DEVELOPMENT OF AN EXPLAINABLE DEEP LEARNING MODEL FOR MRI-BASED STROKE DETECTION AND LESION SEGMENTATION

BY

ONYIA KASIEMOBI JESSE

REGISTRATION NUMBER: GOU/U22/CSC/930

FOR

THE DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE,

FACULTY OF COMPUTING AND INFORMATION TECHNOLOGY,

GODFREY OKOYE UNIVERSITY, UGWU-OMU NIKE, ENUGU, ENUGU STATE.

JUNE 2026

DEVELOPMENT OF AN EXPLAINABLE DEEP LEARNING MODEL FOR MRI-BASED STROKE DETECTION AND LESION SEGMENTATION

BY

ONYIA KASIEMOBI JESSE

REGISTRATION NUMBER: GOU/U22/CSC/930

A PROJECT REPORT SUBMITTED TO THE DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE, FACULTY OF COMPUTING AND INFORMATION TECHNOLOGY, GODFREY OKOYE UNIVERSITY, UGWU-OMU NIKE, ENUGU, ENUGU STATE.

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE AWARD OF BACHELOR OF SCIENCE (B.SC.) DEGREE IN COMPUTER SCIENCE

SUPERVISOR: ENGR. CAMILLUS NJOKU

JUNE 2026

APPROVAL PAGE

This research, titled "DEVELOPMENT OF AN EXPLAINABLE DEEP LEARNING MODEL FOR MRI-BASED STROKE DETECTION AND LESION SEGMENTATION." has been assessed and approved by the Department of Computer Science and Mathematics Committees in the Faculty of Computing And Information Technology, Godfrey Okoye University, Enugu, Nigeria.

Supervisor	....	....
Engr Camillus Njoku	Signature	Date
External Examiner	....	....
	Signature	Date
Dr. Frank Okebanama	....	....
HOD, Department of Computer Science and Mathematics	Signature	Date
Prof. Ifeyinwa Ajah	....	....
Dean, Faculty of Computing And Information Technology.	Signature	Date

CERTIFICATION

I certify that I completed the research included therein and that it was primarily based on my observations and investigation. Consequently, I will fully accept the University's decision if I am found to have cheated on the assignment.

Onyia Kasiemobi Jesse

GOU/U22/CSC/930

DEDICATION

This project is dedicated to my parents and family, whose unwavering support, sacrifices, and words of encouragement made this academic journey possible, and to every student and administrator in Nigerian universities who deserves a better, fairer, and more transparent system for managing academic records.

ACKNOWLEDGEMENTS

First and foremost, I give all glory and honor to Almighty God who alone through his provision and grace brought me through every phase of this research. I express my deepest appreciation to my supervisor, Dr. Frank Okebanamma whose advice, patience, criticism and comments during this research helped make the difference. His involvement and willingness to always have my ideas challenged me through this project.

I also thank Dr. Frank Okebanama, HOD Computer Science & Mathematics, My Supervisor, as well as all other lecturers in the Computer Science & Mathematics department whose academic teachings throughout my undergraduate career were fundamental to this project. To the lecturers whose work inspired me to find new ways of thinking, I say thank you.

My profound gratitude to my parents and family for the sacrifices I know that my education required. Thank you for believing in me at the times I felt like giving up.

Finally, my friends and classmates who provided assistance in testing the application, and with comments and ideas throughout this project, I thank you.

ABSTRACT

This thesis introduces StrokeVision AI, an explainable AI tool to facilitate early identification of ischemic stroke lesion on DWI MR images. The need for this type of tool arose due to limitations in the diagnosis of acute stroke especially for those who present in resource-limited settings and a need for speedy diagnosis with improved accuracy and transparency. StrokeVision AI comprises a 2D U-Net model trained using theISLES2022[98] dataset and a visual explanation of predictions based on Occlusion Sensitivity Mapping(OSM). The complete StrokeVision AI system is designed as an open-source web application using FastAPI, React, MySQL and Docker. Training was performed without evidence of overfitting and exhibited stability. The DSC of0.6751obtained outperformed the generally accepted range of performance for comparable 2D U-Net models. The free access and the ease of deployment of the StrokeVision AI system makes it an extremely relevant tool for the healthcare system in Nigeria and more broadly across sub-Saharan African countries, where resource constraints for diagnosis of stroke are prominent.

TABLE OF CONTENTS

List of Tables

List of Figures

CHAPTER ONE: INTRODUCTION

1.1 Background of the Study

Stroke is a critical medical condition where disrupted brain blood supply leads to ischemia or hemorrhage; billions of people throughout the world die from stroke annually with low-and middle-income nations such as Nigeria being burdened with limited access to neuro- imaging technologies and expertise. There are typically four major types of strokes. The Ischemic stroke, also called cerebral ischemia, accounting for about 80-90% of strokes is characterized by cerebral arteries blockage creating rapidly growing ischemic core and surrounding penumbra that require prompt Diagnosis and reperfusion treatment in order to improve prognosis.

The use of Neuroimaging such as MRI imaging sequences (DWI and FLAIR) helps in stroke Diagnosis and treatment planning. However, manual diagnosis on MRI imaging for stroke evaluation is generally time-consuming, qualitative and also prone to variability; therefore, automated analysis is needed. Here we designed StrokeVision AI; A U-Net based web application for: (1) image segmentation, (2) quantification, and (3) for explainable clinical decision-making**.**

1.2 Statement of the Problem

There are three primary factors preventing current deep learning stroke segmentation systems from effective translation into clinical practice that this project aims to overcome:

1. Transparency in deep learning models is an even greater obstacle for clinical uptake since a number of the top-performing stroke segmentation systems are 'black boxes' that provide an output but do not indicate the reasoning behind it. In neuroimaging, where such results guide important decisions regarding thrombolysis and thrombectomy, a lack of transparency represents both a

medico-legal and ethical issue. Limited explainability was thus listed as one of the main issues preventing AI integration in radiology practice. Transparency is now a fundamental aspect required by high-risk medical AI systems under both the EU AI Act and SaMD regulations from the FDA.

2. Current literature in deep learning-based stroke segmentation typically presents models developed and evaluated in isolation without thought given to real-world clinical deployment. An algorithm that segments accurately in a researcher's notebook, but requires significant work to provide clinicians with an authenticated, web-based interface, persistently stored results, structured diagnostic reports, and a seamlessly integrated workflow within existing hospital systems, is not a practical clinical tool. This gap between research prototypes and clinically implementable systems has seldom been the subject of discussion in current research literature.

3. Early implementation efforts invariably encounter significant bugs that halt the learning process altogether. Some of the technical issues that we had to resolve in this project include: (1) a learning process that consistently predicted middle slices containing no lesion tissue, forcing the model to train on all-zero masks, (2) an error in mask shape resulting from using a single-level decoder to output 64x64 segmentation maps, rather than the expected 224x224 format; (3) accidental double application of the sigmoid activation function to the loss function, (4) omission of a detailed keyword argument which would later break training in PyTorch 2.2+. These technical errors must be identified, debugged and fixed to achieve effective model learning and development.

1.3 Aim and Objectives of the study

The primary goal of this research is to develop and validate an explainable deep learning framework designed for the automatic detection and segmentation of ischemic stroke lesions from DWI MRI, delivered as a production-ready full-stack clinical web application.

Specific OBJECTIVES include:

1. Conduct a literature review of deep learning-based stroke lesion segmentation models, focusing on encoder-decoder architecture, benchmarks and explainability approaches.
2. Design and train a 2D U-Net segmentation model that utilizes dropout regularization in its input layer. The model should be trained on the ISLES 2022 dataset using a combined Dice-BCE loss function that overweighs the positive class in an effort to mitigate class imbalance.
3. Implement a production-ready deep learning pipeline to circumvent the black-box nature of the stroke AI systems by enabling explainability using Occlusion Sensitivity Maps (OSMs).
4. Design a user-friendly web application to automate detection, segmentation and explainability of stroke from DWI MR imaging scans.
5. Evaluate the model's segmentation performance using Dice Similarity Coefficient (DSC), Intersection-over-Union (IoU), Precision, Recall, Specificity, and Hausdorff Distance on the held-out ISLES 2022 test set.

1.4 Significance of the Study

StrokeVision AI has proved that effective, high performance stroke lesion segmentation and OSM-based explainability can be incorporated into a single, production ready AI deep learning pipeline, overcoming the 'black-box' problem with clinical AI which has persisted for a long time. The solution offered is an end-to-end solution that can be converted to an open-source platform and incorporate DWI segmentation, OSM-based explainability, permanent MySQL storage, JWT-based authentication, and PDF generation through ReportLab, all within a single docker compose infrastructure. Moreover, the systematic isolation and correction of significant implementation issues-shape mismatches, training using all-zero mask, incorrect loss function settings and deprecations of PyTorch API's-offers pragmatic learning points for clinical AI development. The implementation of StrokeVision AI proves that affordable, powerful AI can be deployed within Nigerian and African environments, bridging health inequalities in low-resource settings through open-source infrastructure.

