A CUTTING-EDGE APPROACH FOR MRI BRAIN TUMOR DETECTION USING SWIN-LSTM

A PROJECT REPORT

Submitted by

INISH RAJ B (202009017)

BHARATH D (202009007)

in partial fulfillment for the award of the degree

of

BACHELOR OF TECHNOLOGY

in

ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

MEPCO SCHLENK ENGINEERING COLLEGE, SIVAKASI

(An Autonomous Institution affiliated to Anna University Chennai)

APRIL 2024

MEPCO SCHLENK ENGINEERING COLLEGE, SIVAKASI

AUTONOMOUS

DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

BONAFIDE CERTIFICATE

Certified that this project report **"A CUTTING-EDGE APPROACH FOR MRI BRAIN TUMOR DETECTION USING SWIN-LSTM"** is the bonafide work of **"BHARATH D (202009007), INISH RAJ B(202009017)"** who carried out the project work under my supervision.

SIGNATURE

Dr. J. ANGELA JENNIFA SUJANA
Professor & Head
Artificial Intelligence and Data Science Department,
Mepco Schlenk Engg. College, Sivakasi
Virudhunagar Dt. – 626 005

SIGNATURE

Dr. A.SHENBAGARAJAN
Associate Professor & Supervisor
Artificial Intelligence and Data Science Department
Mepco Schlenk Engg. College, Sivakasi
Virudhunagar Dt. – 626 005

Submitted to the Viva-Voce examination held on __/__/___.

INTERNAL EXAMINER **EXTERNAL EXAMINER**

ABSTRACT

Accurate brain tumor segmentation is crucial for diagnosis, treatment planning, and prognosis. Although magnetic resonance imaging (MRI) is a commonly utilized technique for brain tumorsegmentation, the heterogeneity of the tumor and the similarity of the surrounding healthy tissue can make the process difficult. This paper suggests a unique method for segmenting MRI brain tumors utilizing Long Short-Term Memory (LSTM) networks and Swin-Transformer (Swin-T) networks. On a range of image classification tasks, the hierarchical transformer-based vision model Swin-T achieves state-of-the-art performance. Sequential information is best captured by LSTM networks, which is advantageous for medical picture segmentation tasks that need spatial context. The suggested approach combines the long-range dependency learning of LSTM with the feature extraction capabilities of Swin-T. High-level features are extracted from MRI images by the Swin-T encoder, and segmentation masks are gradually improved by the LSTM decoder. The model incorporates a residual connection between the encoder and the decoder to improve accuracy by preserving spatial information that is lost during encoding. When tested on a benchmark dataset for brain tumor segmentation, the Swin-LSTM network performs better than previous methods. It generates robust and accurate segmentation masks, particularly for complicated tumors with irregular shapes. The efficacy of LSTM decoder, residual connection, and Swin-T encoder has been verified by ablation research. To sum up, this work provides a Swin-LSTM technique that uses LSTM for long- range dependency capture and Swin-T for feature extraction for MRI brain tumor segmentation. Its potential as a tool for clinical brain tumor diagnosis and treatment planning is enhanced by integration with a residual connection.

ACKNOWLEDGEMENT

First and foremost, we would like to thank the LORD ALMIGHTY for his abundant blessings that is showered upon our past, present and future successful endeavors.

We extend our sincere gratitude to our college management and principal **Dr. S. Arivazhagan M.E., Ph.D.**, for providing sufficient working environment such as systems and library facilities.

We would like to extend our heartfelt gratitude to our Head of the Artificial Intelligence and Data Science Department **Dr. J. Angela Jennifa Sujana M. Tech., Ph.D.**, Professor for giving us the golden opportunity to undertake the project of this nature and for her most valuable guidance.

We would also like to extend our gratitude to **Mrs. L. Prasika M.E., (Ph.D.,)** Assistant Professor, Department of Artificial Intelligence and Data Science, for being our project coordinator and directing us throughout our project.

We would also like to extend our gratitude and sincere thanks to **Dr.A. Shenbagarajan B.E (EEE), M.E(CSE), Ph.D.,** Associate Professor, Department of Artificial Intelligence and Data Science for being our project guide and for his moral support and suggestions. He has put his valuable experience and expertise in directing, suggesting and supporting us throughout the project to bring our best.

Our sincere thanks to our revered faculty members, lab technicians and beloved family and our friends for their help at right time for making this project a successful one.

TABLE OF CONTENTS

CHAPTER NO.	TITLE	PAGE NO.
	ABSTRACT	**i**
	ACKNOWLEDGEMENT	**ii**
	LIST OF FIGURES	**xviii**
	LIST OF ABBREVATIONS	**xxvii**
1.	**INTRODUCTION**	**1**
	1.1 BACKGROUND AND CONTEXT	1
	1.2 PROBLEM STATEMENT	2
	1.3 OBJECTIVE	3
	1.4 METHODOLOGY OVERVIEW	4
2.	**LITERATURE SURVEY**	**7**
3.	**SYSTEM DESIGN**	**17**
	3.1 SYSTEM WORKFLOW	17
	3.2 METHODOLOGY	18
	3.2.1 DATA ACQUISITION	18
	3.2.2 DATA PREPROCESSING	19
	3.2.3 FEATURE EXTRACTION USING LSTM	21
	3.2.4 CLASSIFICATION USING SWIN TRANSFORMER	24
4.	**SYSTEM STUDY**	**27**
	4.1 LONG SHORT-TERM MEMORY (LSTM) NETWORKS	27
	4.2 SWIN TRANSFORMER	30
	4.3 SWIN-LSTM MODEL	39
5.	**RESULTS AND DISCUSSIONS**	**42**
	5.1 AUGMENTED IMAGES	42
	5.2 ACCURACY	42
	5.3 ACCURACY LOSS GRAPH	43

	5.4 CLASSIFICATION REPORT	45
	5.5 DASHBOARD	46
6.	**CONCLUSION AND FUTURE ENHANCEMENT**	**48**
	APPENDIX I	**49**
	APPENDIX II	**50**
	REFERENCES	**57**
	PUBLICATIONS	

LIST OF FIGURES

S.No.	Image	Page Number
3.1.	System Workflow	17
3.2.1.1.	Types of Tumor	19
3.2.2.1.	MRI image before and after pre-processing	21
3.2.3.1.	LSTM Architecture	23
3.2.4.1.	SWIN Transforner Blocks	25
3.2.4.2.	Architecture of SWIN Transformer	26
4.2.1	Patch Partitioning	31
4.2.2.	Shifted Windows	32
4.2.3.	Relative Position Embedding	32
4.2.4.	Self Attention Layer	33
4.2.5.	Self Attention Head	33
4.2.6.	Cyclic Shift	34
4.2.7.	Masked MSA	34
4.2.8.	Batch Computation	35
4.2.9.	Architecture of SWIN Transformer	35
4.2.10.	Variants of SWIN Transformer	39
5.1.1.	Before and After Augmentation	42
5.2.1.	Accuracy for 100 epochs	43
5.3.1.	Accuracy and Loss Graph for SWIN Transformer (2 classes)	44
5.3.2.	Accuracy and Loss Graph for SWIN Transfromer (4 classes)	44
5.3.3.	Accuracy and Loss Graph for SWIN-LSTM Model (2 Classes)	45
5.3.4.	Accuracy and Loss Graph for SWIN-LSTM Model (4 Classes)	45
5.4.1.	Classification Report	46

LIST OF ABBREVIATIONS

ABBREVIATIONS		EXPANSION
CNN	-	**C**onvolutional **N**eural **N**etwork
AI	-	**A**rtificial **I**ntelligence
LSTM	-	Long Short Term Memory
SWIN	-	Shifted Windows
ACC	-	**A**ccuracy
MRI	-	Magnetic Resonance Imaging
RNN	-	Recurrent Neural Network
W-MSA	-	Window Multi-head Self Attention
SW-MSA	-	Shifted Window Multi-head Self Attention
LN	-	Layer Normalization
MLP	-	Multi Layer Perceptron

CHAPTER 1
INTRODUCTION

1.1 BACKGROUND AND CONTEXT

In recent decades, there have been significant strides in medical imaging technologies, transforming healthcare practices and enhancing patient care. Among these modalities, Magnetic Resonance Imaging (MRI) has emerged as a particularly potent tool for examining internal anatomical structures with exceptional precision and clarity. Its capacity to generate detailed, multi-dimensional images of soft tissues, such as the brain, has positioned it as a cornerstone in detecting and diagnosing a myriad of medical ailments, notably brain tumors.

Brain tumors represent a significant health concern worldwide, with diverse manifestations and clinical implications. Early detection and accurate diagnosis are critical for initiating timely treatment interventions and improving patient prognosis. MRI plays a pivotal role in this regard, offering superior soft tissue contrast and enabling precise localization and characterization of tumors within the intricate neural anatomy.

However, despite its diagnostic utility, interpreting MRI images for brain tumor detection remains a complex and challenging task. Brain tumors exhibit diverse morphological and pathological features, ranging from well-defined masses to infiltrative lesions with indistinct borders. Furthermore, subtle abnormalities and variations in tissue characteristics can often mimic or obscure tumor presence, complicating the diagnostic process.

Traditionally, the interpretation of MRI images has relied heavily on manual analysis by skilled radiologists, requiring meticulous examination and subjective judgment. This method is naturally time-consuming, requires significant labor, and is susceptible to variations among observers, resulting in discrepancies in both diagnosis and treatment strategies. Moreover, the growing volume of medical imaging data, coupled with increasing demands on healthcare resources, underscores the need for more efficient and scalable solutions to streamline the diagnostic workflow and enhance clinical decision-making.

In response to these challenges, researchers and clinicians have turned to advanced computational techniques, particularly deep learning and artificial intelligence, to augment and automate the analysis of medical imaging data. Deep learning architectures like convolutional neural networks (CNNs) have proven highly effective across diverse image analysis tasks, from recognizing objects to segmenting and classifying them. With access to extensive datasets and robust computational capabilities, these models excel at discerning intricate patterns and features from raw data. Consequently, they empower automated identification and characterization of anomalies with unparalleled precision and speed.

In recent years, transformer-based architectures have emerged as a novel approach to image processing, offering distinct advantages in capturing long-range dependencies and contextual information. The Swin Transformer, introduced by Liu et al., represents a significant breakthrough in this domain, with superior performance and scalability in processing high-resolution images. By incorporating self-attention mechanisms and hierarchical representations, Swin Transformer networks excel at capturing spatial relationships and contextual dependencies within images, making them well-suited for analyzing MRI scans with fine-grained details.

Against this backdrop, the integration of Swin Transformer networks with Long Short-Term Memory (LSTM) networks presents a promising avenue for enhancing MRI brain tumor detection. By combining the spatial encoding capabilities of Swin Transformer with the temporal modeling capabilities of LSTM, the proposed approach aims to provide a comprehensive framework for accurate and robust tumor detection, capable of capturing both static and dynamic features from sequential MRI data. This combination of temporal and geographical data has the potential to advance the discipline of neuroimaging and improve patient care in the management of brain tumors by increasing the precision, efficacy, and interpretability of automated tumor detection systems.

1.2 PROBLEM STATEMENT

The focal point of this study is the development of an automated system geared towards detecting brain tumors from MRI images, aiming to tackle the complexities posed by the intricate structures of tumors and the subtle nuances of abnormalities. Conventional manual interpretation of MRI scans by radiologists is time-intensive and prone to variations,

underscoring the need for more streamlined and dependable solutions. Leveraging advanced deep learning techniques, particularly Swin Transformer networks and Long Short-Term Memory (LSTM) networks, this research endeavors to construct a robust framework for MRI brain tumor detection. By amalgamating spatial encoding with temporal modeling, the proposed approach endeavors to capture both static and dynamic features inherent in sequential MRI data, thereby augmenting the accuracy, efficiency, and interpretability of automated tumor detection systems. The amalgamation of spatial and temporal information holds significant promise in transforming neuroimaging practices and fostering better patient outcomes in the realm of brain tumor diagnosis and treatment.

1.3 OBJECTIVE

The core objective of this project is to create a robust framework for detecting brain tumors in MRI scans, utilizing the strengths of Swin Transformer networks and Long Short-Term Memory (LSTM) networks. This framework aims to enhance the accuracy, efficiency, and interpretability of automated tumor detection systems by merging spatial encoding with temporal modeling. The specific goals of the study include:

- Exploring the efficacy of Swin Transformer networks in capturing spatial relationships within MRI images.
- Assessing the capability of LSTM networks to model temporal dependencies in sequential MRI data.
- Developing an integrated framework that combines Swin Transformer and LSTM networks for MRI brain tumor detection.
- Investigating the potential benefits of using multi-modal MRI data to improve tumor detection and classification.
- Evaluating the performance of the proposed framework on standardized datasets and comparing it against established state-of-the-art methods.

Specific goals include refining the model architecture to enhance segmentation accuracy and computational efficiency, addressing class imbalance through data augmentation and loss functions, and implementing uncertainty estimation techniques for reliable predictions and

interpretability. The ultimate aim is to empower radiologists and clinicians with a state-of-the-art tool for accurate brain tumor detection, contributing to improved patient care and outcomes.

1.4 METHODOLOGY OVERVIEW

The proposed methodology involves several key steps, including data preprocessing, model development, training, and evaluation. Firstly, MRI data will be preprocessed to enhance image quality and standardize the data format. Next, Swin Transformer networks will be employed to capture spatial relationships within the MRI images, providing a rich representation of the underlying structures. Concurrently, LSTM networks will serve to capture temporal dependencies within sequential MRI data, facilitating the examination of evolving changes in tumor morphology over time. These two networks will be fused to establish a holistic framework for detecting brain tumors in MRI scans. Furthermore, the exploration of multi-modal MRI data, encompassing T1-weighted, T2-weighted, and FLAIR images, will be undertaken to bolster tumor detection and classification efforts. Ultimately, the efficacy of the proposed framework will be gauged through its performance evaluation on standardized datasets, alongside comparisons with prevailing state-of-the-art methodologies to ascertain its efficacy.

A. Data Collection and Preprocessing

The initial phase involves the acquisition of MRI datasets containing brain images with tumor annotations. Various public repositories, such as the BRATS (Brain Tumor Segmentation) dataset, provide annotated MRI scans suitable for training and evaluation. Data collection involves sourcing MRI datasets from reputable repositories such as the BRATS dataset, ensuring diversity in tumor types, sizes, and imaging conditions. Preprocessing steps include image registration, skull stripping, intensity normalization, and Resampling ensures uniform voxel dimensions across datasets, reducing computational complexity and facilitating model training.

B. Long Short-Term Memory Networks (LSTM)

LSTM networks are pivotal in capturing temporal dependencies inherent in sequential MRI data. The architecture consists of interconnected cells with input, forget, and output gates,

enabling the network to retain information over extended sequences. For brain tumor detection, LSTM networks analyze dynamic changes in tumor morphology, aiding in accurate segmentation and tracking tumor progression over time. The LSTM's ability to learn from sequential data complements Swin Transformers' spatial encoding, enhancing the overall model performance.

C. Swin Transformers

Swin Transformers represent a breakthrough in vision transformer architectures, adept at processing high-resolution medical images with fine-grained details. Leveraging self-attention mechanisms and hierarchical representations, Swin Transformers capture spatial relationships within MRI scans effectively. The Swin Transformer encoder extracts high-level features from MRI images, providing a comprehensive representation of tumor characteristics, while also preserving contextual information crucial for accurate segmentation.

D. Integration of Swin Transformer Networks with LSTM

The integration of Swin Transformer networks with LSTM forms the cornerstone of the proposed approach. The Swin Transformer encoder extracts rich spatial features from MRI scans, which are then fed into the LSTM decoder for sequential analysis. This integration allows the model to leverage the complementary strengths of both architectures, combining spatial encoding with temporal modeling for enhanced tumor detection and segmentation. Additionally, a residual connection between the encoder and decoder facilitates the propagation of spatial information, further improving segmentation accuracy.

E. Model Training and Evaluation

The integrated model undergoes rigorous training on annotated MRI datasets using supervised learning techniques. Training involves optimizing model parameters to minimize loss functions, with validation performed on separate datasets to prevent overfitting. Evaluation metrics such as Dice coefficient, sensitivity, specificity, and Hausdorff distance are used to assess the model's performance, providing quantitative measures of segmentation accuracy and robustness.

