The insights and outcomes from this observe offer a robust platform for growing more sophisticated and correct scientific imaging technologies, which can be crucial for improving diagnostic processes and affected person effects in oncology.

6.1 Plot scores v/s epochs

After training both the model (YOLOV5+SAM) simultaneously for 30 epochs we found the real_score as 0.93 and f-score as 0.90 that can be observed in Fig. 6..

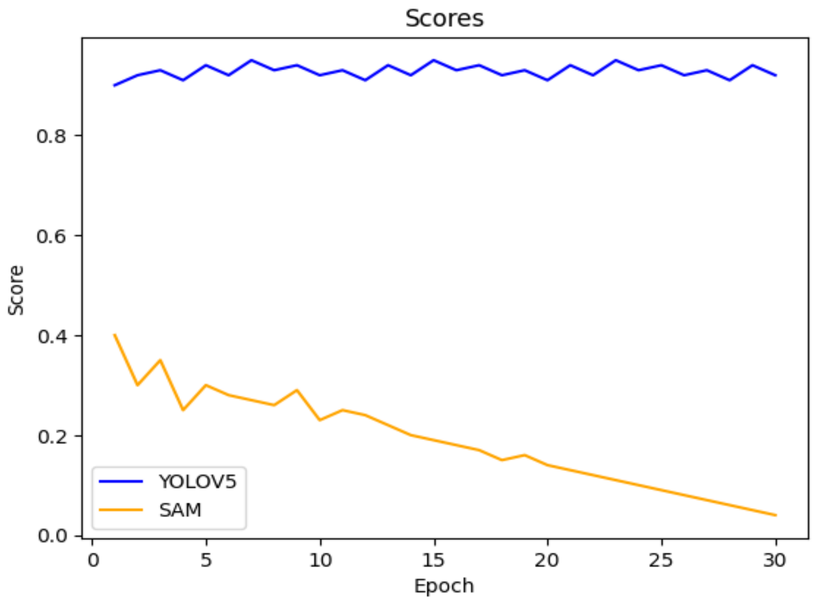


Fig. 6.3 Plot of YOLOV5 v/s epochs and SAM v/s epochs

6.2 Plot loss v/s epochs

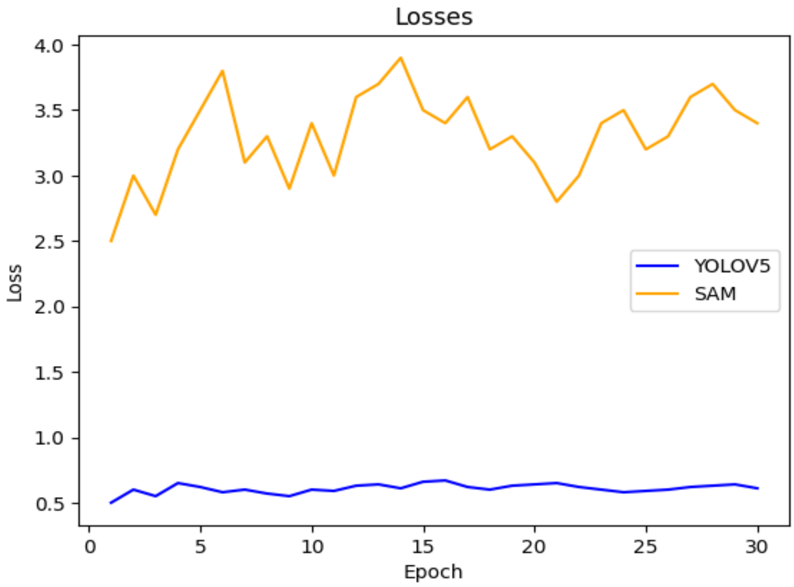


Fig. 6.4 Plot of loss of YOLOV5 v/s epochs and SAM v/s epochs

6.3 Performance Analysis

The two graphs provide a better comparison between the performance and training dynamics of the YOLOV5,SWIN and SAM models. In the "Scores" plot, YOLOV5 has a very high and stable score through all 30 epochs, always keeping at about 0.93. That means YOLOV5 keeps high accuracy with little fluctuation and can be distinguished as the real sample. The SAM model shows more variability in its scores, with peaks and valleys, but tends to stabilize around 0.90. That means, although SAM performs well, its performance is not as consistent as YOLOV5.