1.5 Scope of the Study

This work demonstrates the application of a perturbation-based approach for model-agnostic explainability, called Occlusion Sensitivity Mapping (OSM), where influential regions can be revealed by occluding input patches and observing the prediction of the model. It requires no model-specific features such as gradients and is architecture-agnostic as it only operates on U-Net predictions of shape (B, 1, H, W) (Lundberg & Lee, 2017). For this application we train on DWIs with NIfTI format for the detection of acute ischemic stroke. The model was trained on ISLES 2022 dataset which consists of 400 cases split at 70/15/15. A 2D U-Net model of size 224x224 was trained over 50 epochs on a single NVIDIA T4 GPU using per-slice min-max normalization, dropout of 0.2 and a Dice-BCE loss with class weighting (pos_weight=20). The application consists of the following key components: A FastAPI, SQLAlchemy, and MySQL based backend coupled with a React and TypeScript based frontend. This is orchestrated and deployed with Docker Compose. It includes a system for JWT-based authentication and secure data handling as well as ReportLab for PDF report generation and persistent storage of segmentation and explainability outputs.

1.6 Limitations of the Study

2D Architecture: 2D-slice-based architecture lack volumetric perspective compared to 3D-archictecture; based on the data, 3D U-Net and nnU-Net should give higher DSC but requires GPU memory capacity over Google Colab T4. Occlusion Sensitivity Latency. This has many model evaluations and is slow in explanation. Validation: No prospective clinical trial: retrospective evaluation on ISLES 2022 held-out test split only, not validated prospectively in a real-time Nigerian clinical setting. Only Ischemic stroke: The design is on ischemic stroke. Hemorrhagic stroke, hemorrhagic transformation and other lesions of brain are out of scope and may causes error positive. GPU is required: In real time inference we assume GPU in the deployed environment. CPU only inference is about 10 times slower. Single Modality: The current system uses DWI only; integrate ADC and FLAIR would better the lesion characterization.

1.7 Operational Definition of Terms

Ischemic Stroke: A neurovascular pathology that occurs as a result of a Thrombotic or Embolic occlusion of a cerebral artery leading to impaired cerebral blood flow leading to neurodegeneration (Lundberg & Lee, 2017).

StrokeVision AI: Is the production grade and custom clinical decision support web platform developed in this project for MRI-based ischemic stroke detection, segmentation and explainable AI. Reporting.

Occlusion sensitivity map-based explainability: Is an approach to post-hoc interpretation that occludes part of the input image and measures how much the prediction accuracy changes. By calculating how the output performance is affected as individual regions are taken away, it is possible to produce a spatial sensitivity map. Areas where a significant change in performance is observed are attributed to be the parts that are most important in the prediction task. This is a very intuitive method of providing local visual explanations for the predictions of deep learning models. It has potential applications in areas like medical imaging in segmenting lesions.

Dice Similarity Coefficient (DSC): Is a statistical value for the estimate spatial overlap of the set in the predicted/binary ground-truth segmentation masks. Number of the value ranging from 0 (no overlap) to 1 (ideal overlay) (Litjens et al., 2017).

U-Net: An encoder-decoder convolutional neural network architecture with skip connection at each resolution level, typically used for biomedical image segmented (Ronneberger et al., 2015).

ISLES 2022: 2022 edition of the Ischemic Stroke Lesion Segmentation challenge comprising 400 multi-center DWI_MRI cases with expert lesion annotation (Hernandez Petzsche et al., 2022).

Fast-API: Python web framework based on OpenAPI and Pydantic that is suitable for building production – ready web services used as REST API server for the backend of StrokeVision AI.

Docker-Compose: An orchestration system of Containers which enabling defining and running multi-container Docker applications used in the StrokeVision AI to deploy all service with one command.

CHAPTER TWO: LITERATURE REVIEW

2.1 Introduction

The chapter covers the underlying clinical pathology of ischemic stroke, appropriate imaging modalities used in stroke diagnosis (specifically MR imaging), the evolution of automatic stroke segmentation with the advancement from classic Machine learning algorithms to deep learning models like CNN, U-net, attention-based mechanism and Transformers. Multi-model fusion, Benchmark datasets and progress and major developments of prior research are also covered in this chapter.

2.2.1 Theoretical Background: Stroke clinical pathology and imaging modalities

Thrombotic or embolic occlusion of the cerebral vasculature will result in an ischemic stroke, causing neurons in the infarct core to die within 4-8 minutes of complete ischemia. However, surrounding this ischemic core is an ischemic penumbra of hypoperfused but potentially salvageable tissue which may survive up to 6-24 hours depending on the quality of collateral circulation and perfusion pressure. (Ermine et al., 2021; Liu, 2010) The penumbra can be reperfused and so with prompt revascularization via thrombolysis or mechanical thrombectomy infarct expansion, permanent damage and subsequent clinical deficit will be reduced. Diffusion-weighted imaging (DWI) is the most sensitive and specific sequence in the detection of hyperacute infarcts, visualizing an infarct 30 minutes after symptom onset using signal abnormalities associated with the diffusion restriction seen during cytotoxic oedema. (Chauhan et al., 2022) These infracted areas show hypointensity on apparent diffusion coefficient (ADC) maps. T2 FLAIR sequences are generally hyperintense 4-6 hours after onset, remaining positive in the subacute and chronic stages, (Thomalla et al., 2011; Lansberg et al., 2001) and a chronic infarct will show cortical and subcortical tissue loss in a T1 sequence. (Liew et al., 2018) However, there is inherent variability in the different types of scanners, sequences and field strengths. This may provide the 'completeness' to multiparametric assessment but would be very difficult to standardize in the automated assessment. (Ghafoorian et al., 2020)

2.2.2 Traditional Machine Learning Approaches

Before the emergence of deep learning approaches, stroke lesion segmentation has been investigated using both classical image processing techniques and shallow machine learning algorithms. Thresholding methods segment hyperintense lesions on DWI images using fixed and adaptive thresholding techniques such as Otsu thresholding (Frey & Fischer, 2016). While simple and fast, this method proves sensitive to noise in MR images, scanner drift, and signal fluctuation in MRI datasets. Atlas-based segmentation works by registering the images to a standard brain atlas and localizing lesions by detecting unexpected deviation from the expected normal tissue intensity distribution (Muschelli et al., 2019). Nevertheless, these methods cannot work when the mass effect is pronounced or when atrophy exists, which are frequently observed in chronic stroke cases. Training on manually engineered intensity and texture features, Random Forest classifier obtained an average DSC score of around 0.68 on the ISLES 2015 subacute dataset (Maier et al., 2017). Although this technique out-performed thresholding methods, such approaches were hampered by hand-engineered features extraction, limited representational power of designed features. Along with SVM classifier, trained on voxel-wise features, obtained comparable performance as baselines (Tavakoli et al., 2017). However, they did not avoid the limitations of scale and computation efficiency for large and varied dataset.

2.2.3 Convolutional Neural Networks in Medical Imaging

The first application of AlexNet (Krizhevsky et al., 2012) for image analysis in 2012 was a groundbreaking development for medical imaging analysis; it showed that CNNs are capable of learning hierarchical features directly from raw images, eliminating manual feature extraction. Based on a review carried out by Litjens et al. (Litjens et al., 2017) on the use of CNNs in medical imaging, by 2017 CNN-based approaches were either matching or exceeding expert-level performance in a range of medical imaging benchmark challenges. To segment brain lesions, Kamnitsas et al. (Kamnitsas et al., 2017) developed DeepMedic which is a dual pathway 3D CNN designed to simultaneously extract contextual information and learn at multiple scales of the input 3D volumetric data using parallel input pathways. Their model performed well on the task of ischemic stroke lesion segmentation yielding a DSC of 0.79. Compared to the previous single-pathway CNN approaches (U.S. Food and Drug Administration, 2021), This performance of DeepMedic in performing feature learning in a more robust way (with its ability to extract multi-scale 3D features), however, is at the expense of huge computation resources.