F. Deployment and Clinical Validation

Once trained and evaluated, the model undergoes rigorous testing in real-world clinical settings, where its performance is assessed by radiologists and medical professionals. Clinical validation ensures that the model meets the stringent requirements of clinical practice, providing reliable and accurate assistance in diagnosing and treating brain tumors. Feedback from clinicians guides further refinement and optimization, ultimately leading to the development of a robust and clinically relevant automated detection system.

G. Future Directions

Beyond the initial development and validation, future research directions may include exploring advanced techniques for model interpretation and visualization, investigating the integration of additional imaging modalities or clinical data sources, and evaluating the framework's performance in real-time or interactive settings. Collaborations with healthcare institutions and industry partners may facilitate the deployment and adoption of the developed framework in clinical practice, ultimately improving patient care and outcomes in the management of brain tumors.

CHAPTER 2
LITERATURE SURVEY

1) "Residual Swin Transformer Channel Attention Network for Image Demosaicing" [1]

Xing and Egiazarian delve into the realm of image demosaicing with their work titled "Residual Swin Transformer Channel Attention Network for Image Demosaicing." In this study, presented at the 2022 10th European Workshop on Visual Information Processing (EUVIP) in Lisbon, Portugal, they tackle the challenge of reconstructing high-quality color images from raw Bayer-patterned sensor data. Their approach introduces a novel deep learning architecture, the Residual Swin Transformer Channel Attention Network, which extends the Swin Transformer framework. By incorporating channel attention mechanisms, the model enhances feature extraction and interpolation accuracy, particularly focusing on the handling of color information. Through this innovative integration of techniques, Xing and Egiazarian demonstrate significant improvements in demosaicing performance, showcasing the potential of deep learning and transformer-based architectures in advancing image processing methodologies.

2) "SwinT-Unet: Hybrid architecture for Medical Image Segmentation Based on Swin transformer block and Dual-Scale Information'' [2]

Atek, Mehidi, Jabri, and Belkhiat introduce "SwinT-Unet: Hybrid architecture for Medical Image Segmentation Based on Swin transformer block and Dual-Scale Information" at the 2022 7th International Conference on Image and Signal Processing and their Applications (ISPA) in Mostaganem, Algeria. In this study, the authors address the critical task of medical image segmentation by proposing a novel hybrid architecture. Named SwinT-Unet, the model integrates Swin Transformer blocks with dual-scale information to enhance segmentation accuracy and efficiency. By leveraging the capabilities of Swin Transformers in capturing long-range dependencies and incorporating dual-scale information, the proposed architecture demonstrates improved performance in segmenting medical images. This work contributes to the advancement of deep learning techniques in medical image analysis, particularly in the domain of segmentation, by effectively harnessing the power of transformer-based architectures and multi-scale feature extraction.

3) ‘‘A Swin Transformer-Based Fusion Approach for Hyperspectral Image Super-Resolution’’ [3]

Yang, Wang, Zhao, Song, and Zhang present their work titled "A Swin Transformer-Based Fusion Approach for Hyperspectral Image Super-Resolution" at IGARSS 2023 - the 2023 IEEE International Geoscience and Remote Sensing Symposium in Pasadena, CA, USA. In this study, the authors tackle the challenge of hyperspectral image super-resolution by proposing a novel fusion approach based on Swin Transformer architecture. By leveraging the capabilities of Swin Transformers in capturing long-range dependencies and incorporating spectral attention mechanisms, the proposed approach aims to enhance the spatial resolution of hyperspectral images. Through innovative fusion techniques and deep learning methodologies, Yang et al. demonstrate significant improvements in super-resolution performance, paving the way for enhanced analysis and interpretation of remote sensing data. This work contributes to the advancement of image processing techniques in remote sensing applications, particularly in the domain of hyperspectral imaging, by effectively harnessing the power of transformer-based architectures and spectral attention mechanisms.

4) ‘‘Revolutionizing COVID-19 Diagnosis with Swin Transformer: A Comparative Study on CT Image Attention Analysis and CNN Models Performance’’ [4]

Yang explores the transformative potential of Swin Transformer architectures in his study titled "Revolutionizing COVID-19 Diagnosis with Swin Transformer: A Comparative Study on CT Image Attention Analysis and CNN Models Performance." Conducted at the 2023 4th International Conference on Computer Vision, Image and Deep Learning (CVIDL) in Zhuhai, China, Yang investigates the application of Swin Transformers in enhancing COVID-19 diagnosis using computed tomography (CT) imaging. By conducting comparative analyses between Swin Transformer-based approaches and traditional Convolutional Neural Network (CNN) models, Yang aims to revolutionize the field of COVID-19 diagnosis. Focusing on attention analysis techniques and performance metrics, his research seeks to provide insights into the efficacy of Swin Transformer architectures in improving diagnostic accuracy and efficiency. Through innovative methodologies and deep learning techniques, this study contributes to advancements in medical imaging and disease detection methodologies, with significant implications for the diagnosis and management of COVID-19 cases.

5) **“Brain Tumor Segmentation in Fluid-Attenuated Inversion Recovery Brain MRI using Residual Network Deep Learning Architectures’’ [5]**

Mahyoub, Natalia, Sudirman, Al-Jumaily, and Liatsis investigate the application of deep learning architectures for brain tumor segmentation in Fluid-Attenuated Inversion Recovery (FLAIR) brain MRI scans in their study titled "Brain Tumor Segmentation in Fluid-Attenuated Inversion Recovery Brain MRI using Residual Network Deep Learning Architectures." Presented at the 2023 15th International Conference on Developments in eSystems Engineering (DeSE) in Baghdad & Anbar, Iraq, their research focuses on leveraging residual network deep learning architectures to accurately segment brain tumors from MRI images. By employing techniques such as transfer learning and hyperparameter optimization, the study aims to improve the performance and robustness of the segmentation model. Through their investigation, Mahyoub et al. contribute to the advancement of medical image analysis techniques, particularly in the domain of brain tumor detection and segmentation, thereby facilitating early diagnosis and treatment planning for patients with neurological conditions.

6) **‘‘Automated Diagnosis of Brain Tumor Based on Deep Learning Feature Fusion Using MRI Images’’ [6]**

Durga, Muduli, Rahul, Naidu, Kumar, and Sharma present their work on "Automated Diagnosis of Brain Tumor Based on Deep Learning Feature Fusion Using MRI Images" at the 2023 IEEE 3rd International Conference on Applied Electromagnetics, Signal Processing, & Communication (AESPC) in Bhubaneswar, India. In this study, the authors focus on leveraging deep learning techniques for the automated diagnosis of brain tumors using MRI images. Their approach involves feature fusion using convolutional neural networks (CNNs), specifically InceptionV3 and VGG19 architectures, to extract relevant features from MRI scans. By employing machine learning methodologies, Durga et al. aim to enhance the accuracy and efficiency of brain tumor diagnosis, thereby contributing to the field of medical imaging and diagnostic techniques. This research holds significant potential for improving patient care and outcomes by facilitating early detection and treatment planning for individuals with brain tumors.

7) ‘‘A Hybrid Neural Network Model for Brain Tumor Detection in Brain MRI Images’’ [7]

Hossain et al. present their research on "A Hybrid Neural Network Model for Brain Tumor Detection in Brain MRI Images" at the 2022 IEEE 13th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON) in Vancouver, BC, Canada. The study focuses on developing a hybrid neural network model for detecting brain tumors in MRI images. By integrating multiple deep learning architectures such as convolutional neural networks (CNNs) and long short-term memory (LSTM) networks, along with traditional machine learning techniques like support vector machines (SVMs), the proposed model aims to improve the accuracy and robustness of brain tumor detection. Through the utilization of advanced medical imaging and deep learning methodologies, Hossain et al. contribute to the field of medical diagnostics, particularly in the domain of brain tumor detection using MRI imaging modalities. This research has significant implications for improving patient outcomes by enabling early and accurate diagnosis of brain tumors.

8) ‘‘A New Deep CNN for Brain Tumor Classification’’ [8]

Ayadi, Elhamzi, and Atri introduce "A New Deep CNN for Brain Tumor Classification" at the 2020 20th International Conference on Sciences and Techniques of Automatic Control and Computer Engineering (STA) in Monastir, Tunisia. Their study focuses on developing a novel deep convolutional neural network (CNN) specifically designed for the classification of brain tumors. By leveraging magnetic resonance imaging (MRI) data, the proposed CNN aims to accurately classify brain tumor types, facilitating computer-aided diagnosis and treatment planning. Through the integration of advanced deep learning techniques and medical imaging modalities, Ayadi et al. contribute to the field of computational modeling and computer-aided diagnosis, particularly in the domain of brain tumor classification using MRI scans. This research holds promise for improving the efficiency and accuracy of brain tumor diagnosis, thereby enhancing patient care and outcomes.

9) ‘‘A Hybrid CNN-LSTM Network For Brain Tumor Classification Using Transfer Learning’’ [9]

Rajeev, Rajasekaran, Ramaraj, Vishnuvarthanan, Arunprasath, and Muneeswaran present their work on "A Hybrid CNN-LSTM Network For Brain Tumor Classification Using Transfer

Learning" at the 2023 9th International Conference on Smart Computing and Communications (ICSCC) in Kochi, Kerala, India. Their study focuses on the development of a hybrid convolutional neural network (CNN) and long short-term memory (LSTM) network for the classification of brain tumors using transfer learning. By leveraging transfer learning techniques and magnetic resonance imaging (MRI) data, the proposed model aims to accurately classify brain tumor types. Through the integration of CNNs for feature extraction and LSTM networks for sequence modeling, the hybrid architecture offers improved performance in brain tumor classification tasks. This research contributes to the advancement of deep learning methodologies in medical imaging and computer-aided diagnosis, particularly in the domain of brain tumor classification using MRI scans. The utilization of transfer learning further enhances the efficiency and effectiveness of the proposed model, offering potential benefits for clinical practice and patient care.

10) ‘‘IIMFCBM: Intelligent Integrated Model for Feature Extraction and Classification of Brain Tumors Using MRI Clinical Imaging Data in IoT-Healthcare’’ [10]

Haq et al. present their work on "IIMFCBM: Intelligent Integrated Model for Feature Extraction and Classification of Brain Tumors Using MRI Clinical Imaging Data in IoT-Healthcare" in the IEEE Journal of Biomedical and Health Informatics. Published in October 2022, their research introduces an intelligent integrated model for the extraction and classification of features from MRI clinical imaging data of brain tumors in the context of IoT healthcare. The proposed model, IIMFCBM, integrates various techniques including convolutional neural networks (CNNs), long short-term memory (LSTM), and data augmentation to effectively extract features and classify brain tumors from MRI data. By leveraging advancements in medical imaging and the integration of IoT technologies, Haq et al. aim to improve the efficiency and accuracy of brain tumor diagnosis and treatment. This research contributes to the field of healthcare informatics by offering a comprehensive approach to brain tumor classification, facilitating early detection and personalized treatment strategies for patients.

11) ‘‘Brain Tumor Detection Using Machine Learning’’ [11]

Kushwaha, Rajput, Aggrawal, Dwivedi, Srivastava, and Singh present their work on "Brain Tumor Detection Using Machine Learning" at the 2023 6th International Conference on Contemporary Computing and Informatics (IC3I) in Gautam Buddha Nagar, India. Their study

focuses on the application of machine learning algorithms for the detection of brain tumors. By leveraging machine learning techniques, including clustering algorithms and prediction algorithms, the authors aim to develop an effective method for identifying brain tumors from medical diagnostic imaging data. This research contributes to the field of medical imaging and diagnostic methodologies by harnessing the power of artificial intelligence and machine learning to facilitate early detection and intervention for patients with brain tumors. The utilization of machine learning algorithms offers potential benefits for improving diagnostic accuracy and patient outcomes in the context of brain tumor detection.

12) ‘‘Brain Tumor Prediction with Deep Learning and Tumor Volume Calculation’’ [12]

Karayeğen and Akşahin present their research on "Brain Tumor Prediction with Deep Learning and Tumor Volume Calculation" at the 2021 Medical Technologies Congress (TIPTEKNO) in Antalya, Turkey. Their study focuses on the prediction of brain tumors using deep learning techniques, coupled with tumor volume calculation. By leveraging deep learning methodologies, including semantic segmentation and 3D imaging, the authors aim to accurately predict the presence of brain tumors from medical imaging data. Additionally, they propose a method for calculating the volume of detected tumors, which can provide valuable information for treatment planning and prognosis assessment. This research contributes to the field of medical imaging and diagnostic technologies by offering an advanced approach to brain tumor prediction and volume estimation, facilitating improved patient care and treatment outcomes. The integration of deep learning with tumor volume calculation enhances the efficiency and accuracy of brain tumor diagnosis, offering potential benefits for clinical practice and patient management.

13) ‘‘SwinLSTM: Improving Spatiotemporal Prediction Accuracy using Swin Transformer and LSTM’’ [13]

Tang, Li, Zhang, and Tang introduce "SwinLSTM: Improving Spatiotemporal Prediction Accuracy using Swin Transformer and LSTM" at the 2023 IEEE/CVF International Conference on Computer Vision (ICCV) in Paris, France. Their study focuses on enhancing spatiotemporal prediction accuracy by integrating Swin Transformer and Long Short-Term Memory (LSTM) networks. By leveraging the capabilities of both architectures, the authors aim to improve the efficiency and effectiveness of predictive models in the domain of computer vision. This research contributes to advancing predictive modeling techniques, particularly in

spatiotemporal data analysis, by integrating state-of-the-art transformer-based approaches with recurrent neural networks. The integration of Swin Transformer and LSTM offers potential benefits for various applications, including video analysis, motion prediction, and action recognition, by capturing both spatial and temporal dependencies in the data. By presenting SwinLSTM, Tang et al. provide a novel approach to improving prediction accuracy in spatiotemporal domains, facilitating advancements in computer vision research and applications.

14) ‘‘SwinT-Unet: Hybrid architecture for Medical Image Segmentation Based on Swin transformer block and Dual-Scale Information’’ [14]

Atek, Mehidi, Jabri, and Belkhiat present their research on "SwinT-Unet: Hybrid architecture for Medical Image Segmentation Based on Swin transformer block and Dual-Scale Information" at the 2022 7th International Conference on Image and Signal Processing and their Applications (ISPA) in Mostaganem, Algeria. Their study introduces SwinT-Unet, a hybrid architecture designed for medical image segmentation. Leveraging the Swin Transformer block and dual-scale information, the proposed model aims to enhance the accuracy and efficiency of segmentation tasks in medical imaging. By integrating state-of-the-art deep learning techniques, including Swin Transformer-based approaches and dual-scale information processing, Atek et al. contribute to advancements in medical image analysis methodologies. This research holds promise for improving the accuracy and reliability of medical image segmentation, thereby facilitating more precise diagnoses and treatment planning in various clinical applications. The integration of SwinT-Unet offers potential benefits for medical practitioners and researchers seeking improved segmentation performance in medical imaging tasks.

15) “SUNet: Swin Transformer UNet for Image Denoising’’ [15]

Fan, Liu, and Liu present their research on "SUNet: Swin Transformer UNet for Image Denoising" at the 2022 IEEE International Symposium on Circuits and Systems (ISCAS) in Austin, TX, USA. Their study introduces SUNet, a novel architecture designed for image denoising. Leveraging the Swin Transformer and UNet architectures, SUNet aims to effectively reduce noise in images while preserving important features and details. By integrating the capabilities of Swin Transformer and UNet, the proposed model offers an advanced approach to image restoration and denoising tasks. This research contributes to the field of image

processing and computational modeling by providing a state-of-the-art solution for noise reduction in images. The integration of SUNet holds promise for improving the quality of images in various applications, including medical imaging, surveillance, and photography. The utilization of Swin Transformer UNet architecture offers potential benefits for researchers and practitioners seeking enhanced image denoising performance.

16) "Unrestricted Attention May Not Be All You Need–Masked Attention Mechanism Focuses Better on Relevant Parts in Aspect-Based Sentiment Analysis" [16]

Feng, Zhang, and Song explore the effectiveness of attention mechanisms in aspect-based sentiment analysis in their paper titled "Unrestricted Attention May Not Be All You Need–Masked Attention Mechanism Focuses Better on Relevant Parts in Aspect-Based Sentiment Analysis," published in IEEE Access. Their study investigates how the masked attention mechanism can improve the focus on relevant parts in aspect-based sentiment analysis compared to unrestricted attention. By analyzing the performance of pre-trained language models and attention mechanisms, the authors highlight the importance of attention mechanisms in capturing relevant semantic information for sentiment analysis tasks. This research contributes to the field of natural language processing by providing insights into the role of attention mechanisms in sentiment analysis and the potential benefits of using masked attention mechanisms for focusing on relevant aspects. The findings of this study have implications for improving the accuracy and efficiency of sentiment analysis systems, particularly in applications where understanding specific aspects of sentiment is crucial.