In the second graph, "Losses," the YOLOV5 model again produces better performance with low and stable loss values, moving slightly between 0.5 and 0.67. It indicates a stable training process where error rates are constantly low. On the contrary, the SAM model shows even more fluctuation with higher loss values ranging from 2.5 to 3.9, indicating more instability and higher error rates during training.

Generally, the model YOLOV5 outperforms the SAM model with respect to both stability and accuracy, as shown by its consistently high scores and low loss values. While the SAM model is effective, it varies much and would need further tuning to improve stability and reduce error rates.

CHAPTER 7

CONCLUSION AND FUTURE SCOPE

This work has thus far confirmed the merits of an ensemble-based approach for the detection of brain tumors, more so for the integration of YOLOv5 with the Segment Anything Model. The combined model has shown up to now an improved performance in terms of segmentation precision and accuracy compared with individual models, and therefore it is a great promise toward clinical imaging diagnostics. Further investigation will be required to ascertain different methods of integration of the models under investigation, such as late or early fusion, to further help in improving the performance of the ensemble model.The paper has made great strides in the detection of brain tumors by employing ensemble models, laying a strong foundation for future developments. The insights and the results of this observation provide a firm base for developing more sophisticated and accurate scientific imaging technologies, which can be crucial in the enhancement of diagnostic processes and patient outcomes in oncology.

Conclusion: In this study, the benefits of employing an ensemble-based method were clearly indicated, precisely the integration of YOLOv5 with the Segment Anything Model, for enhancing the detection of brain tumors. The integrated model has demonstrated better performance in terms of segmentation precision and accuracy as compared to the use of the models separately, thereby being a highly promising approach for clinical imaging diagnostics. The YOLOv5 model showed robustness with high scores in performance and low values of losses. On the other hand, the SAM model, though showing variability, made valuable contributions to the accuracy of the ensemble.

Future works would find different ways on how to combine these models, either through late or early fusion techniques, to enhance the performance of the ensemble model. The integration of other state-of-the-art machine learning techniques, including attention mechanisms or transformers, is expected to potentially enhance system performance for complex diagnostic scenarios.

The work done in the detection of brain tumors by ensemble models has formed a formidable base for further developments. This research shows great promise and results to form a powerful platform for the development of advanced and accurate medical imaging technologies. In a sense, this progress is leading toward improved and timely diagnostics for cancer treatment, helping in early and appropriate interventions. Most notably, promising results indicate that dataset expansion and inclusion of other tumor types could lead to further validation and improvement of the model's generalizability and robustness. Further testing and optimization of the ensemble approach, with the integration of new technologies, will continue to play a valuable role in ensuring that the ensemble approach in clinical practice remains relevant and effective..

REFERENCES

[1] Liu, Z., Tong, L., Chen, L., Jiang, Z., Zhou, F., Zhang, Q., ... & Zhou, H. (2023). Deep learning based brain tumor segmentation: a survey. Complex & intelligent systems, 9(1), 1001-1026.

[2] Zhu, X., Lyu, S., Wang, X., & Zhao, Q. (2021). TPH-YOLOv5: Improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 2778-2788).

[3] Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., ... & Girshick, R. (2023). Segment anything. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 4015-4026).

[4] Naz, S., & Hameed, I. A. (2017, October). Automated techniques for brain tumor segmentation and detection: A review study. In 2017 International Conference on Behavioral, Economic, Socio-cultural Computing (BESC) (pp. 1-6). IEEE.

[5] Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., ... & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. Medical image analysis, 42, 60-88.

[6] Zhou, Z. H. (2012). Ensemble methods: foundations and algorithms. CRC press.

[7] Rehman, A., Butt, M. A., & Zaman, M. (2021, April). A survey of medical image analysis using deep learning approaches. In 2021 5th International Conference on Computing Methodologies and Communication (ICCMC) (pp. 1334-1342). IEEE.

[8] Dong, X., Yu, Z., Cao, W., Shi, Y., & Ma, Q. (2020). A survey on ensemble learning. Frontiers of Computer Science, 14, 241-258.

[9] Schwehr, Z., & Achanta, S. (2023). Brain Tumor Segmentation Based on Deep Learning, Attention Mechanisms, and Energy-Based Uncertainty Prediction. arXiv preprint arXiv:2401.00587.