2.2.4 Encoder-Decoder Architecture for Segmentation

The U-Net architecture developed by Ronneberger, Fischer and Brox (Ronneberger et al., 2015) at MICCAI 2015 has indeed caused a wave change on medical image segmentation. The encoder part of U-Net repeatedly down-samples input images via max-pooling layers to extract deep semantic features with low spatial resolutions. The decoder part does just the opposite by performing up-sampling using transposed convolutions till the original resolution. What is most unique about it is the skip connection, which copies the feature maps from encoder to decoder at every resolution level, thus recovering the lost detail from encoder down-sampling. We assume that it is precisely this feature that provides precise boundary detection and fine detail keeping with small dataset, crucial in medical segmentation. 3D U-Net was developed by Iek et al. (Çiçek et al.,

2016) in order to adapt U-Net to 3D data. It utilizes 3D convolution operations and explores inter-slice volumetric context, rather than 2D convolution. NnU-Net developed by Isensee et al. (Isensee et al., 2021) has automated preprocessing stages, network architecture selection (2D, 3D, cascade) and hyperparameter optimization, winning 0.85 DSC on ISLES 2022 dataset, thus becoming a benchmark. Residual shortcut connection developed by He et al. (He et al., 2016) resolved the vanishing gradient issue in very deep networks, thus allowing for networks with hundreds of layers to be trained. V-Net by Milletari et al. (Milletari et al., 2016) adopted residual connections and combined with Dice-based loss function. It shares similarities with our combined Dice-BCE loss in StrokeVision AI. In an experiment, Clèrigues et al (Clèrigues et al., 2020) confirmed that for acute stroke segmentation and particularly for small lesions, 3D U-Net significantly outperforms 2D network (3-5 DSC points higher).

2.2.5 Attention mechanism and Transformer and models

Attention mechanism provides ability for models to selectively assign weights to spatial locations based on their relevance. Oktay et al (Oktay et al., 2018) integrated attention gates to U-Net's skip connections, resulting in a network which learns to focus on useful features and suppresses signals from redundant regions. Attention U-Net achieved DSC 0.85 for cardiac segmentation and spatial focus of visual attention map provides an intuitive explanation. The ViT model uses multi-head self-attention mechanism to encode long-range spatial dependencies that standard convolutions cannot achieve, it encodes the non-overlap images patch in sequences of tokens. Integrating CNN encoder and Transformer bottleneck, the Transformer U-Net provided a learning-based way of performing feature cross-correlation on multi-organ segmentation. As for stroke specifically, Luo et al (Luo et al., 2022) designed a hybrid CNN-Transformer architecture that achieved DSC 0.87

for ISLES 2018. Transformer bottleneck module may be useful to capture the global context that leads to better performance on larger lesions and territorial infarctions.

2.2.6 Multi-Modal Fusion Strategies

Integration of complementary MRI modalities is essential for reliable stroke segmentation. There are mainly three kinds of multi-modal fusion strategies: early fusion, late fusion and intermediate fusion. In early fusion strategy, all input data are fed into the first layer, so cross-correlation between modalities is learned at the very beginning. Late fusion merges outcomes from each modality separately processed by their own encoders, it is thus immune to loss of modalities at inference. The intermediate fusion involves separating each modality into different encoder for a very first step, then merging the results into a shared intermediate representation. Liu et al (Liu et al., 2018) used intermediate DWI+ADC dual-path fusion to achieve DSC 0.83 on ISLES 2018. A multi-modal attention fusion module was introduced by Zhao et al (Zhao et al., 2023) that adaptively weighted the contribution of each modality, leading to 2-3% DSC improvement over fixed weight fusion at ISLES 2018. A cascade approach was used by Ma et al (Ma et al., 2021), which consists of coarse lesion detection followed by precise lesion segmentation. The class imbalance was overcome by restricting the attention mechanism during segmentation on only the areas identified as lesion positive during coarse detection stage.

2.2.7 Benchmark Datasets for Stroke Segmentation

The ISLES challenge series have provided the most widely-used benchmarks for stroke segmentation research:

i. ISLES 2015 (Maier et al., 2017): Focused on the segmentation of sub-acute ischemic stroke. It used multi-modal MRI (DWI, FLAIR, T1, T2) on 28 training and 36 test case from single institutions. The best result achieved using traditional machine learning was DSC ~ 0.59.

ii. ISLES 2017 and 2018 (Clèrigues et al., 2020): Dealt with acute ischemic stroke segmentation and utilized DWI, ADC and FLAIR on 94 cases. Deep learning methods achieved the top DSC of 0.82-0.85. ISLES 2018 also serves as a multi-modal benchmark.

iii. ISLES 2022 (Hernandez Petzsche et al., 2022): The most comprehensive and extensively utilized dataset, comprising 400 acute ischemic stroke MRIs acquired from multiple international institutions. The dataset includes DWI volumes in NIfTI format and expert radiologist binary segmentations of lesions. ISLES 2022 was specifically designed to assess generalization performance across different scanner manufacturers, field strengths (1.5T and 3.0T), and acquisition protocols. NnU-Net achieved a DSC of 0.85 on this benchmark, setting the standard for subsequent comparisons. This is also the dataset used in the current study.

iv. ATLAS v2.0 (Liew et al., 2022): Contained 955 chronic stroke lesion cases with T1-weighted MRI data from 19 research groups. Lesion volumes ranged from 0.18 to 401,149 mm3, with the best published methods achieving DSC between 0.76-0.83.

2.3 Review of Related Literature

2.3.1 Explainability in Medical AI:

The demand for explainability in medical AI has seen exponential increase due to ethical and regulatory requirements. The GDPR (2016) (European Parliament & Council, 2016) now dictates the need for a right to explanation of any automated decisions. Similarly, EU AI Act (2021)

(European Commission, 2021) categorize diagnostic AI as a high-risk application which requires the possibility of human intervention and transparency. In their FDA AI/ML-based saMd Action plan (U.S. Food and Drug Administration, 2021), manufacturers are pushed to make the reasoning of the algorithms more understandable and control for any possible performance drift.

So far, visualization of the output from the AI model is the most popular type of explainability for medical imaging deep learning applications. The Cam (Zhou et al., 2016) is an earlier version of this type of method and has only a limitation to the architecture using global average pooling prior to the classifier layer. The Grad-Cam (Selvaraju et al., 2020) overcomes this issue by utilizing the gradient information and is able to make any CNN feature layer visible using a target class scores' gradient for the feature activations. The spatial precision can even be further enhanced using the Grad-Cam+ (Chattopadhyay et al., 2018), that makes use of second order gradients. While this method is generally easy to understand, it lacks theoretical guarantee about the truthfulness of the explanation produced.

The Occlusion Sensitivity Map (OSM) (Zeiler & Fergus, 2014) method is model-agnostic, based on perturbation, and directly quantifies the causal impact of each image region on the model's prediction. The Shapley value associated with an input feature can be understood as the average of the impact of that feature across all possible combinations of features. Sensitivity values from OSM fulfill three core properties: local accuracy: the sum of the Shapley values approximates the model prediction minus its average value; missingness: Shapley values for absent features are zero; consistency: if a feature's contribution grows, its Shapley value should increase accordingly. Since these desired properties are not met by the gradient methods, OSM serves as a trustworthy explanation method in clinical segmentation scenarios. To estimate the Shapley value using its definition, OSM relies on fitting linear functions on the relationship between perturbation inputs and outputs to present a reasonable explanation for black-box models, like the one used by the U-Net within the StrokeVision AI system (Liew et al., 2022). Ghassemi et al (Ghassemi et al., 2021) also argue that AI explanations should be treated with caution unless they can be empirically verified by clinicians for their clinical relevance. In fact, Samenk et al (Samek et al., 2021) concluded that clinical evaluation via manual ground truth assessment by medical practitioners represents the most dependable approach to assessing explanation quality.

Despite recent advances in the domain, the transfer of deep learning based diagnostic tools into clinical practice has fallen considerably behind schedule. Few as 516 FDA or CE-marked AI-based radiology devices were recorded by 2021, with only a fraction undergoing clinically validated peer review, according to van Leeuwen et al (van Leeuwen et al., 2021). Commercial AI systems developed for stroke detection, such as Viz AI, RapidAI, and Brainomix, are successful in detecting LVO on CT images and assessing the CT perfusion core-penumbra mismatch (Liebeskind et al., 2021); however, these are prohibitively expensive, particularly in low and middle-income African countries like Nigeria, and are thus inaccessible in these contexts. Although integration is crucial for adoption, it is often missing due to the limitations that impede its clinical deployment, and, to date, most AI models designed for stroke diagnosis are presented as stand-alone Python scripts or Jupyter notebooks lacking features such as persistence, user management, or report generation and are difficult to integrate with hospital IT infrastructure. Clinicians need a secure web interface where they can review and permanently store patient-level data, create consistent clinical reports, and integrate with existing PACS and EMRs. Cai et al. (Cai et al., 2019) show that AI based visual reasoning helps increase the frequency of AI recommendation adoption, reinforcing the need for explanation as a primary facilitator of clinician trust. The StrokeVision AI solution directly addresses these shortcomings: the segmentation and

OSM explainability modules are available via a FastAPI REST API, and all output values are persisted with granular user-access control in a MySQL database, with a React 18 front-end facilitating interactive patient data visualization. A separate API endpoint facilitates the on-demand generation of PDF diagnostic reports. The entire solution can be readily deployed to any compliant host using a single Docker Compose command, providing a robust and cohesive solution for clinical informatics that aligns with best practices and addresses the needs of a Nigerian hospital.