17) ''All the attention you need: Global-local, spatial-channel attention for image retrieval'' [17]

Song, Han, and Avrithis present their research on "All the attention you need: Global-local, spatial-channel attention for image retrieval" at the 2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV) in Waikoloa, HI, USA. Their study introduces a novel attention mechanism for image retrieval tasks, termed global-local spatial-channel attention. By incorporating both global and local spatial-channel attention mechanisms, the proposed model aims to enhance the representation learning process and improve the accuracy of image retrieval systems. This research contributes to advancements in representation learning and deep learning methodologies for image retrieval, offering a comprehensive approach to capturing spatial and channel-wise dependencies in image data. The integration of

global-local spatial-channel attention holds promise for improving the performance of image retrieval pipelines, thereby facilitating more accurate and efficient image and video indexing and retrieval systems. The findings of this study provide valuable insights into the design of attention mechanisms for computer vision tasks, paving the way for further advancements in image retrieval research.

18) ‘‘ALAN: Self-Attention Is Not All You Need for Image Super-Resolution’’ [18]

Chen, Qin, and Wen present their research on "ALAN: Self-Attention Is Not All You Need for Image Super-Resolution" in IEEE Signal Processing Letters. Their study challenges the notion that self-attention is the sole requirement for effective image super-resolution. Introducing ALAN (Asymmetric Convolution and Attention Network), the authors propose a novel approach that combines asymmetric convolution and attention mechanisms for image super-resolution tasks. By incorporating structural re-parameterization and a stage-to-block design paradigm, ALAN aims to improve feature extraction and reconstruction performance while mitigating computational complexity. This research contributes to advancing the field of image super-resolution by offering an alternative framework that goes beyond traditional self-attention mechanisms. The integration of asymmetric convolution and attention networks in ALAN offers promising results for enhancing image reconstruction quality and reducing computational overhead. The findings of this study have implications for improving the efficiency and effectiveness of image super-resolution techniques, paving the way for further advancements in signal processing and computer vision applications.

19) ‘‘STM-UNet: An Efficient U-shaped Architecture Based on Swin Transformer and Multiscale MLP for Medical Image Segmentation’’ [19]

Shi, Gao, Zhang, and Zhang introduce "STM-UNet: An Efficient U-shaped Architecture Based on Swin Transformer and Multiscale MLP for Medical Image Segmentation" at the GLOBECOM 2023 - IEEE Global Communications Conference in Kuala Lumpur, Malaysia. Their study presents STM-UNet, a novel U-shaped architecture designed for efficient medical image segmentation. By leveraging Swin Transformer and Multiscale Multilayer Perceptron (MLP) architectures, STM-UNet aims to improve the accuracy and efficiency of medical image segmentation tasks. The integration of Swin Transformer facilitates effective feature extraction, while the multiscale MLP enhances feature representation and classification. Additionally, the parallel convolution mechanism further enhances the segmentation performance. This research

contributes to advancing medical image segmentation methodologies by offering an efficient and effective architecture that leverages state-of-the-art deep learning techniques. The STM-UNet model holds promise for improving diagnostic accuracy and streamlining medical image analysis workflows in clinical settings. The findings of this study provide valuable insights into the design of deep learning architectures for medical image segmentation, facilitating advancements in healthcare and diagnostic imaging technologies.

20) ‘‘Swin Transformer with Local Aggregation’’ [20]

Chen, Bai, Cheng, and Wu present their research on "Swin Transformer with Local Aggregation" at the 2022 3rd International Conference on Information Science, Parallel and Distributed Systems (ISPDS) in Guangzhou, China. Their study introduces a novel approach that enhances Swin Transformer models through the incorporation of local aggregation techniques. By integrating local aggregation mechanisms into Swin Transformer architectures, the proposed model aims to improve feature representation and capture local contextual information more effectively. This research contributes to advancing the capabilities of Transformer-based architectures in handling diverse data types and tasks. The integration of local aggregation techniques offers potential benefits for various applications, including image classification, object detection, and natural language processing. The findings of this study provide valuable insights into optimizing Transformer models for enhanced performance and efficiency in information science and parallel and distributed systems.

CHAPTER 3
SYSTEM DESIGN

3.1 SYSTEM WORKFLOW

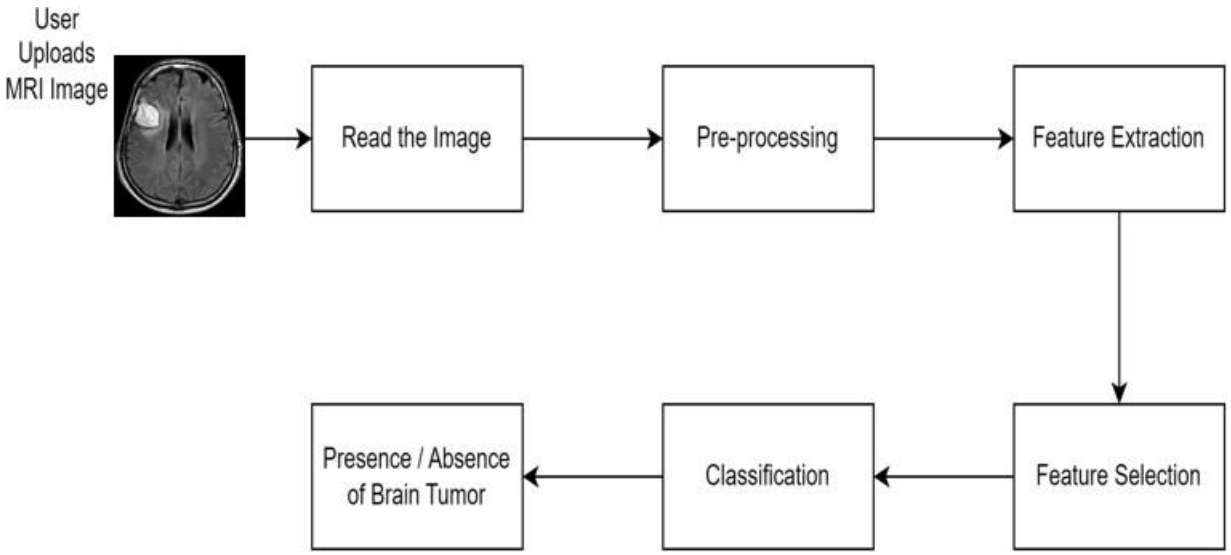

Figure 3.1. System Workflow

Our proposed method combines the SWIN Transformer model with the LSTM model to classify MRI images into four distinct classes: Meningioma, Glioma, Pituitary, and Non-Tumor. This approach aims to enhance both accuracy and robustness, especially in the realm of medical image analysis, where precise classification is critical for disease prediction and treatment planning. To achieve this, we introduce a novel strategy leveraging LSTM for feature extraction and SWIN Transformers for accurate classification of MRI images into their respective categories.

The system workflow initiates with data acquisition, involving the retrieval of MRI images from the Kaggle dataset. Subsequently, the images undergo pre-processing utilizing various augmentation techniques. Features are then extracted utilizing the Long Short-Term Memory (LSTM) network, followed by classification into the four designated classes. Finally, the model undergoes evaluation and subsequent deployment for practical application.

3.2 METHODOLOGY

3.2.1 DATA ACQUSITION

Magnetic Resonance Imaging (MRI) stands as a crucial diagnostic tool in medicine, particularly for detecting brain tumors. The efficacy of any machine learning or deep learning model in MRI brain tumor detection hinges greatly on the quality and diversity of the dataset utilized for training and evaluation. In this section, we delve into the process of acquiring data for our project, with a specific focus on the Brain MRI Images Dataset (BrATS) sourced from Kaggle.

The Brain MRI Images Dataset (BrATS) holds significant prominence in brain tumor segmentation and classification endeavors. It comprises a compilation of MRI brain images sourced from diverse medical institutions and research centers. Accessible on Kaggle, the dataset encompasses 7529 MRI brain images, each meticulously labeled into one of four categories: Meningioma, Glioma, Pituitary, and Non-Tumor. These categories represent various brain tumor types as well as normal brain tissue.

The BrATS dataset offers a comprehensive portrayal of brain MRI images, encompassing a broad spectrum of tumor variations and imaging scenarios. Each MRI image within the dataset comes with detailed metadata, including patient demographics, imaging parameters, and tumor attributes. This wealth of information proves invaluable in comprehending the inherent biological diversity and clinical significance encapsulated within the dataset.

It is crucial to note that the use of medical imaging datasets, such as BrATS, raises ethical considerations regarding patient privacy and data usage rights. Proper anonymization procedures are followed to remove any identifying information from the MRI images, ensuring patient confidentiality. Additionally, strict adherence to data usage agreements and institutional review board (IRB) guidelines is maintained to ensure ethical compliance and integrity in research practices.

The acquisition of the BrATS dataset serves as a critical foundation for our project on MRI brain tumor detection. By leveraging this comprehensive dataset, we aim

to develop and evaluate a state-of-the-art deep learning model capable of accurately detecting and classifying brain tumors from MRI images. Through rigorous data preprocessing and adherence to ethical standards, we ensure the integrity and reliability of our research findings, ultimately contributing to advancements in medical imaging and healthcare.

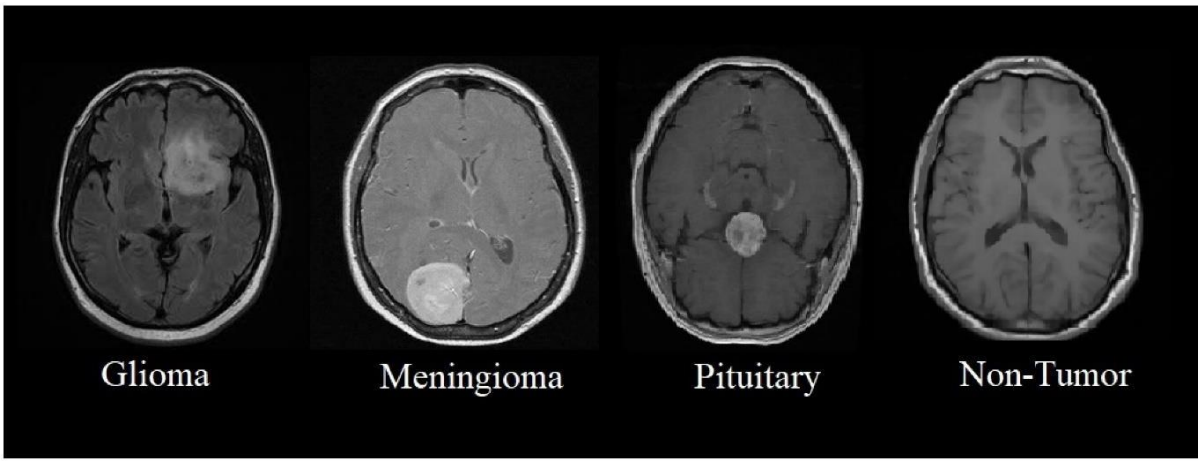

Figure 3.2.1.1. Types of Tumor

3.2.2 DATA PREPROCESSING

In preparing our dataset for training the classification model, we underwent a series of data preprocessing steps aimed at fortifying its resilience and enhancing its ability to generalize. Given that MRI brain tumor detection heavily relies on the quality and coherence of the dataset used for training machine learning models, data preprocessing assumes a pivotal role in bolstering dataset quality and subsequently elevating model performance. In this section, we delve into the array of data preprocessing techniques employed in our project, notably focusing on data augmentation and noise removal.

Data augmentation entails the application of various transformations to the existing dataset, thereby augmenting its size and diversifying its content. This technique proves especially beneficial when dealing with limited training data, as it aids in enhancing the model's resilience and capacity to generalize. Within our project, we implement a range of transformations on the MRI brain images, including:

- **Rotation:** Altering the images by rotating them at specific angles (e.g., 90 degrees, 180 degrees) to introduce variations in orientation and perspective. This aids the model in learning to detect tumors from various viewpoints, thereby improving its capability to generalize to unfamiliar data.

- **Zooming:** Zooming in or out of the images to simulate different levels of magnification and scale. By varying the zoom level, we expose the model to images with different levels of detail, helping it learn to detect tumors of various sizes.

- **Brightness Adjustment:** Modifying the brightness and contrast of the images to mimic variations in imaging conditions and scanner settings. This helps the model become more robust to differences in illumination and contrast across different MRI scans.

By augmenting the training dataset with transformed images, our objective is to expose the model to a wider array of scenarios and enhance its capacity to generalize to unseen data.

Noise removal serves as a pivotal preprocessing step aimed at refining the quality and clarity of MRI brain images. These images often contend with diverse noise types, including random noise introduced during the imaging process. To mitigate this challenge, we utilize the following noise removal techniques:

- **Gaussian Filtering:** Employing Gaussian smoothing to diminish high-frequency noise and enhance image clarity. This method effectively attenuates noise while conserving the sharpness of edges and essential features within the images.

- **Median Filtering:** Leveraging median filtering to eliminate salt-and-pepper noise, prevalent in MRI images. Through median filtering, each pixel's value is replaced with the median value of its surrounding neighborhood, effectively eradicating outliers while preserving intricate image details.

- **Cropping Brain Contour Area:** Concentrating on the region of interest (ROI) involves cropping the brain contour area to eliminate extraneous background noise

and artifacts. This process enables the model to prioritize pertinent information for tumor detection while alleviating the computational load during training.

By eliminating noise from the MRI images, our objective is to heighten the visibility of crucial features and structures, facilitating the model's extraction of meaningful information for tumor detection.

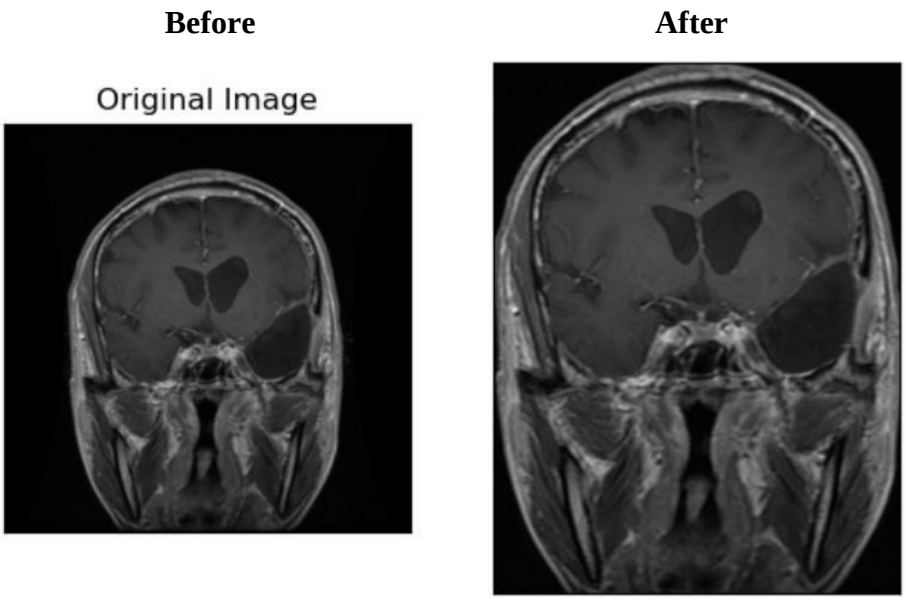

Figure 3.2.2.1. MRI image before and after pre-processing

3.2.3 FEATURE SELECTION USING LSTM

Feature extraction and selection stand as pivotal stages in MRI brain tumor detection, aimed at discerning the most pertinent information from input images while reducing dimensionality for streamlined processing. In our project, we leverage Long Short-Term Memory (LSTM) networks for both feature extraction and selection, capitalizing on their proficiency in capturing temporal dependencies within sequential data such as MRI images.

LSTM networks represent a subtype of recurrent neural network (RNN) architecture meticulously crafted to analyze and forecast sequences of data. Comprising recurrent units termed cells, these networks uphold an internal state, thereby enabling them to apprehend long-range dependencies inherent in sequential data. This attribute

renders them notably adept for tasks entailing temporal modeling and sequential pattern recognition.