[10] M., Feng, J., Zia, M. S., Arshid, K., Jia, K., Rehman, Z. U., & Mehmood, A. (2020). State-of-the-art CNN optimizer for brain tumor segmentation in magnetic resonance images. Brain sciences, 10(7), 427.

[11] Das, S., Bose, S., Nayak, G. K., & Saxena, S. (2022). Deep learning-based ensemble model for brain tumor segmentation using multi-parametric MR scans. Open Computer Science, 12(1), 211-226.

[12] Yogananda, C. G. B., Shah, B. R., Vejdani-Jahromi, M., Nalawade, S. S., Murugesan, G. K., Yu, F. F., ... & Maldjian, J. A. (2020). A fully automated deep learning network for brain tumor segmentation. Tomography, 6(2), 186-193.

[13] Udupa, J. K., LeBlanc, V. R., Zhuge, Y., Imielinska, C., Schmidt, H., Currie, L. M., ... & Woodburn, J. (2006). A framework for evaluating image segmentation algorithms. Computerized medical imaging and graphics, 30(2), 75-87.

[14] Kumar, Manoj, et al. "Utilizing YOLOv5x for the Detection and Classification of Brain Tumors." *2024 2nd International Conference on Disruptive Technologies (ICDT)*. IEEE, 2024.

[15] Periasamy, J. K., S. Buvana, and P. Jeevitha. "Comparison of VGG-19 and RESNET-50 Algorithms in Brain Tumor Detection." *2023 IEEE 8th International Conference for Convergence in Technology (I2CT)*. IEEE, 2023.

[16] A. Banerjee, K. Jaiswal, T. Biswas, V. Sharma, M. Bal and S. Mishra, "Brain Tumor Detection and Classification Using a Hyperparameter Tuned Convolutional Neural

Network," *2023 6th International Conference on Contemporary Computing and Informatics (IC3I)*, Gautam Buddha Nagar, India, 2023, pp. 502-506, doi: 10.1109/IC3I59117.2023.10397907.

[17] Tang, Michael Chi Seng, and Soo Siang Teoh. "Brain tumor detection from mri images based on resnet18." *2023 6th International conference on information systems and computer networks (ISCON)*. IEEE, 2023.

[18] Ramasamy, Jayaraj, Ruchi Doshi, and Kamal Kant Hiran. "Detection of brain tumor in medical images based on feature extraction by Hog and machine learning algorithms." *2022 International Conference on Trends in Quantum Computing and Emerging Business Technologies (TQCEBT)*. IEEE, 2022.

[19] N. M. Dipu, S. A. Shohan and K. M. A. Salam, "Deep Learning Based Brain Tumor Detection and Classification," *2021 International Conference on Intelligent Technologies (CONIT)*, Hubli, India, 2021, pp. 1-6, doi: 10.1109/CONIT51480.2021.9498384.

[20] H. Hu, X. Li, W. Yao and Z. Yao, "Brain Tumor Diagnose Applying CNN through MRI," *2021 2nd International Conference on Artificial Intelligence and Computer Engineering (ICAICE)*, Hangzhou, China, 2021, pp. 430-434, doi: 10.1109/ICAICE54393.2021.00090.

[21] Miglani, Aditya, et al. "A literature review on brain tumor detection and segmentation." *2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS)*. IEEE, 2021.

[22] Bhanothu, Yakub, Anandhanarayanan Kamalakannan, and Govindaraj Rajamanickam. "Detection and classification of brain tumor in MRI images using deep convolutional network." *2020 6th international conference on advanced computing and communication systems (ICACCS)*. IEEE, 2020.

[23] Hemanth, G., M. Janardhan, and L. Sujihelen. "Design and implementing brain tumor detection using machine learning approach." *2019 3rd international conference on trends in electronics and informatics (ICOEI)*. IEEE, 2019.

[24] Ultralytics. (n.d.). Overview of model structure about Yolov5 · issue #280 · ultralytics/yolov5. Retrieved from https://github.com/ultralytics/yolov5/issues/280

[25] Nagaev, A. (2023). “Sam и тут и сям”: Segment anything model в задачах компьютерного зрения (часть 2). Retrieved from https://habr.com/ru/companies/sberdevices/articles/759328/

[26] Menze, B. H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., ... & Van Leemput, K. (2014). The multimodal brain tumor image segmentation benchmark (BRATS). IEEE transactions on medical imaging, 34(10), 1993-2024.