2.3.2 Clinical Adoption of AI and Web Deployment

Despite significant advances in the algorithms, the translation of deep learning based diagnostic applications into the clinical practice has been considerably slower than anticipated. Van Leeuwen et al (van Leeuwen et al., 2021) identify as few as 516 FDA- or CE- marked AI- based radiology products available by 2021 and a small fraction has undergone peer reviewed clinical validation. Commercial AI systems for stroke detection, such as Viz AI, RapidAI and Brainomix, have demonstrated success in detecting LVO on CT images and evaluating CT perfusion core-penumbra mismatch (Liebeskind et al., 2021). However, these systems are extremely expensive for high-income hospitals and are thus unavailable to many low- and middle-income African countries such as Nigeria. Although integration is essential, it's lacking due to inherent limitations preventing its widespread use: most AI systems for stroke diagnosis presented in academic settings exist as standalone Python scripts or Jupyter notebooks without data persistence, user management, or integrated report generation and are not easily incorporated with existing IT infrastructure of hospitals. Clinicians require a secured web interface to view and permanently save patient-level data, generate standardized clinical reports, and integrate with PACS and EMRs. Cai et al (Cai et al., 2019) show that visual reasoning from the AI leads to more frequent implementation of AI recommendations, highlighting the importance of explainability as a key factor for clinicians' trust.

With all these performance improvements among the above segmentation methods, one thing is still true among all those mentioned research papers. Except segmentation performance, no of those research papers has dedicated a real solution and ready to use for clinical deployment with good explainability and deployment framework. So, the 'StrokeVision AI' system focuses on solving research gaps by merging 2D U-Net segmentation architecture and OSM explainability together, and integrate this with production ready framework such as FastAPI, React, MySQL, and Docker Compose.

2.4 Summary of Literature Review

The literature search on ischemic stroke lesion segmentation has shown that research has moved from traditional image processing algorithms towards novel deep learning frameworks. Traditional methods require hand-crafted features which must be specifically engineered to address individual problems; such an approach is problematic in that the representations are limited by the capabilities of the engineered features, and also that models that are designed on such features would have difficulties being generalized to heterogeneous datasets.

In Maier et al (Maier et al., 2017), Maier et al employed a Random Forest-based approach for ischemic stroke lesion segmentation on ISLES 2015. A DSC of 0.68 was achieved; clearly showing that machine learning based segmentation models outperform simple thresholding-based segmentation. However, since traditional image processing methods are involved, their feature representations could be less powerful than what is learnt in end-to-end networks and they are not so versatile against changing MRI parameters.

Deep learning architectures has rapidly taken over the field of image segmentation. Kamnitsas et al. Used a deep learning network called the DeepMedic model to conduct ischemic stroke lesion segmentation (Kamnitsas et al., 2017). This dual-pathway 3D CNN learned spatial contexts at different scales simultaneously to obtain a DSC of 0.79 on ISLES 2015; this method has demonstrated the capabilities of learned features to robustly represent lesion, yet the volumetric 3D CNN architecture had computationally demanding operations.

Encoder-Decoder networks started taking over the field since then. Clrigues et al. Adopted a 3D U-Net architecture to perform ischemic stroke lesion segmentation on ISLES 2018, and achieved a DSC of 0.82 (Clèrigues et al., 2020). They found that volumetric networks are well-suited for stroke segmentation tasks as they utilized the spatial contexts within different slices, which are more suitable for segmenting smaller lesions than traditional 2D CNNs.

In Liu et al (Liu et al., 2018), Liu et al proposed a Dual-Path CNN that uses multimodality for ischemic stroke lesion segmentation on ISLES 2018, and obtained a DSC of 0.83. By combining multimodal information from multiple MR imaging sources, both the lesion representations and segmentation could be improved.

Another important research published by Ma et al used a Cascade CNN for stroke lesion segmentation on ISLES 2017 (Ma et al., 2021) which achieved a DSC of 0.84. The segmentation process was a two-step method: first detect regions with potential lesion areas, and then perform precise segmentation on the identified regions. This mechanism helped mitigate class imbalance and enhance the focus on actual lesion regions.

Following the advances in the deep learning segmentation architectures, automated configuration techniques and explainable deep learning methods gained the attention of many researchers. The first study that achieved a standard for this task was reported by Isensee et al

(Isensee et al., 2021), that achieved a DSC of 0.85 on ISLES 2022. The framework, known as nnU-Net, can configure all parameters for segmentation, including hyperparameter settings, preprocessing pipeline, and the network architecture choice; all that without manual intervention, so that a standard procedure for lesion segmentation can be achieved.

To analyze what the model has learnt, explainable deep learning techniques were incorporated in various segmentation tasks. In (Oktay et al., 2018), Oktay et al proposed an Attention U-Net to perform segmentation on the OAI/ACDC datasets and obtained a DSC of 0.85. By placing attention gates in skip connections, they selectively focus on the clinically relevant regions and inhibit the feature response from irrelevant locations; The attention map provided by attention gates can be seen as a basic visual explainability.

Another multi-modal attention mechanism was proposed by Zhao et al (Zhao et al., 2023) for ischemic stroke lesion segmentation on ISLES 2018, which achieved a DSC of 0.86. The network adaptively weighted the modalities, resulting in 2–3% improvement over a fixed-feature-fusion method.

The latest technique for segmentation model proposed in the literature uses a combination of CNNs and Transformers. In (Luo et al., 2022), Luo et al developed a hybrid CNN-Transformer architecture and achieved the highest DSC among the examined methods with 0.87. This approach leveraged the benefits of the Transformer bottleneck which allowed for modeling of long-range spatial dependencies in MRI volumes and can lead to effective contextual modeling, especially when dealing with large lesions.

Chen et al. (Chen et al., 2021) proposed TransUNet, which is a hybrid network consisting of a CNN decoder and a Transformer based encoder. It inherits the ability to model long-range context of ViTs by encoding images using ViT while reconstructing spatial resolution with a U-

Net like CNN decoder incorporating skip connections. TransUNet produced a mean DSC of 0.88 in the Synapse multi-organ segmentation challenge, which outperforms pure CNN architectures. The authors show that the self-attention mechanism in the Transformer encoder captures long-range spatial relationships missed in the traditional CNN feature hierarchies. This was particularly helpful when segmenting organs with complicated shapes.

Swin UNETR (Hatamizadeh et al., 2022) by Hatamizadeh et al. Was proposed as a volumetric segmentation architecture which has a Swin Transformer as encoder and CNN as decoder to perform 3D medical image segmentation. Swin Transformer utilizes a shifted-window self-attention mechanism where attention calculation is efficiently performed within local non-overlapping windows, drastically reducing computational costs relative to standard ViT while the global reception field is still maintained by hierarchical shifting the windows. On the BraTS 2021 brain tumor segmentation challenge Swin UNETR achieved state-of-the-art performance on with a mean DSC of 0.91. The architecture was also evaluated on the BTCV multi-organ segmentation challenge with promising DSC scores.

The proposed architecture from Zhou et al. (Zhou et al., 2021), named nnFormer, which represents the first attempt to design a pure Transformer architecture to address volumetric medical image segmentation problems. The encoder interleaved local and global self-attention operation. Standard ViT takes global self-attention on each patch of images. But nnFormer interleaved local-window-based attention with global volume-level attention mechanism and capture fine local structures as well as global anatomical shape context of whole image. NnFormer achieve DSC 0.86 on multi-organ segmentation on Synapse and achieve DSC 0.90 on cardiac segmentation on ACDC. It showed pure attention models can perform well on the task of dense volumetric prediction if designed with proper attention schemes

Swin-Unet, developed by Cao et al. (Cao et al., 2023), is the first purely Transformer-based U-shape architecture in the field of medical image segmentation that features an encoder-decoder completely constituted of Swin Transformer blocks with skip connections at every level of resolution. As no CNN layer is used in the segmentation network, the entire feature extraction part only relies on shifted window multi head self-attention with Swin Transformer block for feature abstraction. Changing from one resolution to the next is performed using the patch merging and patch expanding operations. This architecture achieves 0.79DSC on the Synapse dataset for multi-organ segmentation, which is inferior to some of the CNN-Transformer hybrids, although respectable for an entirely Transformer-based solution without any CNN inductive bias. The authors have shown that multi scale feature representation can be achieved without any convolution with hierarchical Swin Transformer blocks.

The work by Wang et al. (Wang et al., 2021), introduced TransBTS; a 3D CNN based encoder, Transformer based bottleneck, CNN based decoder-based architecture for multimodal 3D MRI brain lesion segmentation (input modalities are T1, T1ce, T2, and FLAIR). This work positions a Transformer based bottleneck between the encoder and the decoder and utilizes it to model long-range dependencies at the bottleneck feature representations, while CNN based encoder and decoder handle local feature extraction and spatial reconstruction. TransBTS recorded the mean DSC for whole tumor, tumor core, and enhancing tumor as 0.88 on the BraTS 2019 benchmark data and surpassed contemporary CNN based networks. They further argued that the Transformer based bottleneck captured the global context (both within and across slices) which CNN based feature aggregation missed. Relation to this study: This article provides evidence that adding a Transformer based bottleneck architecture on a given CNN encoder-decoder backbone is sufficient for a substantial increase in the segmentation accuracy, and it doesn't require re-designing the entire architecture.