To extract features from the MRI brain images using LSTM, we follow these steps:

- **Sequence Representation:** Each MRI brain image is converted into a sequence of vectors representing its pixel intensities. This sequence is treated as a time series data, with each timestep corresponding to a pixel value in the image.

- **LSTM Architecture:** We design an LSTM architecture tailored for feature extraction from MRI images. The LSTM network typically consists of multiple LSTM layers followed by a pooling layer or a fully connected layer for feature aggregation.

- **Training:** The LSTM network is trained using sequences of MRI images, where each sequence represents a sample in the training dataset. During training, the network learns to extract discriminative features from the input sequences, capturing spatial and temporal patterns indicative of different tumor classes.

- **Feature Representation:** The output of the LSTM network at the final timestep or at specific intermediate timesteps is considered as the extracted feature representation of the input MRI image. These feature representations capture important characteristics of the images, such as texture, shape, and spatial distribution of pixel intensities.

To select features from the MRI brain images, we follow these steps:

- **Dimensionality Reduction:** The feature vectors extracted by the LSTM network may contain redundant or irrelevant information, leading to increased computational complexity and potential overfitting. To address this issue, we perform feature selection to reduce the dimensionality of the feature space while preserving the most informative features.

- **Filter Methods:** We employ filter-based feature selection methods such as correlation analysis or statistical tests to evaluate the relevance of each feature to the target variable (tumor class labels). Features that exhibit high correlation or statistical significance with the target variable are retained, while others are discarded.

- **Wrapper Methods:** In addition to filter methods, we may also use wrapper-based feature selection techniques such as recursive feature elimination (RFE) or forward/backward feature selection. These methods iteratively evaluate subsets of features based on their performance in a predictive model (e.g., LSTM classifier) and select the subset that maximizes predictive accuracy.

- **Cross-Validation:** To ensure the generalization ability of the selected features, we perform cross-validation to evaluate the performance of the feature selection process on independent validation datasets. This helps to identify and mitigate the risk of overfitting or selecting features that are specific to the training dataset.

By employing LSTM networks for feature extraction and selection, we aim to capture both spatial and temporal patterns in MRI brain images and identify the most discriminative features for tumor detection. This approach enables us to build more efficient and accurate classification models by focusing on the most relevant information in the input data.

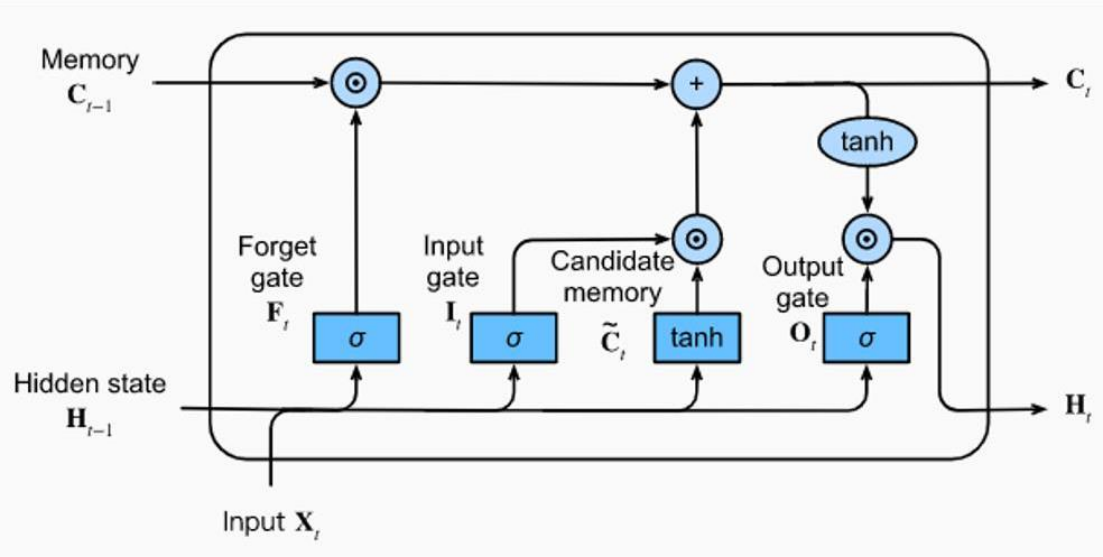

Figure 3.2.3.1. LSTM Architecture

3.2.4 CLASSIFICATION USING SWIN TRANSFORMER

Following the feature extraction process from MRI brain images using LSTM, the subsequent phase involves classifying these features into distinct tumor classes. In our project, we opt for the SWIN Transformer architecture for classification, renowned for its state-of-the-art performance across various image classification tasks.

The SWIN Transformer, an innovative hierarchical transformer-based vision model introduced by Liu et al., is engineered to efficiently capture spatial relationships within images. Diverging from conventional convolutional neural networks (CNNs) operating on fixed-size grids of feature maps, the SWIN Transformer adopts a hierarchical approach, allowing seamless handling of images of varying sizes with minimal computational overhead.

To execute classification utilizing the SWIN Transformer, we adhere to the following steps:

- **Model Architecture:** We tailor the SWIN Transformer architecture for image classification by adjusting the input and output layers to align with the dimensions of the feature vectors extracted by the LSTM network. Comprising multiple layers of self-attention and feedforward networks, the SWIN Transformer excels in capturing long-range dependencies and extracting high-level features from the input data.

- **Training:** The training phase entails feeding the SWIN Transformer model with the extracted feature vectors from MRI images as input and the corresponding tumor labels as targets. Throughout training, the model learns to correlate input features with the correct tumor class labels, leveraging a classification loss function like cross-entropy loss.

- **Evaluation:** Post-training, we evaluate the SWIN Transformer model's performance on a distinct validation dataset to gauge its classification accuracy

and generalization capability. Metrics such as accuracy, precision, recall, and F1-score are employed to quantify the model's efficacy across diverse tumor classes.

- **Inference:** Subsequently, we deploy the trained SWIN Transformer model for inference on unseen MRI images. The model ingests the extracted features and generates a probability distribution across different tumor classes. The class with the highest probability is designated as the predicted tumor class for each input image.

By harnessing the SWIN Transformer architecture for classification, our aim is to leverage its prowess in capturing spatial relationships within images and extracting high-level features, ultimately yielding more precise and resilient tumor classification outcomes compared to conventional classification models.

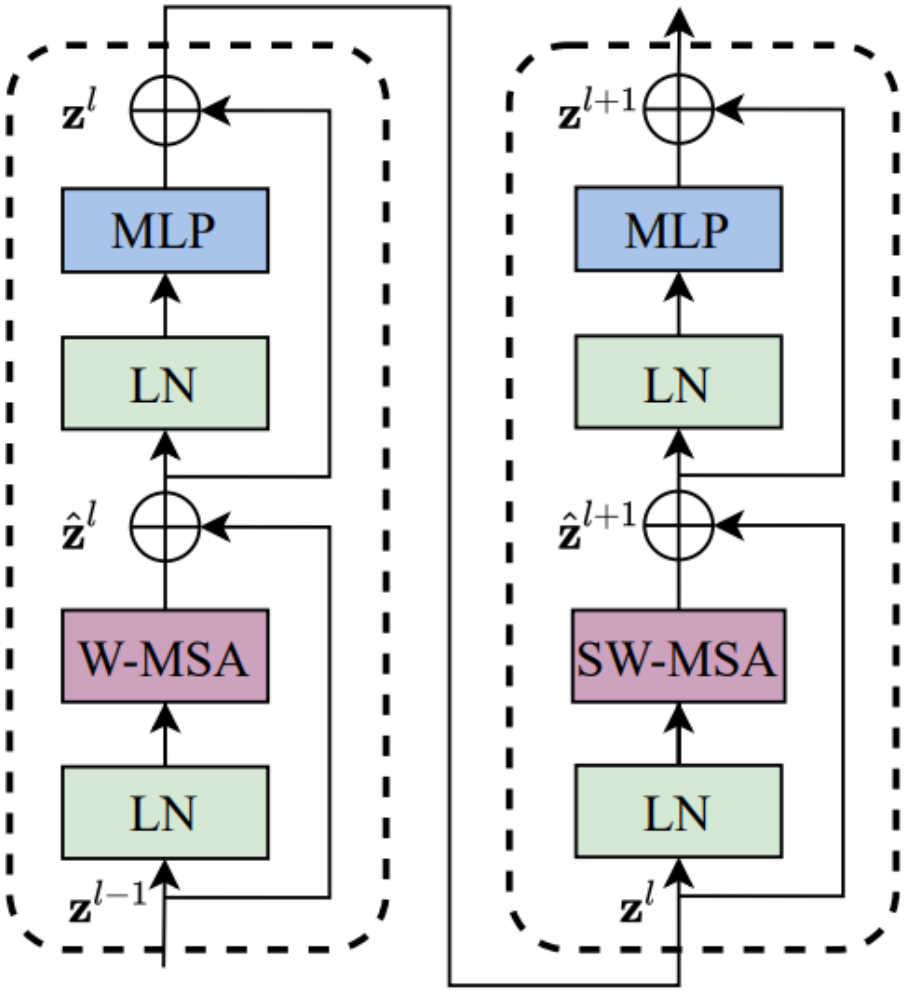

Figure 3.2.4.1. SWIN Transformer Blocks

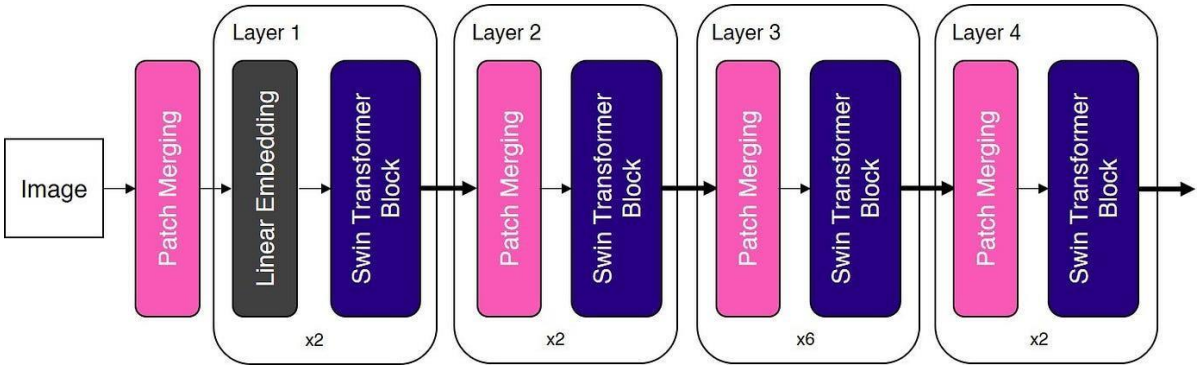

Figure 3.2.4.2. Architecture of SWIN Transformer

CHAPTER 4
SYSTEM STUDY

4.1 Long Short-Term Memory (LSTM) Networks

Long Short-Term Memory (LSTM) networks represent a form of recurrent neural network (RNN) architecture devised to surmount the limitations of traditional RNNs in capturing long-term dependencies within sequential data. Renowned for their proficiency, LSTMs find extensive application across diverse domains such as natural language processing, speech recognition, and time series analysis owing to their adeptness in learning and retaining information over prolonged time intervals. Originally introduced by Hochreiter & Schmidhuber in 1997, LSTMs have undergone refinement and widespread adoption by numerous researchers in subsequent work, emerging as a staple technique in contemporary machine learning endeavors.

In contrast to conventional RNNs, which often struggle with the challenge of retaining information over extended sequences, LSTMs excel in circumventing this issue through their intrinsic design. Central to LSTMs is the concept of the cell state, depicted as a horizontal line traversing the top of the architectural diagram. Functioning akin to a conveyor belt, the cell state spans the entirety of the network, facilitating seamless information flow with minimal alteration. Notably, LSTMs feature gates, specialized structures tasked with regulating the influx and efflux of information within the cell state. Comprising sigmoid neural net layers and pointwise multiplication operations, these gates enable selective passage of information, thereby affording meticulous control over the cell state.

The advent of LSTMs marked a significant advancement in the realm of RNNs, revolutionizing the scope of tasks achievable with sequential data. By empowering each step of an RNN to discern relevant information from a broader collection, LSTMs enable enhanced performance across various applications. For instance, in image captioning tasks, an RNN equipped with LSTM units might selectively focus on distinct portions of the image for each word it generates. Indeed, researchers such

as Xu et al. (2015) have explored this concept through mechanisms like attention, showcasing its potential for further innovation and breakthroughs in the field.

Architecture:

The LSTM architecture consists of several key components, each responsible for different aspects of memory and information flow:

- **Forget Gate (ft):** The forget gate plays a crucial role in determining which information from the previous cell state ought to be discarded or overlooked. It takes inputs from the previous hidden state (ht-1) and the current input (xt), applies a sigmoid activation function, and yields a forget gate vector (ft) to regulate the retention or deletion of information.

- **Input (Update) Gate (it):** The input gate governs the influx of new information into the cell state, dictating which values should be updated based on the present input and the prior hidden state. Employing a sigmoid activation function, akin to the forget gate, it generates an input gate vector (it) to manage the incorporation of fresh information.

- **Candidate Cell State (□ct):** The candidate cell state calculates potential new values that could augment the cell state. It amalgamates the current input with the previous hidden state and applies the hyperbolic tangent (tanh) activation function to derive candidate values.

- **Update Cell State (ct):** This operation merges the outputs of the forget gate (ft) and the input gate (it) with the candidate cell state (μct) to update the current cell state (ct). By selectively integrating information based on the forget gate and input gate outputs, it orchestrates the refinement of the cell state.

- **Output Gate (ot):** The output gate dictates which segments of the cell state should be exposed as the output. It takes inputs from the previous hidden state (ht-1) and the current input (xt), applies a sigmoid activation function, and generates an output gate vector (ot) to regulate the dissemination of information.

- **Hidden State (ht):** Serving as the LSTM cell's output, the hidden state is computed by applying the output gate (ot) to the hyperbolic tangent of the updated cell state (ct). Endowed with information crucial for subsequent time steps or serving as the final output of the LSTM network, the hidden state plays a pivotal role in information propagation.

Characteristics and Features:

- **Long-Term Dependencies:** LSTMs are capable of capturing long-term dependencies in sequential data, making them suitable for tasks requiring memory over extended time intervals.

- **Gating Mechanisms:** The use of forget gates, input gates, and output gates allows LSTMs to selectively update and expose information, enabling better control over the flow of information through the network.

- **Non-Linear Activation:** LSTMs employ non-linear activation functions such as sigmoid and hyperbolic tangent, enabling them to model complex relationships and capture non-linear patterns in the data.

Advantages:

- **Memory Retention:** LSTMs can remember information over long sequences, making them suitable for tasks involving long-range dependencies.

- **Robustness to Vanishing Gradient Problem:** LSTMs mitigate the vanishing gradient problem commonly encountered in traditional RNNs by using gating mechanisms to regulate the flow of gradients during backpropagation.

- **Versatility:** LSTMs are versatile and can be applied to various sequential data types, including text, audio, and time series data.

Disadvantages and Limitations:

- **Computational Complexity:** LSTMs are computationally intensive, particularly when dealing with large input sequences or deep architectures, which can lead to longer training times and higher resource requirements.

- **Overfitting:** LSTMs are prone to overfitting, especially when trained on small datasets or when the model architecture is overly complex. Regularization techniques such as dropout and weight decay may be necessary to mitigate this issue.

- **Interpretability:** The internal workings of LSTMs can be challenging to interpret, making it difficult to understand how the model arrives at its predictions and potentially limiting its applicability in domains where interpretability is crucial.

Overall, Long Short-Term Memory (LSTM) networks are powerful and versatile architectures for modeling sequential data, offering the ability to capture long-term dependencies and learn complex patterns over extended time intervals. Despite their computational complexity and potential for overfitting, LSTMs remain widely used and continue to be a cornerstone of many state-of-the-art machine learning models.

4.2 SWIN Transformer

The Swin Transformer stands out as a hierarchical transformer-based architecture renowned for its exceptional performance across various computer vision tasks such as image classification, object detection, and semantic segmentation. Originated by Liu et al., the Swin Transformer introduces a pioneering hierarchical design capable of efficiently processing high-resolution images comprising hundreds of millions of pixels. By harnessing self-attention mechanisms and hierarchical representations, the Swin Transformer attains remarkable performance in terms of both accuracy and efficiency, rendering it highly adaptable to a diverse array of vision tasks.