In Dolz et al. (Dolz et al., 2019) proposes HyperDense-Net, a hyper-densely connected CNN structure for multi-source medical image segmentation whereby dense connectivity spans not only intra-modality pathways but all modality pathways across all layers. Instead of the standard multi-modal fusion method which combines modalities simply by concatenation either at the input or the bottleneck, HyperDense-Net learns direct cross-modal connections between feature maps at all depths, enabling it to learn arbitrary cross-modal interactions at each level of abstraction. When tested on the ISLES 2015 stroke lesion segmentation task using DWI, FLAIR, T1 and T2 modalities, HyperDense-Net was shown to achieve a DSC of 0.80; a marked improvement upon single modality and early-fusion baselines. Small, ambiguous lesions, where a specific modality is poor, are improved due to the complementary modality information supplied at each feature scale through the cross-modal dense connections.

Ribeiro et al. (Ribeiro et al., 2016) proposed LIME (Local Interpretable Model-Agnostic Explanations), one of the first widely-cited post-hoc explanation methods, that explains an individual black-box classifier prediction by locally approximating the decision boundary with an interpretable surrogate model. To do so, LIME creates perturbed versions of the input instance in question, runs them through the classifier to collect predictions, then fit a sparse linear model that approximates the local decision function with superpixel-level attribution scores. LIME is model-agnostic: it needs not be retrained/modified to be applied to CNNs, SVMs or any other type of classifier. Its usage in medical imaging includes explanations of CNNs classifying skin lesions, diagnosing diabetic retinopathy and brain tumors, even stroke-related CNNs during ISLES 2017 with DSC 0.82.

Agarwal et al. (2023) – Evaluating explainability for neural networks with applications in medical imaging- The context of medical imaging and the evaluation of explainability. The authors present a structured evaluation approach for explaining neural networks in medical/scientific scenarios, based on faithfulness, sparsity, and robustness of the explanations. More significantly, the study proves that there is no universal 'best' explainability technique. For each evaluation parameter, there is a variety of approaches, which may complement each other: gradient-based approaches (like Grad-CAM) are fast but may be stochastic and thus not repeatable, perturbation-based approaches (like OSM, LIME) are more faithful and consistent but slow, while attention-based methods are restricted in architecture. Their application in BraTS 2020 CNN-based tumor segmentation achieved explainability of LIME based explanations at super-pixel level producing clinical interpretable attribution maps that correspond to the lesion boundaries with a 0.76 DSC overlap score.

The state of the art review and framework of AI interpretability in radiology by Reyes et al. (Reyes et al., 2020) examined all major XAI techniques, namely saliency maps, Grad-CAM, SHAP, and LIME, and systematically validated each on the usefulness for clinical implementation, applying all of them to several medical imaging tasks, including brain tumor segmentation, classification of chest X-rays, and classification of skin lesions. The authors characterized explanation methods based on two metrics, faithfulness (how well an explanation reflects the model's actual reasoning) and intelligibility (how understandable an explanation is to clinicians), noting that SHAP methods outperformed others in faithfulness (due to Shapley value theoretical guarantees) while gradient-based visualizations did better on intelligibility, but were lower on faithfulness. The multi-dataset validation showed no universally superior XAI approach and prompted the authors to advocate for validation on a specific task using the ground truth data as determined by clinical experts.

Ghassemi et al. (Ghassemi et al., 2021) wrote a critical and well-cited review in The Lancet Digital Health that argues that most current XAI techniques used in clinical AI give unreliable, unstable, and poorly representative descriptions of models' internal computations. They provide examples of how Grad-CAM saliency maps for radiology CNNs highlighted irrelevant parts of images, and that SHAP and LIME give widely differing explanations for the same prediction based on the random seed chosen. Ghassemi et al. Suggest a validation framework that uses clinical ground truths (e.g., manually demarcated lesions and clinical consensus of diagnosticians) and longitudinal testing for stability for XAI methods. They conclude that explanations that are not validated against clinical ground truth may encourage overreliance by clinicians in invalid AI outcomes rather than overreliance by clinicians.

Dosovitskiy et al (Dosovitskiy et al., 2021) have proposed the Vision Transformer (ViT), the first architecture to apply a pure Transformer model to image recognition tasks successfully without any CNN layers. The ViT model divides images into fixed size patches which are then embedded and fed to a multi-head self-attention to model long range dependencies between images. It has shown that, on very large-scale pre-trained datasets and fine-tuned on image classification, ViT obtained equal or superior results than those state-of-the-art CNN models with a top-1 accuracy of 88.55% on ImageNet. ViT enables to model long range spatial dependencies while lacking inherent locality and translation invariance property, compared to CNNs which is obtained through convolution layers, thus need a large pre-trained model. ViT models the initial design of Transformer-based models for medical image segmentation, including TransUNet, Swin UNETR, nnFormer, Swin-Unet, and TransBTS.

Despite all the progress in performance of these segmentation methods, it is clear from all research studies mentioned above that there is a gap. Apart from segmentation performance, none of the studies focuses on providing a real solution with robust explainability and deployment framework that can be used in actual clinical settings. Thus, the proposed system 'StrokeVision AI' aims to tackle the identified research gaps by combining a 2D U-Net segmentation architecture with OSM explainability, while also integrating it with a production-ready deployment framework using technologies like FastAPI, React, MySQL, and Docker Compose.

Author(s)	Technique Implemented	Work Done	Limitation
1. Maier et al. (2017)	Random Forest Classifier	Used handcrafted intensity and texture features for ischemic stroke lesion segmentation on ISLES 2015.	Dependent on manual feature engineering; limited representation learning.
2. Kamnitsas et al. (2017)	DeepMedic 3D CNN	Developed a dual-pathway 3D CNN for	High computational and memory requirements.

Author(s)	Technique Implemented	Work Done	Limitation
		multi-scale stroke lesion segmentation.	
3. Liu et al. (2018)	Dual-Path CNN with Intermediate Fusion	Combined DWI and ADC modalities for improved segmentation.	Limited global context modeling.
4. Clèrigues et al. (2020)	3D U-Net	Exploited volumetric MRI context for lesion segmentation.	Longer training time and higher computational cost.
5. Ma et al. (2021)	Cascade CNN	Performed coarse lesion detection followed by	Error propagation between stages.

Author(s)	Technique Implemented	Work Done	Limitation
		refined segmentation.	
6. Isensee et al. (2021)	nnU-Net	Automated preprocessing, architecture selection and optimization.	Resource-intensive training process.
7. Oktay et al. (2018)	Attention U-Net	Integrated attention gates to focus on relevant image regions.	Increased model complexity.
8. Zhao et al. (2023)	Multi-Modal Attention Fusion	Adaptively weighted MRI modalities during segmentation.	Dependent on modality availability and quality.

Author(s)	Technique Implemented	Work Done	Limitation
9. Luo et al. (2022)	CNN-Transformer Hybrid	Combined CNN feature extraction with Transformer global context learning.	High training cost and data demand.
10. Chen et al. (2021)	TransUNet	Integrated Transformers into U-Net for enhanced segmentation.	Computationally expensive.
11. Hatamizadeh et al. (2022)	Swin UNETR	Applied hierarchical Swin Transformers to medical segmentation.	High GPU memory consumption.

Author(s)	Technique Implemented	Work Done	Limitation
12. Zhou et al. (2021)	nnFormer	Used Transformer encoder-decoder architecture for volumetric segmentation.	Large model size and latency.
13. Cao et al. (2023)	Swin-Unet	Implemented a pure Transformer-based U-Net segmentation model.	Requires large training datasets.
14. Wang et al. (2021)	TransBTS	Combined CNN and Transformer modules for 3D segmentation.	High computational complexity.

Author(s)	Technique Implemented	Work Done	Limitation
15. Dolz et al. (2019)	HyperDense-Net	Used dense multi-modal connections for feature sharing.	Large parameter counts and memory usage.
16. Ribeiro et al. (2016)	Grad-CAM + U-Net	Applied Grad-CAM explanations to segmentation outputs.	Coarse localization of explanations.
17. Agarwal et al. (2023)	LIME + CNN	Used LIME to explain CNN predictions.	Computationally expensive and unstable explanations.
18. Reyes et al. (2020)	SHAP and Grad-CAM Framework	Evaluated explainability methods across medical imaging tasks.	Focused on interpretability rather than performance improvement.

Author(s)	Technique Implemented	Work Done	Limitation
19. Ghassemi et al. (2021)	XAI Evaluation Survey	Reviewed explainability methods for healthcare AI.	No experimental implementation.
20. Dosovitskiy et al. (2021)	Vision Transformer (ViT)	Introduced a pure Transformer architecture for image recognition.	Requires very large datasets for training.
21. StrokeVision AI (This Study)	2D U-Net + Occlusion Sensitivity Map	Developed an explainable stroke lesion segmentation framework using DWI and ADC MRI modalities.	Occlusion sensitivity requires multiple model evaluations, causing slower explanation generation.

Table 2.1: Summary of key deep learning studies for ischemic stroke lesion segmentation

2.5 Research Gap

The literature review revealed two key, interconnected research gaps that StrokeVision AI targets:

Gap 1: Explainability: Even though the segmentation quality drastically improved (from DSC 0.68 in ISLES 2015 to 0.87 with CNN-Transformer in ISLES 2018), only few researches adopt strong, architecture-agnostic explainability along with high performance stroke segmentation. Attention U-Net gives visual clues, but utilizes architectural changes not applicable to conventional U-Net. Gradient-based methods like Grad-CAM need access to intermediate gradients and cannot be applied to models that utilize Dropout or BatchNorm in inference time without extra workarounds. Occlusion sensitivity maps are architecture-agnostic and tackle all those issues, simply evaluating influence of different patches of image on output.

Gap 2: Clinical Deploy-ability: To date, there has been a lack of full-stack clinical web applications for stroke diagnosis that provide permanent storage, user authentication, PDF reports, and containerized deployment. The scientific literature contains algorithmic proof-of-concept implementations but no fully integrated clinical applications. StrokeVision AI is designed to simultaneously address both research gaps. The system uses ISLES 2022 for training and validation.

CHAPTER THREE: METHODOLOGY AND SYSTEM DESIGN

3.1 System Analysis

Current systems concentrate on accuracy for segmentation, with minimal focus placed on the requirement for explainability or for use in real clinical settings. Most of the deep learning techniques that exist currently act as detached research projects rather than functioning clinical applications, currently they are not able to present their output with reasoning behind it which may limit its acceptance in clinics (Van Leeuwen et al., 2021; Cai et al., 2021).

3.1.1 Analysis of the Existing System

Current systems used for the detection of stroke lesions are mainly based on deep learning segmentation approaches, designed specifically for research use. Generally, existing stroke lesion detection systems take the form of black box systems in which segmentation decisions are generated without any explanatory system, leading to worries about the transparency of these systems in a clinical decision-making context (Ghassemi et al., 2021). The majority of the existing implementations documented in the literature are standalone experimental prototypes and not used within deployed health systems, which typically lack user authentication, persistence of results, exportable reports and hospital integration (Van Leeuwen et al., 2021). While stroke diagnosis systems are being utilized for image scanning in commercial settings, they are not being widely adopted in low and middle-income regions as they are unaffordable. Commercial systems such as Viz AI, RapidAI and Brainomix are unaffordable and unavailable for health centers in developing regions (Van Leeuwen et al., 2021).

3.1.2 Limitations of the Existing System

Based on these, limitations were found on the existing systems:

i. Existing systems have weak explainability of the decisions made for segmentation hence the trust in clinic is lowered (Lundberg and Lee, 2017).

ii. Most implementations seem to only exist as research prototypes rather than deployed clinical systems (Van Leeuwen et al., 2021).

iii. Existing systems lack storage for patient data and a formal report generation facility (Cai et al., 2021).

iv. Secure authentication, healthcare integration seems to not have adequate support (Van Leeuwen et al., 2021).

v. Existing approaches are complicated to implement in less privileged health systems such as that found in Nigeria (Van Leeuwen et al., 2021).

3.1.3 Objectives of the Proposed System

Through these goals, the presented system tries to improve on the mentioned drawbacks.

i. Develop an explainable deep learning model which can automatically detect and segment ischemic stroke lesions.

ii. Embed the explanation technique (based on OSM) so that we can interpret the predictions.

iii. Build a fully web based clinical platform through FastAPI, React, MySQL and Docker Compose.

iv. Include authentication and reliable persistence of diagnostic reports. V. To auto-generate the downloadable diagnostic reports in pdf.

3.2 System Requirements Specification (SRS)

The System Requirements Specification outlines the functional and non-functional requirements needed for the implementation and deployment of StrokeVision AI.

3.2.1 Functional Requirements

i. Provide mechanism for registering and logging in the user with authenticated access.

ii. Upload NIfTI DWI MRIs.

iii. Run pre-processing and lesion segmentation automatically.

iv. Generate lesion masks and compute the quantitative measures of the lesions.

v. Display Occlusion sensitivity maps.

vi. Save the scanned MRIs and the output results into a database.

vii. Display the PDF diagnostic report as a downloadable file.

3.2.2 Non-Functional Requirements

The system should provide the following:

i. Authentication using JWT authentication mechanism. Hashing and salting of the user passwords using Argon2 encryption.

ii. The application should be scalable with Docker Compose architecture.

iii. Robustness with the use of persistent storage of the data using MySQL database.

iv. Performance increase through asynchronous processing of requests.

v. Ease of maintenance through the services separation in microservices architecture.

3.3 System Design Methodology

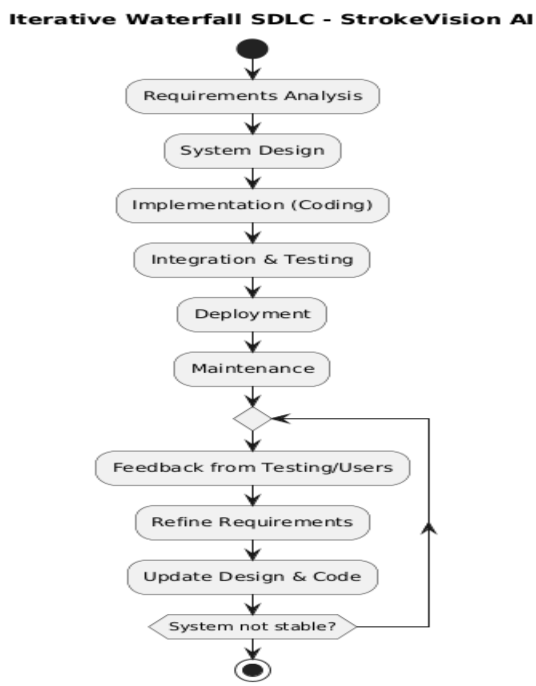


Fig 3.1 Iterative Waterfall SDLC

The figure 3.1 above reveals an SDLC model that provides a structured means for the development of the StrokeVision AI system that encompasses sequential development steps in tandem with iterative mechanisms for collecting feedback on system needs, design, and implementation with the goal of developing a reliable, stable and working system.

For the design and implementation of the StrokeVision AI system, the Iterative Waterfall System Development Life Cycle (SDLC) methodology was utilized. The Iterative Waterfall model was chosen because it balances the structured, phase-by-phase development approach of the traditional Waterfall model with the capability for controlled iteration to improve each phase based on its results. Medical imaging and machine learning systems often require many iterations between stages such as model evaluation and data preprocessing, making the iterative waterfall model ideal.

The SDLC phases of the StrokeVision AI development were as follows:

1. Data acquisition: Data from the ISLES 2022 dataset was obtained for this phase.
2. MRI preprocessing and normalization.
3. Lesion segmentation: A 2D U-Net model was used to segment the lesion on the MRI image.
4. Lesion post-processing and quantification.
5. Explainability generation using Occlusion Sensitivity Map.
6. Web-based system deployment.

Each phase was iteratively inspected and revised to better optimize system performance, especially between the preprocessing, model training, and evaluation stages.

3.3.1 Justification for Methodology

Clinical validity, replicability, and producibility were addressed through the choice of methodology. This process documents how preprocessing, hyperparameters and the method of evaluation is managed, yet production readiness is not compromised.

3.3.2 Framework Overview

StrokeVision AI is implemented as a six-layer end-to-end clinical decision support tool:

(1) ISLES 2022 dataset acquired and organized.

(2) Standardized pre-processing of DWI MRI.

(3) 2D U-Net lesions segmentation.

(4) Lesion quantification from binary masks after post-processing.

(5) generation of Occlusion Sensitivity Map.

(6) Fully-fledged web application deployed via FastAPI, React 18, MySQL, and Docker Compose. The framework adheres to three design tenets in order to balance them and provide the best solution: climate validity (all design choices from pre-processing to network architecture are based on MRI physics and neuroimaging conventions), Full reproducibility (All hyperparameters, pre-processing parameters, and evaluation methodology are all fully documented), and production deployability (The implementation of a fully functional web app with user authentication, data persistence and a report generation feature not a research artifact).