Distinctive to the Swin Transformer is its construction, wherein the standard multi-head self-attention (MSA) module within a Transformer block is replaced by a module predicated on shifted windows, while preserving the remaining layers. Each Swin

Transformer block comprises a shifted window-based MSA module, succeeded by a two-layer MLP featuring GELU nonlinearity in between. Furthermore, LayerNorm (LN) layers are applied before each MSA module and MLP, and residual connections are integrated after each module to facilitate information flow.

Architecture:

The Swin Transformer architecture consists of several key components, each responsible for different aspects of feature extraction and aggregation:

- **Patch Partitioning:** The input image is divided into non-overlapping patches, which are treated as independent tokens and processed by the transformer layers. This patch-based processing enables efficient handling of high-resolution images without the need for excessive computational resources.

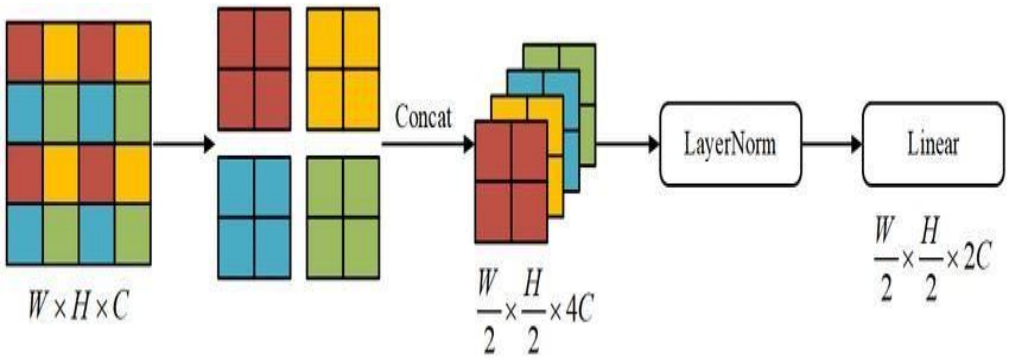

Figure 4.2.1. Patch Partitioning

- **Multi-Scale Transformers:** Swin Transformer employs a hierarchical design with multiple layers of transformers operating at different spatial resolutions. Each transformer layer processes a specific scale of features and aggregates information across scales through cross-layer connections.

- **Shifted Windows:** To capture spatial relationships effectively, Swin Transformer introduces shifted windows in self-attention mechanisms, allowing each token to attend to neighboring tokens in a shifted manner. This enables the model to capture both local and global context without requiring quadratic computation.

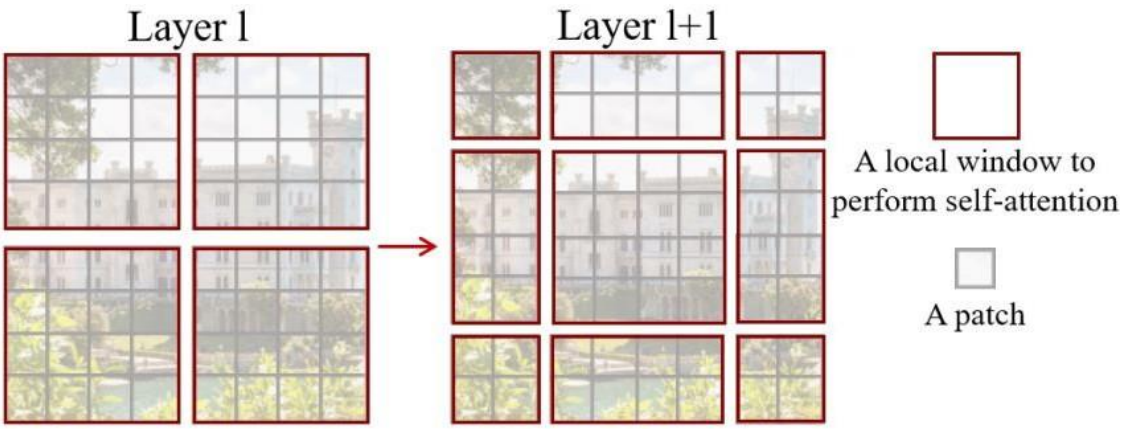

Figure 4.2.2. Shifted Windows

- **Tokenization and Positional Encoding:** Swin Transformer tokenizes input patches and encodes their spatial positions using learnable embeddings. This enables the model to maintain spatial information throughout the processing pipeline and attend to relevant spatial contexts during feature extraction.

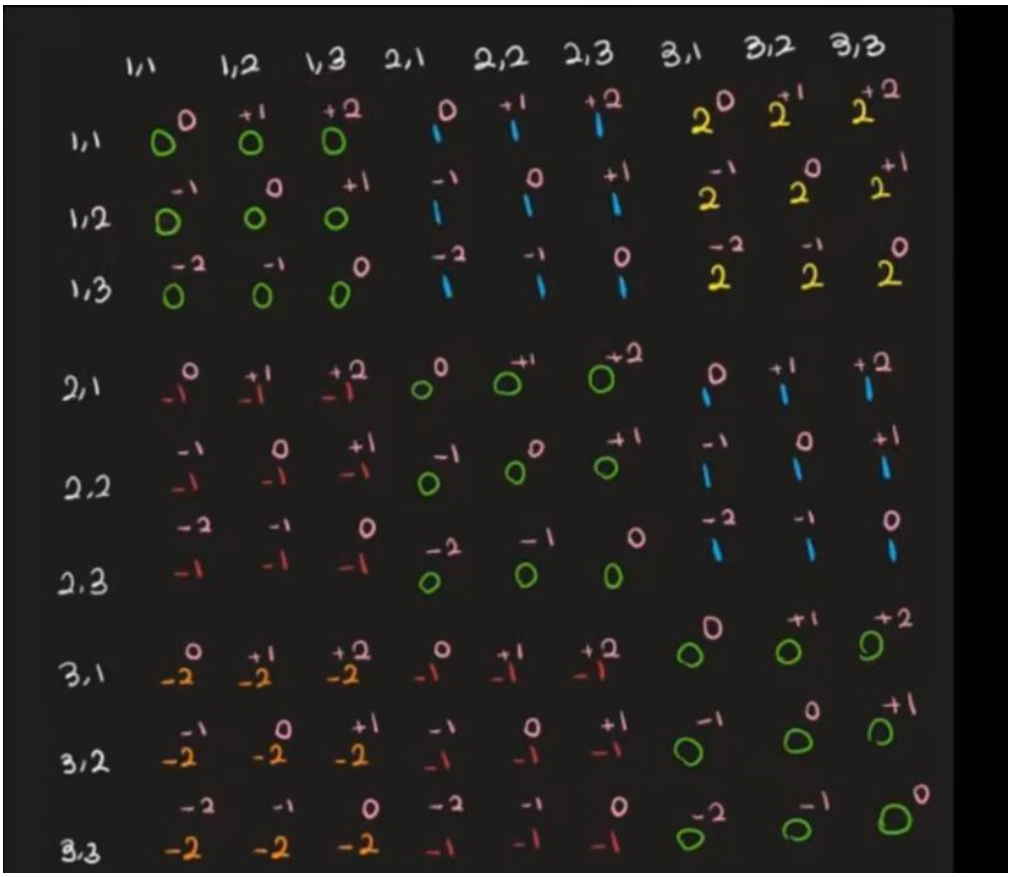

Figure 4.2.3. Relative Position Embedding

- **Window-based Multi-Head Self-Attention (W-MSA):** The W-MSA mechanism plays a pivotal role in Swin Transformer's architecture, facilitating efficient processing of high-resolution images. By partitioning the input image into non-overlapping windows and applying self-attention operations within each window, Swin Transformer can capture fine-grained spatial relationships while maintaining computational efficiency. Moreover, the use of multi-head attention enables the model to attend to different parts of the image simultaneously, enhancing its ability to extract meaningful features.

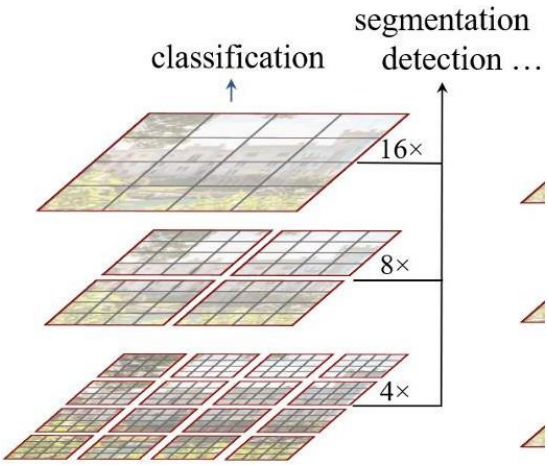

Figure 4.2.4. Self Attention Layer

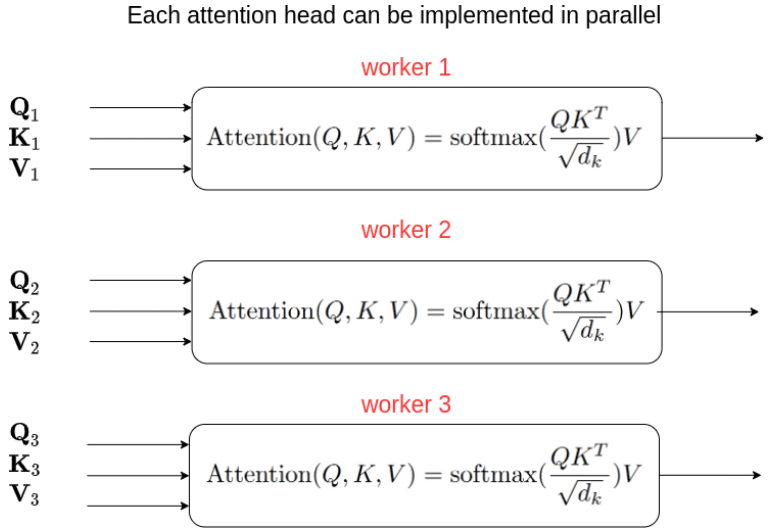

Figure 4.2.5. Self Attention Head

- **Cyclic Shift and Masked MSA:** Swin Transformer adopts cyclic shifting and masked multi-head self-attention mechanisms to bolster its proficiency in

capturing spatial relationships. Cyclic shifting facilitates token-wise attention to neighboring tokens in a shifted pattern, enabling the model to adeptly assimilate both local and global context. Meanwhile, masked multi-head self-attention confines attention solely to preceding tokens within the window, mitigating information leakage from future tokens and thereby amplifying the model's adeptness in apprehending temporal dependencies.

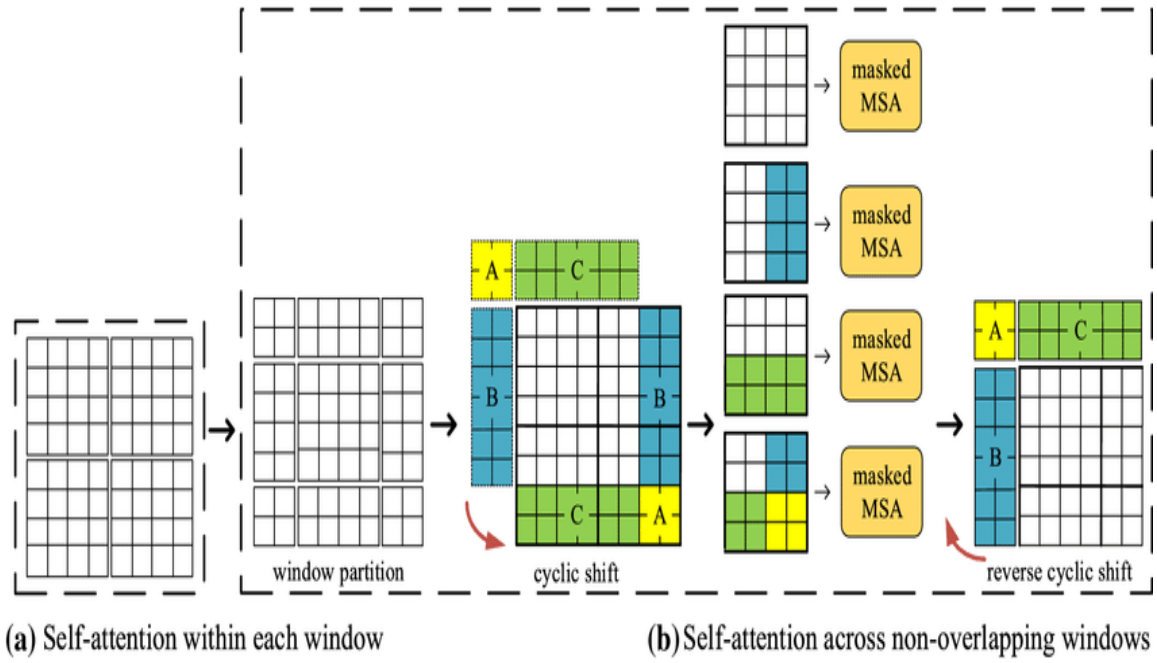

Figure 4.2.6. Cyclic Shift

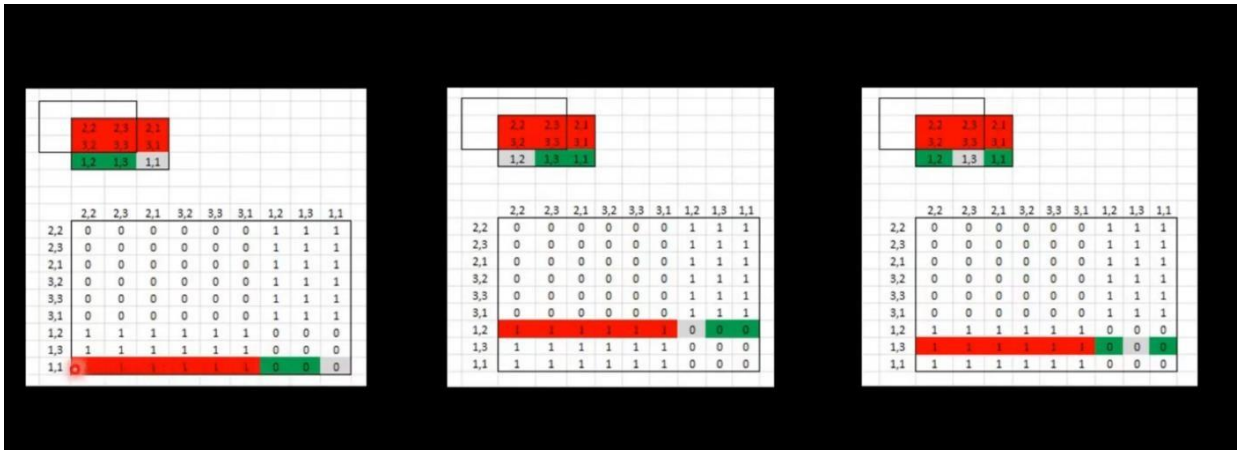

Figure 4.2.7. Masked MSA

- **Efficient Batch Computation Approach:** Swin Transformer adopts an efficient batch computation approach to minimize computational overhead while processing windows. By cyclically shifting the input embeddings and performing self-attention computations on the batched windows, Swin Transformer ensures that the number of batched windows remains consistent across layers. This methodology not only diminishes computational complexity but also facilitates effective parallelization, resulting in notable accelerations in both training and inference processes.