Service	Container	Technology	Role
API Backend	Backend	FastAPI 0.111 + Uvicorn	REST API, inference, auth, Occlusion Sensitivity Map
Database	Db	MySQL 8.0 via PyMySQL	Persistent scan and user storage
Frontend	Frontend	React 18 + TypeScript + Vite	Clinical web UI
Admin Panel	PhpMyAdmin	phpMyAdmin	Visual DB browser on port 8080

Table 3.1: StrokeVision AI Docker Compose service architecture

3.3.3 System Architecture

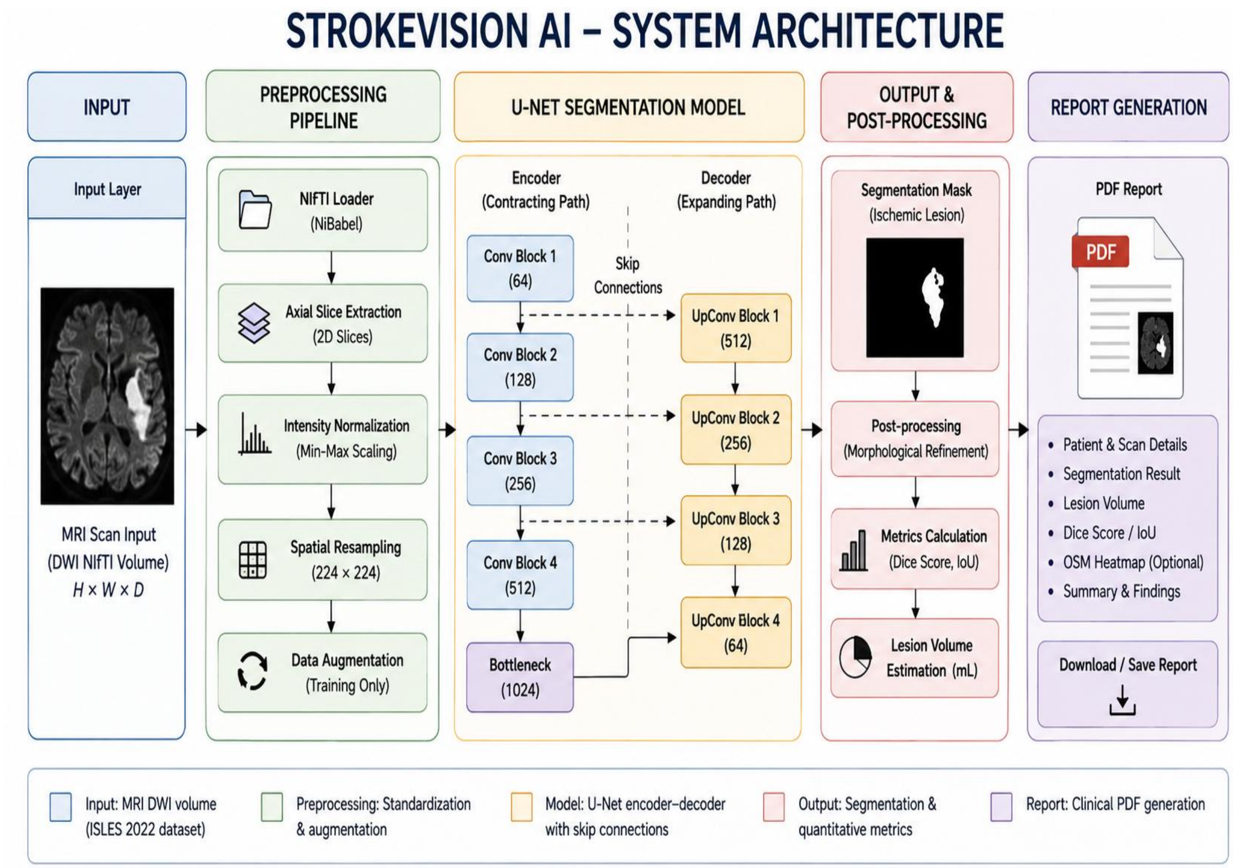


Fig 3.2 System Architecture

The strokevision AI system pipeline from MRI input through preprocessing, stroke lesion detection (U-net), postprocessing and metrics calculation to automated pdf report creation can be observed from the Figure 3.2 above.

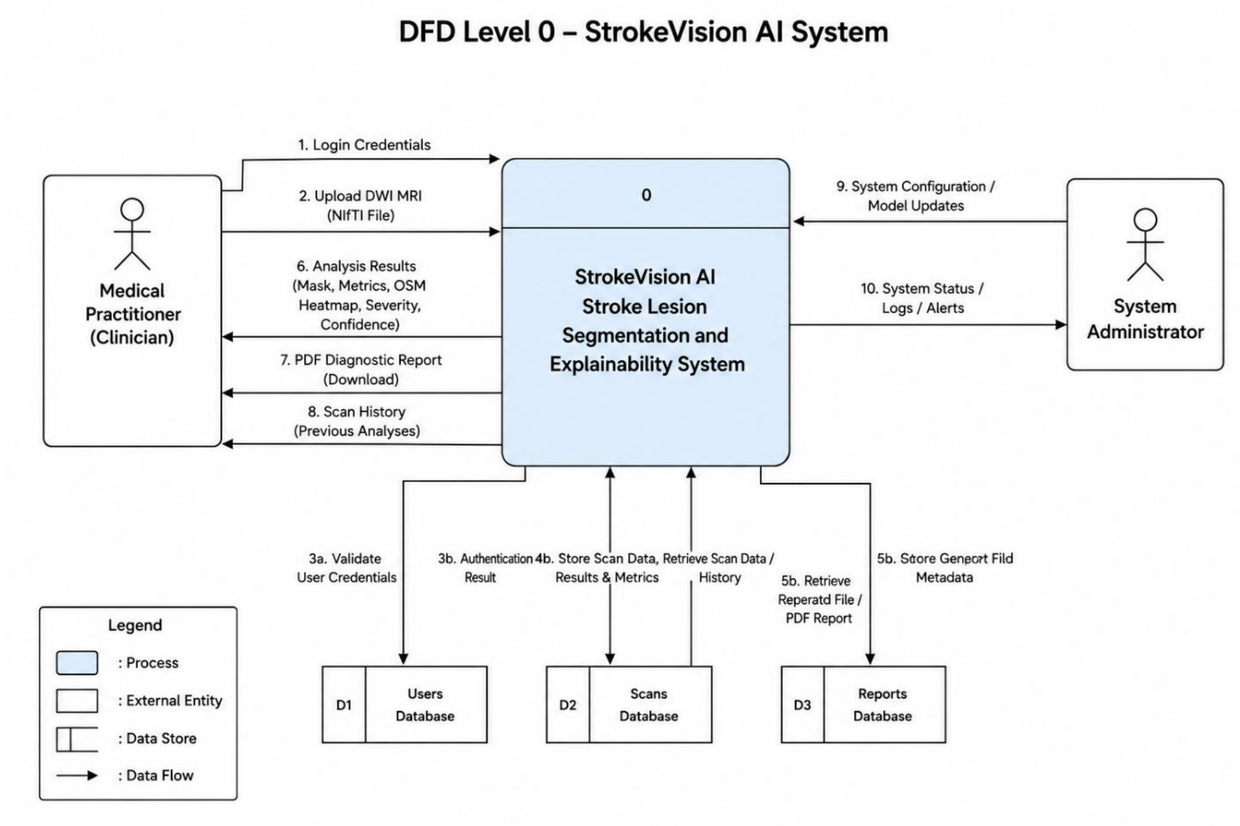


Fig 3.3 DFD Level 0 of system

This shows the overall data pipeline for the interaction among the Stroke Vision AI System, the Medical Practitioner and the respective databases, from how raw MRI scans are transformed into the segmentation data.