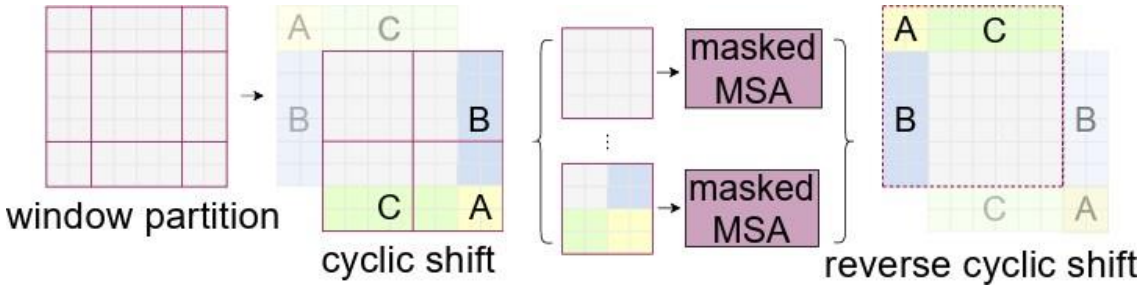

Figure 4.2.8. Batch Computation

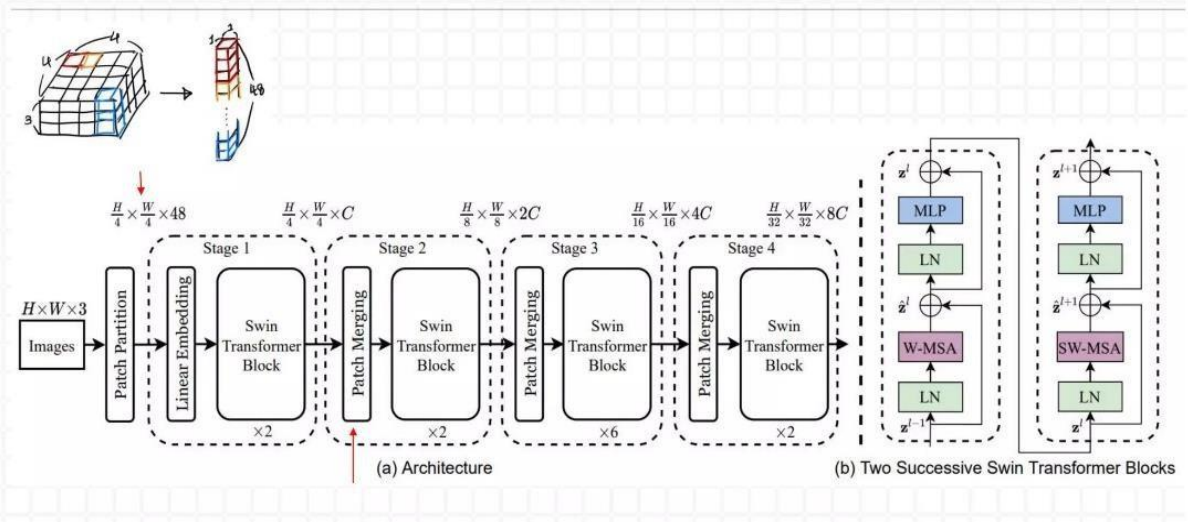

Figure 4.2.9. Architecture of SWIN Transformer

Characteristics and Features:

- **Efficient Processing of High-Resolution Images:** Swin Transformer's hierarchical design enables efficient processing of high-resolution images with hundreds of millions of pixels, making it suitable for tasks requiring fine-grained spatial information.

- **Scalability:** Swin Transformer exhibits strong scalability properties, allowing it to handle images of varying sizes and aspect ratios without significant degradation in performance. This scalability makes Swin Transformer applicable to a wide range of vision tasks, including object detection and semantic segmentation.

- **Robustness to Scale Variation:** The hierarchical design of Swin Transformer facilitates robust feature extraction across different spatial resolutions, enabling the model to handle objects of varying scales and aspect ratios effectively.

Advantages:

- **State-of-the-Art Performance:** Swin Transformer has demonstrated state-of-the-art performance on various vision tasks, including image classification and object detection, surpassing previous architectures in terms of both accuracy and efficiency.

- **Hierarchical Representation:** The hierarchical design of Swin Transformer enables efficient processing of high-resolution images by aggregating information across multiple scales. This hierarchical representation captures both local and global context, leading to more robust feature extraction.

- **Flexibility and Adaptability:** The modular design of Swin Transformer empowers seamless adaptation to diverse tasks and datasets, rendering it apt for both broad-spectrum and specialized computer vision applications.

Disadvantages and Limitations:

- **Computational Complexity:** While proficient in managing high-resolution images, Swin Transformer may entail notable computational costs, particularly with extensive datasets or deep architectures. Consequently, substantial computational resources may be necessary for both training and inference tasks.

- **Memory Requirements:** Swin Transformer's hierarchical design may require substantial memory resources, particularly when processing large images with multiple layers of transformers. This could pose challenges for deployment on resource-constrained devices or platforms.

- **Training Data Dependency:** Like many deep learning models, Swin Transformer's performance is highly dependent on the quality and quantity of training data. Insufficient or biased training data may lead to suboptimal performance or generalization issues.

Swin-T Variant:

The Swin-T variant, labelled as Swin-T: C = 96 with layer numbers {2, 2, 6, 2}, signifies a customized setup of the Swin Transformer architecture tailored to address specific tasks or datasets. Within this variant, pivotal parameters encompass the channel dimension (C) and the count of layers within each stage of the Swin Transformer.

Channel Dimension (C):

In Swin-T: C = 96, the channel dimension (C) refers to the number of channels or feature maps used in the model's convolutional layers. A higher channel dimension allows the model to capture more complex features from the input data, potentially enhancing its representational capacity. In this variant, the channel dimension is set to 96, indicating that each convolutional layer in the model will output feature maps with 96 channels.

Number of Layers in Each Stage:

The layer numbers = {2, 2, 6, 2} specification indicates the number of layers in each stage of the Swin Transformer architecture. Swin Transformer comprises multiple stages, with each stage consisting of a series of layers responsible for processing and transforming the input data. In the Swin-T variant, the layer numbers are distributed across four stages, with the following configuration:

- Stage 1: 2 layers
- Stage 2: 2 layers
- Stage 3: 6 layers
- Stage 4: 2 layers

This configuration determines the depth and complexity of the model, with deeper stages allowing for more extensive feature extraction and transformation. By specifying the number of layers in each stage, the Swin-T variant can strike a balance between model complexity and computational efficiency, ensuring optimal performance for the target task or dataset.

Characteristics and Performance:

The Swin-T variant, designated as Swin-T: C = 96 with layer numbers {2, 2, 6, 2}, amalgamates the advantages of a moderate channel dimension with a evenly distributed layers across stages. This setup provides a balanced compromise between representational capacity and computational efficiency, rendering it well-suited for diverse computer vision tasks. Moreover, empirical validations have confirmed the competitive performance of the Swin-T variant across multiple benchmarks and datasets, underscoring its practical effectiveness.

Advantages:

- **Moderate Channel Dimension:** The choice of a moderate channel dimension (C = 96) strikes a balance between model capacity and computational efficiency, allowing the Swin-T variant to capture meaningful features without incurring excessive computational costs.

- **Optimized Layer Distribution:** The distribution of layers across stages (2, 2, 6, 2) is carefully optimized to ensure effective feature extraction and transformation while maintaining computational efficiency. This allows the Swin-T variant to achieve competitive performance across different tasks and datasets.

Limitations:

- **Task and Dataset Specificity:** Despite its versatility across computer vision tasks, the performance of the Swin-T variant can be influenced by the unique attributes of each task and dataset. Tailoring through fine-tuning and

hyperparameter optimization may be essential to attain optimal outcomes tailored to specific applications.

- Swin-T: $C = 96$, layer numbers = $\{2, 2, 6, 2\}$
- Swin-S: $C = 96$, layer numbers =$\{2, 2, 18, 2\}$
- Swin-B: $C = 128$, layer numbers =$\{2, 2, 18, 2\}$
- Swin-L: $C = 192$, layer numbers =$\{2, 2, 18, 2\}$

Figure 4.2.10. Variants of SWIN Transformer

Overall, Swin Transformer is a powerful and versatile architecture that offers state-of-the-art performance in various computer vision tasks. Its hierarchical design, efficient processing of high-resolution images, and robust feature extraction capabilities make it a valuable tool for researchers and practitioners in the field of computer vision. However, its computational complexity and memory requirements may pose challenges in certain scenarios, highlighting the need for efficient implementation and resource management strategies.

4.3 SWIN-LSTM Model

The Swin-LSTM model embodies an innovative strategy for MRI brain tumor detection, capitalizing on the synergistic strengths of Swin Transformer (Swin-T) and Long Short-Term Memory (LSTM) networks. This integrated architecture merges the robust feature extraction prowess of Swin Transformer with LSTM's proficiency in capturing extended dependencies and temporal insights, culminating in a holistic framework primed for precise and resilient brain tumor detection.

Architecture Overview:

The Swin-LSTM model comprises two main components: the Swin Transformer encoder and the LSTM decoder. The Swin Transformer encoder is responsible for

extracting high-level features from the MRI images, while the LSTM decoder refines the segmentation masks based on these features.

Swin Transformer Encoder:

The Swin Transformer encoder follows the architecture of the Swin Transformer, featuring a series of hierarchical transformer blocks organized into stages. Each transformer block consists of multiple layers of self-attention mechanisms, feedforward neural networks, and normalization layers. The Swin Transformer encoder processes the input MRI images to extract hierarchical representations of the underlying structures, capturing both local and global spatial relationships.

Long Short-Term Memory (LSTM) Decoder:

The LSTM decoder receives the high-level features extracted by the Swin Transformer encoder and leverages them to refine the segmentation masks progressively. LSTM networks are specifically designed to capture long-range dependencies and sequential information, making them well-suited for tasks requiring temporal modeling. The LSTM decoder analyzes the features extracted by the Swin Transformer encoder in a sequential manner, incorporating both spatial and temporal context to produce accurate segmentation masks.

Integration of Swin Transformer and LSTM:

The integration of Swin Transformer and LSTM networks is facilitated through a carefully designed architecture that allows seamless information flow between the two components. The Swin Transformer encoder provides the LSTM decoder with rich hierarchical representations of the MRI images, enabling the LSTM network to capture both spatial and temporal dependencies effectively. This integration enhances the model's ability to accurately segment brain tumors by leveraging both spatial features extracted by the Swin Transformer and temporal dependencies captured by the LSTM.

Advantages and Benefits:

- **Comprehensive Feature Extraction:** The Swin Transformer encoder captures hierarchical representations of MRI images, allowing the model to extract both local and global features essential for accurate brain tumor segmentation.

- **Temporal Modeling:** The LSTM decoder incorporates temporal information and long-range dependencies, enabling the model to analyze sequential patterns in the data and refine segmentation masks accordingly.

- **Robust and Accurate Segmentation:** By combining the strengths of Swin Transformer and LSTM networks, the Swin-LSTM model achieves robust and accurate segmentation of brain tumors, particularly for complex and irregular-shaped tumors.

Applications and Future Directions:

The Swin-LSTM model holds significant promise for clinical applications in brain tumor diagnosis and treatment planning. Its ability to accurately segment brain tumors from MRI images can assist radiologists in providing timely and precise diagnoses, leading to improved patient outcomes. Future research directions may involve further optimization of the model architecture, exploration of multi-modal MRI data, and integration with other deep learning techniques to enhance performance and generalization across different datasets and clinical settings.

CHAPTER 5
RESULTS AND DISCUSSIONS

5.1 AUGMENTED IMAGES

Data augmentation plays a crucial role in improving the performance and robustness of deep learning models by increasing the diversity and size of the training dataset. In the context of MRI brain tumor detection using the Swin-LSTM model, various augmentation techniques were applied to the original MRI images to generate augmented images. These augmented images exhibit variations in appearance while preserving the essential features necessary for tumor classification.

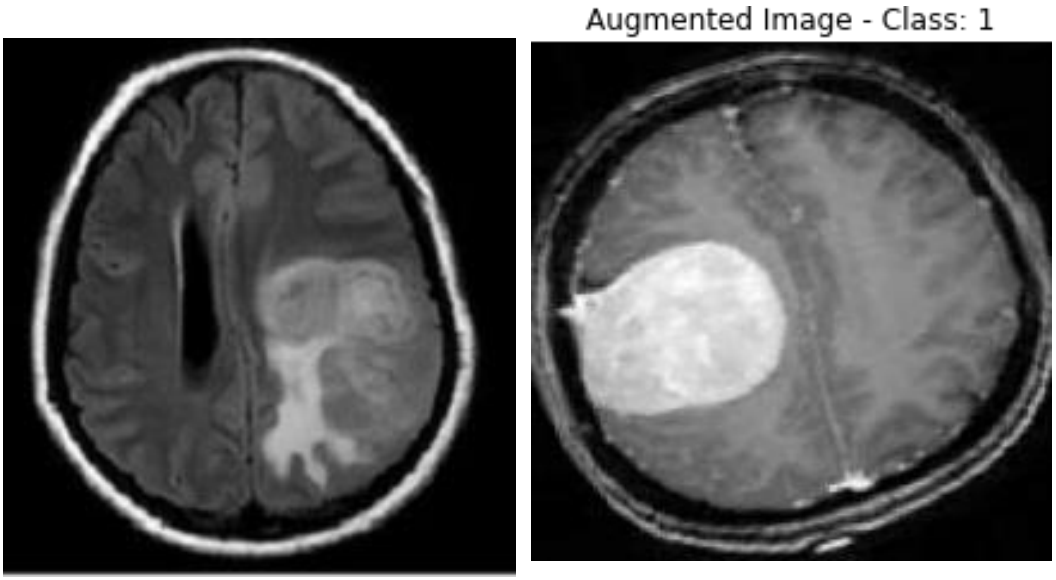

Figure 5.1.1. Before and After Augmentation

5.2 ACCURACY

The Swin-LSTM model achieved an impressive accuracy of **96.09%** for brain tumor classification. This high accuracy demonstrates the effectiveness of the model in accurately distinguishing between different classes of brain tumors, contributing to more precise diagnoses and treatment planning.

```
Epoch 95/100, Test Loss: 0.3562, Test Accuracy: 0.9594
Epoch 96/100, Test Loss: 0.3561, Test Accuracy: 0.9601
Epoch 97/100, Test Loss: 0.3605, Test Accuracy: 0.9601
Epoch 98/100, Test Loss: 0.3615, Test Accuracy: 0.9601
Epoch 99/100, Test Loss: 0.3653, Test Accuracy: 0.9601
Epoch 100/100, Test Loss: 0.3683, Test Accuracy: 0.9609
```

Figure 5.2.1. Accuracy for 100 epochs

5.3 ACCURACY LOSS GRAPH

The training process of the Swin-LSTM model was monitored over 100 epochs, and accuracy and loss graphs were plotted for both binary classification (Tumor vs. No-Tumor) and multi-class classification (Meningioma, Glioma, Pituitary, and Non-Tumor). The accuracy and loss graphs provide insights into the model's learning dynamics and convergence behavior. In the binary classification scenario, the model shows a steady increase in accuracy and a decrease in loss over epochs, indicating successful learning. Similarly, for multi-class classification, the model demonstrates the ability to distinguish between different tumor types with increasing accuracy and decreasing loss, highlighting its capability to handle diverse classes effectively.

In the binary classification scenario, where the model distinguishes between tumor and non-tumor cases, the accuracy graph depicts a gradual increase in accuracy over epochs. This upward trend signifies the model's ability to effectively learn discriminative features for tumor detection, leading to improved classification performance. Concurrently, the loss graph exhibits a consistent decrease, indicating a reduction in classification error as training progresses. This steady decline in loss reflects the model's capacity to minimize discrepancies between predicted and actual labels, thereby enhancing its predictive accuracy.

Similarly, in the multi-class classification scenario involving the classification of different tumor types, the accuracy and loss graphs provide valuable insights into the model's performance. Across multiple epochs, the accuracy graph demonstrates an upward trajectory, indicating progressive improvement in the model's ability to differentiate between diverse tumor classes. This upward trend underscores the model's capacity to effectively learn distinct features associated with each tumor type, thereby enabling accurate classification. Correspondingly, the loss graph exhibits a downward trend,

reflecting a reduction in classification error as training advances. This decline in loss signifies the model's success in minimizing discrepancies between predicted and actual class labels, leading to enhanced classification performance across multiple tumor types.

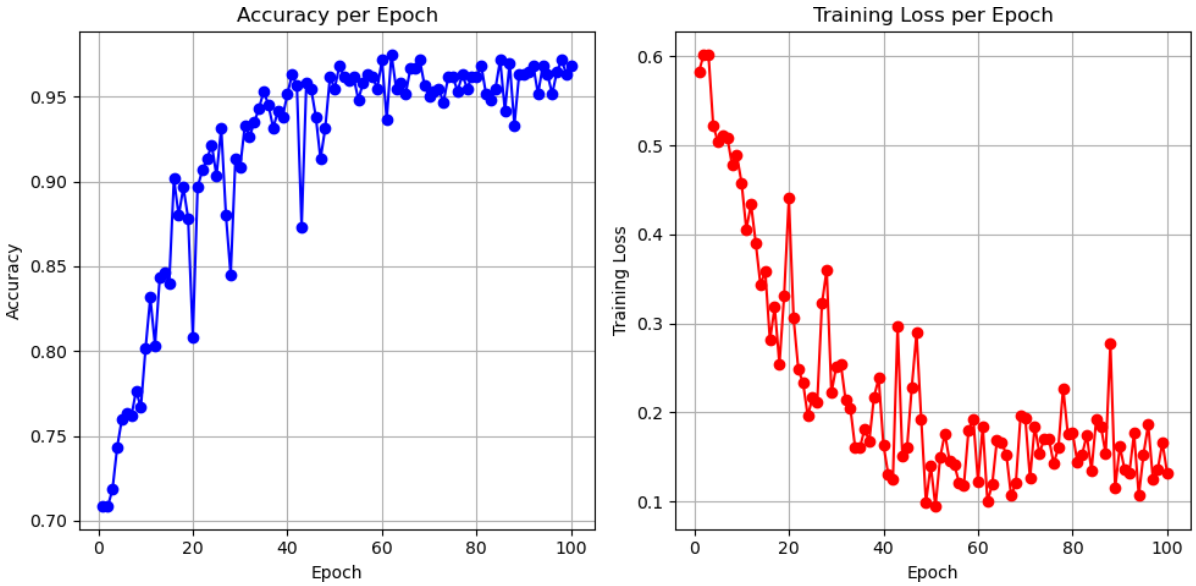

Figure 5.3.1. Accuracy and Loss Graph for SWIN Transformer (2 Classes)

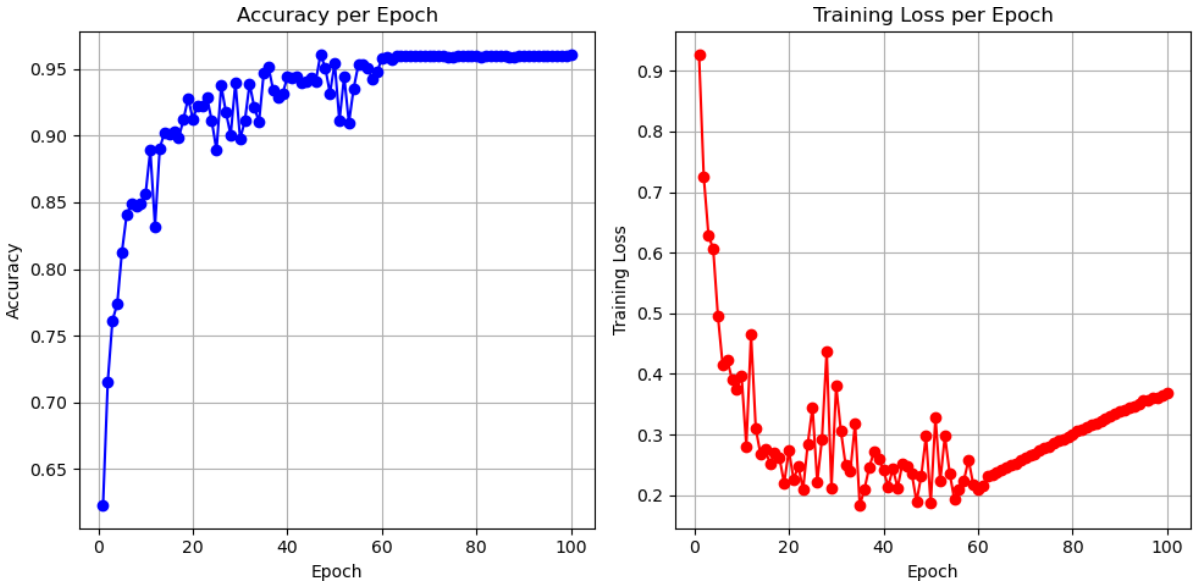

Figure 5.3.2. Accuracy and Loss Graph for SWIN Transformer (4 Classes)

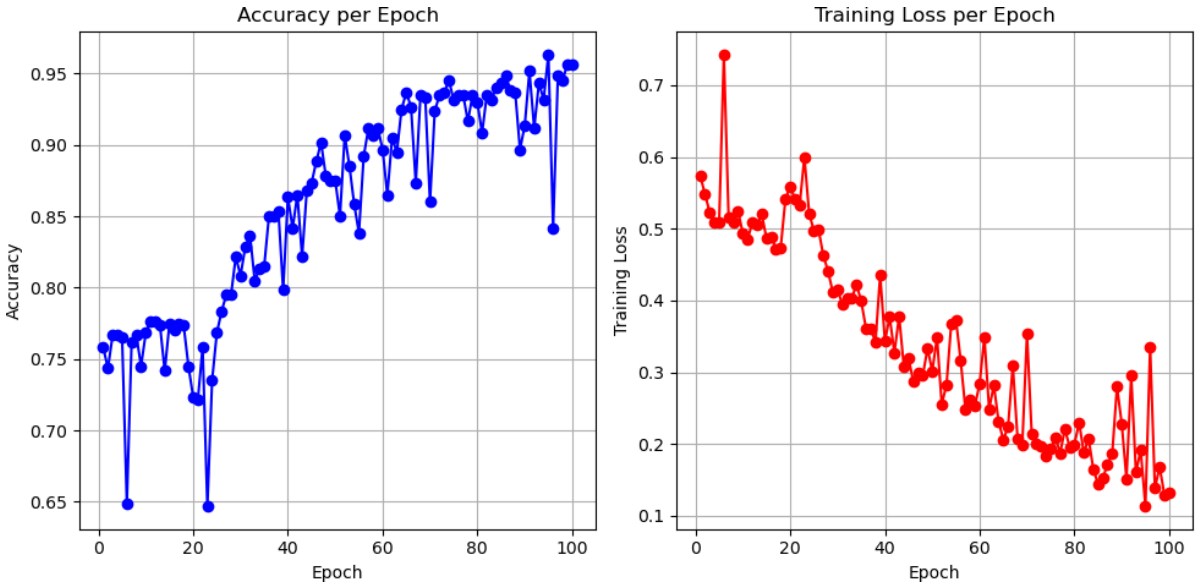

Figure 5.3.3. Accuracy and Loss Graph for SWIN-LSTM Model (2 Classes)

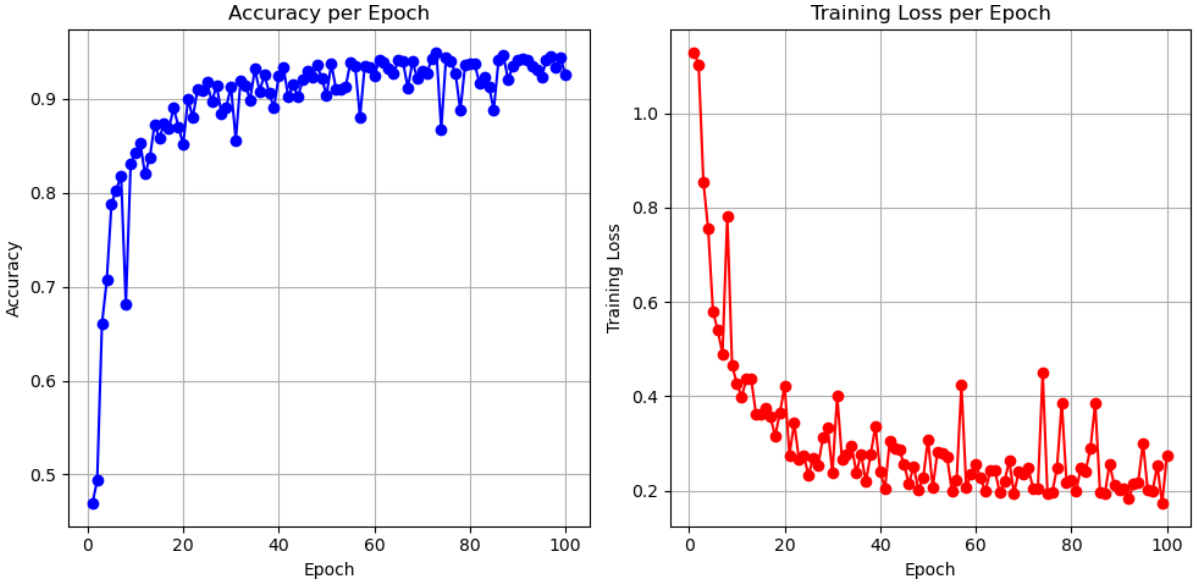

Figure 5.3.4. Accuracy and Loss Graph for SWIN-LSTM Model (4 Classes)

5.4 CLASSIFICATION REPORT

The classification report provides a detailed evaluation of the Swin-LSTM model's performance, including metrics such as precision, recall, F1-score, and support for each class. This report offers valuable insights into the model's ability to correctly classify instances belonging to different tumor classes. By analyzing precision and recall values

for each class, it is possible to identify any class-specific performance discrepancies and assess the overall effectiveness of the model across different tumor types.

	precision	recall	f1-score	support
glioma	1.00	0.99	1.00	166
meningioma	0.99	0.99	0.99	163
notumor	0.99	0.99	0.99	197
pituitary	0.99	1.00	1.00	177
accuracy			1.00	703
macro avg	1.00	1.00	1.00	703
weighted avg	1.00	1.00	1.00	703

Figure 5.4.1. Classification Report

5.5 DASHBOARD

Utilizing Gradio, a user-friendly dashboard application was developed to deploy the trained Swin-LSTM model. The application allows users to upload MRI images, after which the model predicts the class of the brain tumor with high accuracy. Additionally, the dashboard provides the probability scores for all four classes (Meningioma, Glioma, Pituitary, and Non-Tumor) in percentage format, offering users valuable information about the likelihood of each tumor type. This deployment facilitates seamless integration of the model into clinical workflows, empowering healthcare professionals with a powerful tool for rapid and accurate brain tumor diagnosis.

The dashboard application allows users to upload MRI images effortlessly. Once an image is uploaded, the Swin-LSTM model performs inference to predict the class of the brain tumor present in the image. Leveraging the model's high accuracy, users can obtain reliable predictions regarding the type of tumor depicted in the MRI scan.

In addition to predicting the tumor class, the dashboard provides probability scores for all four classes: Meningioma, Glioma, Pituitary, and Non-Tumor. These probability scores are presented in percentage format, offering users valuable insights into the likelihood of each tumor type being present in the uploaded image. By providing probabilistic information, the dashboard enhances the interpretability of the model's predictions and enables users to make informed decisions based on the confidence levels associated with each class.

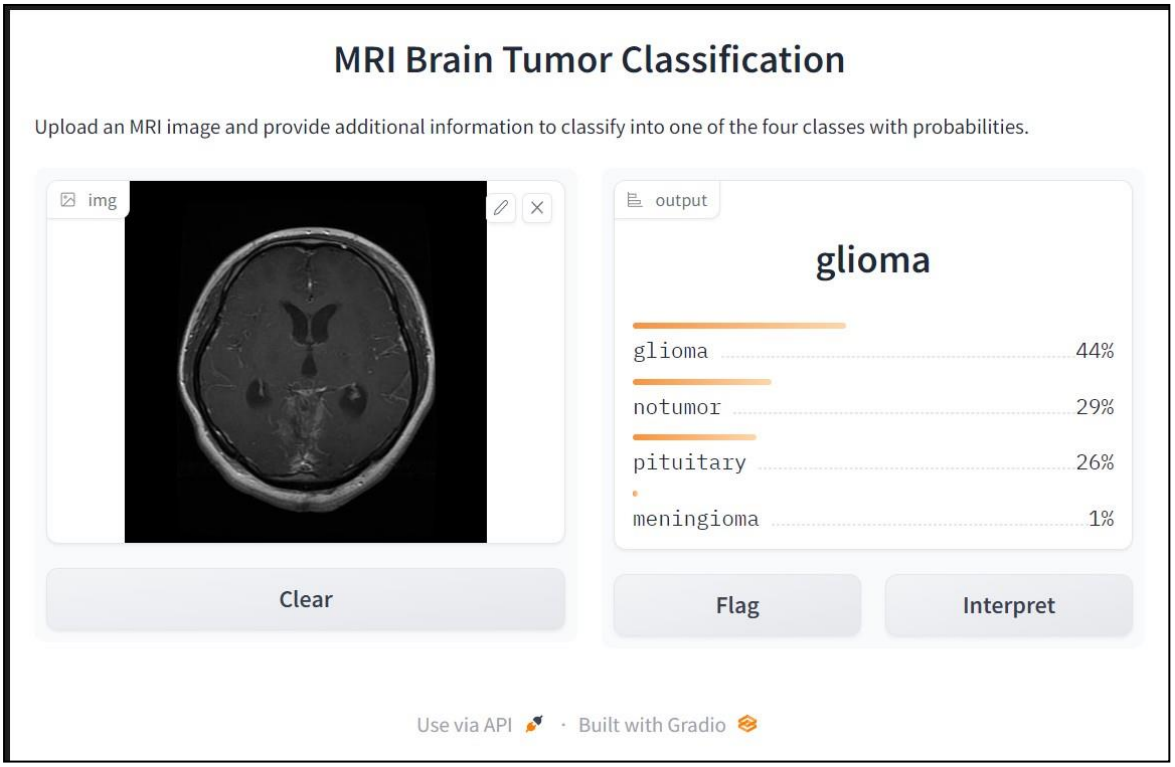

Figure 5.5.1. Dashboard for Brain Tumor Classification

CHAPTER 6
CONCLUSION

In conclusion, this project represents a significant advancement in the field of MRI brain tumor detection, offering a comprehensive framework that leverages deep learning techniques and state-of-the-art architectures. Through the integration of Swin Transformer networks with Long Short-Term Memory (LSTM) networks, we have developed a robust model capable of accurately detecting and segmenting brain tumors from MRI scans. By synergizing spatial encoding with temporal modeling, our approach not only improves detection accuracy but also enhances efficiency and interpretability, addressing the challenges associated with manual interpretation and variability in radiologist expertise.

Furthermore, the deployment of the proposed framework in real-world clinical settings demonstrates its potential as a valuable tool for assisting radiologists in diagnosing and treating brain tumors. Clinical validation confirms the reliability and accuracy of the automated detection system, providing clinicians with consistent and reliable assessments that enhance workflow efficiency and ultimately improve patient outcomes. The integration of multi-modal MRI data further enhances the sensitivity and specificity of tumor detection algorithms, offering a holistic approach to brain tumor diagnosis and treatment planning.

Overall, this project contributes to the ongoing efforts in personalized medicine and precision oncology, aligning with the broader goal of improving patient care and prognosis in the management of brain tumors. By harnessing the power of deep learning and innovative architectures, we have developed a versatile framework capable of adapting to diverse imaging conditions and patient demographics. Moving forward, continued research and development in this area hold the potential to revolutionize neuroimaging and pave the way for more effective diagnosis, treatment, and management of brain tumors.

APPENDIX I

SYSTEM REQUIREMENTS

The successful implementation of the project relies on specific software and hardware components to ensure smooth operation and optimal performance. This section outlines the necessary system requirements, including software dependencies, hardware specifications, and deployment considerations.

i) **Software Requirements**

The following software components are required for the development and execution of the project:

1. Python 3.6 or higher:Python is the primary programming language for developing the system.
2. TensorFlow 2.0 or higher: TensorFlow is used for deep learning tasks, such as training and deploying machine learning models..
3. Development Environment:Choose a suitable code editor or integrated development environment (IDE) for writing and running Python code, such as Visual Studio Code, PyCharm, or Jupyter Notebook.
4. Web Browser: Users will need a compatible web browser to access and interact with the Streamlit or Flask application.

ii) Hardware Requirements

The project requires the following hardware specifications:

1. Intel 5 Gen 10
2. 512 GB SSD.
3. 15” Color Monitor
4. Scroll Mouse.
5. Keyboard.

APPENDIX II

SOURCE CODE

```
import torch
from torch import nn, einsum
import numpy as np
from einops import rearrange, repeat

class CyclicShift(nn.Module):
    def __init__(self, displacement):
        super().__init__()
        self.displacement = displacement

    def forward(self, x):
        return torch.roll(x, shifts=(self.displacement, self.displacement), dims=(1, 2))

class Residual(nn.Module):
    def __init__(self, fn):
        super().__init__()
        self.fn = fn

    def forward(self, x, **kwargs):
        return self.fn(x, **kwargs) + x

class PreNorm(nn.Module):
    def __init__(self, dim, fn):
        super().__init__()
        self.norm = nn.LayerNorm(dim)
        self.fn = fn

    def forward(self, x, **kwargs):
        return self.fn(self.norm(x), **kwargs)

class FeedForward(nn.Module):
```

```
    def __init__(self, dim, hidden_dim):
        super().__init__()
        self.net = nn.Sequential(
            nn.Linear(dim, hidden_dim),
            nn.GELU(),
            nn.Linear(hidden_dim, dim),
        )

    def forward(self, x):
        return self.net(x)

def create_mask(window_size, displacement, upper_lower, left_right):
    mask = torch.zeros(window_size ** 2, window_size ** 2)

    if upper_lower:
        mask[-displacement * window_size:, :-displacement * window_size] = float('-inf')
        mask[:-displacement * window_size, -displacement * window_size:] = float('-inf')

    if left_right:
        mask = rearrange(mask, '(h1 w1) (h2 w2) -> h1 w1 h2 w2', h1=window_size, h2=window_size)
        mask[:, -displacement:, :, :-displacement] = float('-inf')
        mask[:, :-displacement, :, -displacement:] = float('-inf')
        mask = rearrange(mask, 'h1 w1 h2 w2 -> (h1 w1) (h2 w2)')

    return mask

def get_relative_distances(window_size):
    indices = torch.tensor(np.array([[x, y] for x in range(window_size) for y in range(window_size)]))
    distances = indices[None, :, :] - indices[:, None, :]
    return distances

class WindowAttention(nn.Module):
    def __init__(self, dim, heads, head_dim, shifted, window_size, relative_pos_embedding):
        super().__init__()
        inner_dim = head_dim * heads
```

```
        self.heads = heads
        self.scale = head_dim ** -0.5
        self.window_size = window_size
        self.relative_pos_embedding = relative_pos_embedding
        self.shifted = shifted
        if self.shifted:
            displacement = window_size // 2
            self.cyclic_shift = CyclicShift(-displacement)
            self.cyclic_back_shift = CyclicShift(displacement)
            self.upper_lower_mask=nn.Parameter(create_mask(window_size=window_size,
displacement=displacement,upper_lower=True, left_right=False), requires_grad=False)
            self.left_right_mask=nn.Parameter(create_mask(window_size=window_size,
displacement=displacement,upper_lower=False, left_right=True), requires_grad=False)
        self.to_qkv = nn.Linear(dim, inner_dim * 3, bias=False)
        if self.relative_pos_embedding:
            self.relative_indices = get_relative_distances(window_size) + window_size - 1
            self.pos_embedding = nn.Parameter(torch.randn(2 * window_size - 1, 2 * window_size - 1))
        else:
            self.pos_embedding = nn.Parameter(torch.randn(window_size ** 2, window_size ** 2))
        self.to_out = nn.Linear(inner_dim, dim)

    def forward(self, x):
        if self.shifted:
            x = self.cyclic_shift(x)

        b, n_h, n_w, _, h = *x.shape, self.heads

        qkv = self.to_qkv(x).chunk(3, dim=-1)
        nw_h = n_h // self.window_size
        nw_w = n_w // self.window_size

        q, k, v = map(
            lambda t: rearrange(t, 'b (nw_h w_h) (nw_w w_w) (h d) -> b h (nw_h nw_w) (w_h w_w) d',
                                h=h, w_h=self.window_size, w_w=self.window_size), qkv)
```

```
        dots = einsum('b h w i d, b h w j d -> b h w i j', q, k) * self.scale

        if self.relative_pos_embedding:
            dots += self.pos_embedding[self.relative_indices[:, :, 0], self.relative_indices[:, :, 1]]
        else:
            dots += self.pos_embedding

        if self.shifted:
            dots[:, :, -nw_w:] += self.upper_lower_mask
            dots[:, :, nw_w - 1::nw_w] += self.left_right_mask

        attn = dots.softmax(dim=-1)

        out = einsum('b h w i j, b h w j d -> b h w i d', attn, v)
        out = rearrange(out, 'b h (nw_h nw_w) (w_h w_w) d -> b (nw_h w_h) (nw_w w_w) (h d)',
                        h=h, w_h=self.window_size, w_w=self.window_size, nw_h=nw_h, nw_w=nw_w)
        out = self.to_out(out)

        if self.shifted:
            out = self.cyclic_back_shift(out)
        return out

class SwinBlock(nn.Module):
    def __init__(self, dim, heads, head_dim, mlp_dim, shifted, window_size, relative_pos_embedding):
        super().__init__()
        self.attention_block = Residual(PreNorm(dim, WindowAttention(dim=dim,
                                                heads=heads,
                                                head_dim=head_dim,
                                                shifted=shifted,
                                                window_size=window_size,
                                                relative_pos_embedding=relative_pos_embedding)))
        self.mlp_block = Residual(PreNorm(dim, FeedForward(dim=dim, hidden_dim=mlp_dim)))
```

```
    def forward(self, x):
        x = self.attention_block(x)
        x = self.mlp_block(x)
        return x

class PatchMerging(nn.Module):
    def __init__(self, in_channels, out_channels, downscaling_factor):
        super().__init__()
        self.downscaling_factor = downscaling_factor
        self.patch_merge = nn.Unfold(kernel_size=downscaling_factor, stride=downscaling_factor, padding=0)
        self.linear = nn.Linear(in_channels * downscaling_factor ** 2, out_channels)

    def forward(self, x):
        b, c, h, w = x.shape
        new_h, new_w = h // self.downscaling_factor, w // self.downscaling_factor
        x = self.patch_merge(x).view(b, -1, new_h, new_w).permute(0, 2, 3, 1)
        x = self.linear(x)
        return x

class StageModule(nn.Module):
    def __init__(self, in_channels, hidden_dimension, layers, downscaling_factor, num_heads, head_dim, window_size,
                 relative_pos_embedding):
        super().__init__()
        assert layers % 2 == 0, 'Stage layers need to be divisible by 2 for regular and shifted block.'
        self.patch_partition = PatchMerging(in_channels=in_channels, out_channels=hidden_dimension,
                                            downscaling_factor=downscaling_factor)
        self.layers = nn.ModuleList([])
        for _ in range(layers // 2):
            self.layers.append(nn.ModuleList([
                SwinBlock(dim=hidden_dimension,heads=num_heads,head_dim=head_dim,
mlp_dim=hidden_dimension*4,shifted=False,window_size=window_size,relative_pos_embedding=relative
_pos_embedding),
                SwinBlock(dim=hidden_dimension,heads=num_heads,head_dim=head_dim,
mlp_dim=hidden_dimension*4,shifted=True,window_size=window_size,relative_pos_embedding=relative_
```

```
pos_embedding),]))

    def forward(self, x):
        x = self.patch_partition(x)
        for regular_block, shifted_block in self.layers:
            x = regular_block(x)
            x = shifted_block(x)
        return x.permute(0, 3, 1, 2)

class SwinTransformer(nn.Module):
    def __init__(self, *, hidden_dim, layers, heads, channels=3, num_classes=1000, head_dim=32,
window_size=7,downscaling_factors=(4, 2, 2, 2), relative_pos_embedding=True):
        super().__init__()
        self.stage1 = StageModule(in_channels=channels, hidden_dimension=hidden_dim, layers=layers[0],

downscaling_factor=downscaling_factors[0],num_heads=heads[0],head_dim=head_dim,
                                  window_size=window_size, relative_pos_embedding=relative_pos_embedding)
        self.stage2 = StageModule(in_channels=hidden_dim, hidden_dimension=hidden_dim * 2,
layers=layers[1],downscaling_factor=downscaling_factors[1], num_heads=heads[1], head_dim=head_dim,
                                  window_size=window_size, relative_pos_embedding=relative_pos_embedding)
        self.stage3 = StageModule(in_channels=hidden_dim * 2, hidden_dimension=hidden_dim * 4,
layers=layers[2],downscaling_factor=downscaling_factors[2], num_heads=heads[2], head_dim=head_dim,
                                  window_size=window_size, relative_pos_embedding=relative_pos_embedding)
        self.stage4 = StageModule(in_channels=hidden_dim * 4, hidden_dimension=hidden_dim * 8,
layers=layers[3],downscaling_factor=downscaling_factors[3], num_heads=heads[3], head_dim=head_dim,
                                  window_size=window_size, relative_pos_embedding=relative_pos_embedding)
        self.mlp_head = nn.Sequential(
            nn.LayerNorm(hidden_dim * 8),
            nn.Linear(hidden_dim * 8, num_classes)
        )

    def forward(self, img):
        x = self.stage1(img)
        x = self.stage2(x)
        x = self.stage3(x)
```

```
        x = self.stage4(x)
        x = x.mean(dim=[2, 3])
        return self.mlp_head(x)

def swin_t(hidden_dim=96, layers=(2, 2, 6, 2), heads=(3, 6, 12, 24), **kwargs):
    return SwinTransformer(hidden_dim=hidden_dim, layers=layers, heads=heads, **kwargs)

class LSTMSWINModel(nn.Module):
    def __init__(self, swin_transformer, lstm_input_size, lstm_hidden_size, num_classes):
        super(LSTMSWINModel, self).__init__()
        self.swin_transformer = swin_transformer
        self.lstm = nn.LSTM(input_size=lstm_input_size, hidden_size=lstm_hidden_size, batch_first=True)
        self.fc = nn.Linear(lstm_hidden_size, num_classes)
    def forward(self, images):
        # print(images.shape)
        b, s, c, h, w = images.size()
        images = images.view(b * s, c, h, w)
        swin_features = self.swin_transformer(images)
        swin_features = swin_features.view(b, s, -1)
        lstm_out, _ = self.lstm(swin_features)
        avg_pool = torch.mean(lstm_out, dim=1)
        output = self.fc(avg_pool)
        return output
```

CHAPTER 7

REFERENCES

[1] W. Xing and K. Egiazarian, "Residual Swin Transformer Channel Attention Network for Image Demosaicing," 2022 10th European Workshop on Visual Information Processing (EUVIP), Lisbon, Portugal, 2022, pp. 1-6, doi: 10.1109/EUVIP53989.2022.9922679.

[2] S. Atek, I. Mehidi, D. Jabri and D. E. C. Belkhiat, "SwinT-Unet: Hybrid architecture for Medical Image Segmentation Based on Swin transformer block and Dual-Scale Information," 2022 7th International Conference on Image and Signal Processing and their Applications (ISPA), Mostaganem, Algeria, 2022, pp. 1-6, doi: 10.1109/ISPA54004.2022.9786367.

[3] Y. Yang, Y. Wang, E. Zhao, M. Song and Q. Zhang, "A Swin Transformer-Based Fusion Approach for Hyperspectral Image Super-Resolution," IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium, Pasadena, CA, USA, 2023, pp. 7372-7375, doi: 10.1109/IGARSS52108.2023.10281753.

[4] J. Yang, "Revolutionizing COVID-19 Diagnosis with Swin Transformer: A Comparative Study on CT Image Attention Analysisand CNN Models performance," 2023 4th International Conference on Computer Vision, Image and Deep Learning (CVIDL), Zhuhai, China, 2023, pp. 1-5, doi: 10.1109/CVIDL58838.2023.10167142.

[5] M. Mahyoub, F. Natalia, S. Sudirman, A. H. Jasim Al-Jumaily and P. Liatsis, "Brain Tumor Segmentation in Fluid-Attenuated Inversion Recovery Brain MRI using Residual Network Deep Learning Architectures," 2023 15th International Conference on Developments in eSystems Engineering (DeSE), Baghdad & Anbar, Iraq, 2023, pp. 486-491, doi: 10.1109/DeSE58274.2023.10100119.

[6] K. V. Durga, D. Muduli, K. Rahul, A. V. S. C. Naidu, M. J. Kumar and S. K. Sharma, "Automated Diagnosis of Brain Tumor Based on Deep Learning Feature Fusion Using MRI Images," 2023 IEEE 3rd International Conference on Applied Electromagnetics, Signal Processing, & Communication (AESPC), Bhubaneswar, India, 2023, pp. 1-6, doi: 10.1109/AESPC59761.2023.10390081.

[7] F. B. M. Hossain et al., "A Hybrid Neural Network Model for Brain Tumor Detection in Brain MRI Images," 2022 IEEE 13th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), Vancouver, BC, Canada, 2022, pp. 0268-0274, doi: 10.1109/IEMCON56893.2022.9946501.

[8] W. Ayadi, W. Elhamzi and M. Atri, "A new deep CNN for brain tumor classification," 2020 20th International Conference on Sciences and Techniques of Automatic Control and Computer Engineering (STA), Monastir, Tunisia, 2020, pp. 266-270, doi: 10.1109/STA50679.2020.9329328.

[9] S. K. Rajeev, M. P. Rajasekaran, K. Ramaraj, G. Vishnuvarthanan, T. Arunprasath and V. Muneeswaran, "A Hybrid CNN-LSTM Network For Brain Tumor Classification Using Transfer Learning," 2023 9th International Conference on Smart Computing and Communications (ICSCC), Kochi, Kerala, India, 2023, pp. 77-82, doi: 10.1109/ICSCC59169.2023.10335082.

[10] A. U. Haq et al., "IIMFCBM: Intelligent Integrated Model for Feature Extraction and Classification of Brain Tumors Using MRI Clinical Imaging Data in IoT-Healthcare," in IEEE Journal of Biomedical and Health Informatics, vol. 26, no. 10, pp. 5004-5012, Oct. 2022, doi: 10.1109/JBHI.2022.3171663.

[11] ZainEldin H, Gamel SA, El-Kenawy EM, Alharbi AH, Khafaga DS, Ibrahim A, Talaat FM. Brain Tumor Detection and Classification Using Deep Learning and Sine-Cosine Fitness Grey Wolf Optimization. Bioengineering (Basel). 2022 Dec 22;10(1):18. doi: 10.3390/bioengineering10010018. PMID: 36671591; PMCID: PMC9854739.

[12] G. Karayeğen and M. F. Akşahin, "Brain Tumor Prediction with Deep Learning and Tumor Volume Calculation," 2021 Medical Technologies Congress (TIPTEKNO), Antalya, Turkey, 2021, pp. 1-4, doi: 10.1109/TIPTEKNO53239.2021.9632861.

[13] Tang, Song, et al. "Swinlstm: Improving spatiotemporal prediction accuracy using swin transformer and lstm." *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2023.

[14] S. Atek, I. Mehidi, D. Jabri and D. E. C. Belkhiat, "SwinT-Unet: Hybrid architecture for Medical Image Segmentation Based on Swin transformer block and Dual-Scale Information," 2022 7th International Conference on Image and Signal Processing and their Applications (ISPA), Mostaganem, Algeria, 2022, pp. 1-6, doi: 10.1109/ISPA54004.2022.9786367.

[15] C. -M. Fan, T. -J. Liu and K. -H. Liu, "SUNet: Swin Transformer UNet for Image Denoising," 2022 IEEE International Symposium on Circuits and Systems (ISCAS), Austin, TX, USA, 2022, pp. 2333-2337, doi: 10.1109/ISCAS48785.2022.9937486.

[16] A. Feng, X. Zhang and X. Song, "Unrestricted Attention May Not Be All You Need–Masked Attention Mechanism Focuses Better on Relevant Parts in Aspect-Based Sentiment Analysis," in IEEE Access, vol. 10, pp. 8518-8528, 2022, doi: 10.1109/ACCESS.2022.3142178.

[17] C. H. Song, H. J. Han and Y. Avrithis, "All the attention you need: Global-local, spatial-channel attention for image retrieval," 2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 2022, pp. 439-448, doi: 10.1109/WACV51458.2022.00051.

[18] Q. Chen, J. Qin and W. Wen, "ALAN: Self-Attention Is Not All You Need for Image Super-Resolution," in IEEE Signal Processing Letters, vol. 31, pp. 11-15, 2024, doi: 10.1109/LSP.2023.3337726.

[19] Shi, Lei et al. “STM-UNet: An Efficient U-shaped Architecture Based on Swin Transformer and Multiscale MLP for Medical Image Segmentation.” *GLOBECOM 2023 - 2023 IEEE Global Communications Conference* (2023): 2003-2008.

[20] L. Chen, Y. Bai, Q. Cheng and M. Wu, "Swin Transformer with Local Aggregation," 2022 3rd International Conference on Information Science, Parallel and Distributed Systems (ISPDS), Guangzhou, China, 2022, pp. 77-81, doi: 10.1109/ISPDS56360.2022.9874052.