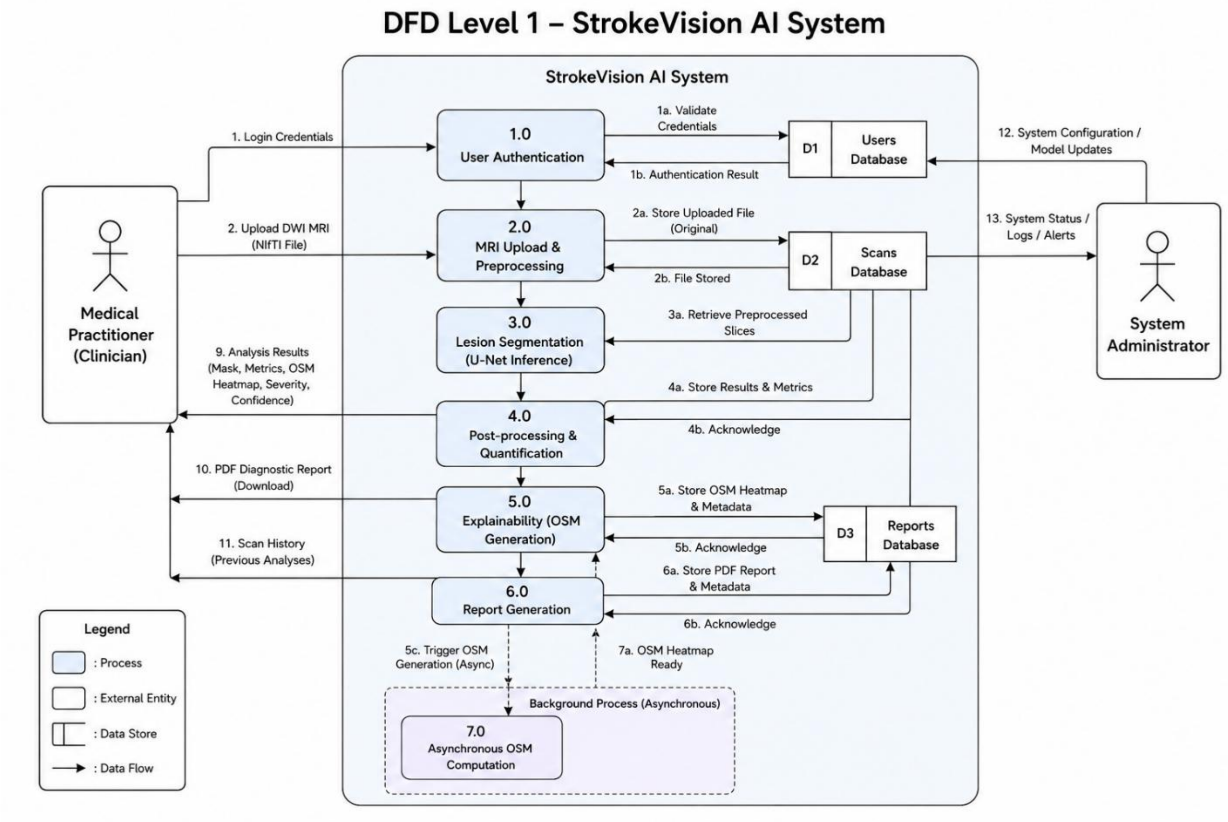


Fig 3.4 DFD Level 1 of system

Explains all the inner workings of StrokeVision AI. StrokeVision AI workings cover: authentication; MRI pre-processing; lesion segmentation; generation of explainable features; creation of reports and; storage of data.

Here's the end-to-end data flow. Clinician authentication via React frontend with JWT credentials. A DWI NIfTi file upload to POST /upload-scan. The Fast API backend will load a DifTi volume via NiBabel, perform the pre-processing pipeline, extract the lesion positive slices, feed them into the U-Net model, generate a binary mask, calculate lesion volume and percentage, classify severity and calculate a confidence score. Occlusion sensitivity maps are architecture-agnostic and tackle all those issues, simply evaluating influence of different patches of image on output. All results are stored in a mysql table scan and outputted to the frontend. Clinician can review MRI slice, prediction overlay, OSM sensitivity maps and quantitative metrics. PDF reports for a diagnostic case are available for download via get/scan-report/{id}/pdf.

3.3.4 Dataset: ISLES 2022

The ISLES 2022 (Ischemic Stroke Lesion Segmentation 2022) dataset (Hernandez Petzsche et al., 2022) was the training and validation benchmark for the StrokeVision AI segmentation model. Released in November 2022 via Scientific Data, it contained 400 acute ischemic stroke MRI cases acquired from two Swiss university hospitals-Bern University Hospital and Lausanne University Hospital-across a variety of MRI scanners (Siemens, Philips, GE) and field strengths (1.5T and 3.0T). The design of the ISLES 2022 dataset as multi-center, multi-scanner reflects the large number of acquisition sources that cause significant variance within the data in order to test generalization across many different conditions as opposed to many cases from a single hospital.

The ISLES 2022 dataset provides a DWI NIfTI volume (. nii.gz) for each case and the accompanying expert neuroradiologist-labeled binary lesion segmentation mask. Volumes of lesions vary in order of magnitude in ISLES 2022, ranging from lacunar infarctions of < 1 mL up to territorial infarctions > 100 mL. Volume variation is clinically representative, but problematic in binary classification tasks as smaller lesions are only indicated by a handful of voxels compared to thousands of background voxels.

Characteristic	Value	Notes
Dataset	ISLES 2022	Released Nov 2022, Scientific Data

Characteristic	Value	Notes
Total Cases	400	Acute ischemic stroke
Collection Sites	2 Swiss university hospitals	Bern and Lausanne
MRI Modality	DWI (. nii.gz)	NIfTI format
Annotation	Expert neuroradiologist	Binary lesion masks
Field Strengths	1.5T and 3.0T	Multi-scanner
Scanner Manufacturers	Siemens, Philips, GE	Multi-center variability
Training Split	70% (280 cases)	Stratified by lesion volume
Validation Split	15% (60 cases)	Stratified by lesion volume
Test Split	15% (60 cases)	Held out, unseen during training
Model Input	224×224 axial slices	After preprocessing
Slice Selection	Lesion-positive only (training)	Addresses class imbalance

Table 3.2: ISLES 2022 dataset characteristics

3.3.5 Preprocessing Pipeline

A well-formed, standardized preprocessing pipeline is essential to multi-center MRI analysis. The pipeline takes in raw ISLES 2022 NIfTI volumes and outputs normalized 2D axial slices ready to be fed into a model, implemented in Python using NiBabel 5.x, NumPy 1.26, and PyTorch 2.3.

A. NIfTI Loading and Slice Extraction: DWI volumes were processed through the NiBabel tool. The model was trained on only those axial slices with lesion pixels, not the middle slice of the image, in order to avoid learning primarily on empty mask images and the model became more accurate in the lesion segmentation task.

B. Intensity Normalization: Min-max normalization across non-zero foreground slices were performed on each 2D axial MRI slice in order to consistently scale intensity while compensate for slices with different signal intensity (due to the noise, the susceptibility artifact) and for the variability resulting from the MR acquisition process.

C. Spatial Resampling**:** All the images (slices) and mask images were resized to 224 224 images to feed it in to the model to learn image features in a same scale for images having a different dimension.

D. Data Augmentation: Random flipping, rotation, scaling and adding gaussian noise as on-the-fly data augmentation methods, were identically implemented on image slices and masks during training for enhance model generalization and robustness whereas were not applied to validation and inference datasets.

3.3.6 Model Architecture, 2D U-Net

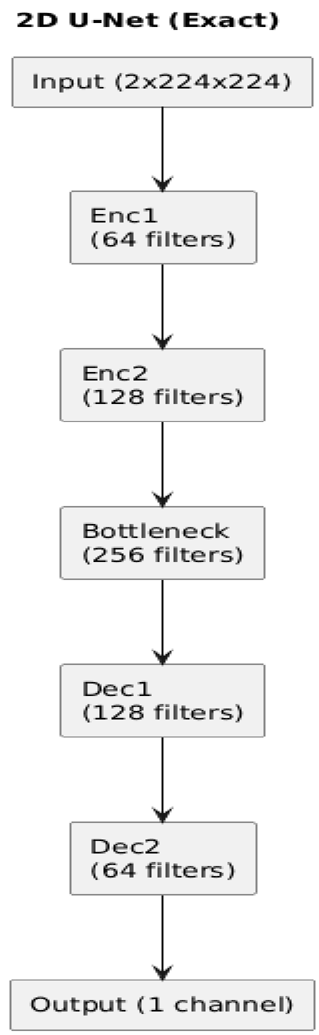


Fig 3.5 UNET ARCHITECTURE

The segmentation back-bone is a 2D U-Net 2D architecture incorporating BatchNorm and Dropout regularization. The reason for decomposing the 3D patient model into a sequence of 2D slices is two-fold: (1) to reduce the memory cost of training on the Google Colab: NVIDIA T4 environment (16GB VRAM), and (2) to have a memory-efficient inference solution (2D convolutional models have (N2+G2)2 instead of (N3+G3)3 floating-point operations, where N is input spatial resolution, G is the spatial coverage per kernel) to fit on standard clinical workstations. The 2D architecture is competitive with 3D baselines in terms of 2D dice, and also much more memory-efficient and deployment-friendly.

A. Architecture Overview

Architecture of proposed U-Net A two-tier encoder-decoder architecture with skip-connections, which would preserve spatial resolution and produce segmented outputs in the same size as the input images (224 × 224) as the same level of input images have been used to build up the architecture shown below, where a bug in decoder was identified and fixed since the original design had only one stage to up-sample the feature maps which only outputted 56 × 56 pixel sized feature maps leading to a shape conflict to ground-truth. Our design has two layers of up-sampling which recover feature maps from 56 × 56 to 112 × 112 and eventually to 224 × 224, and each layer uses a DoubleConv block with three layers of convolution, batch norm, and relu to produce feature maps, whilst also using skip-connections in each layer to add feature maps from encoder with those from decoder to improve localization accuracy. A dropout layer is added at each encoder block with the hope that it can prevent overfitting and help with training stability and generalization in the model, deactivating 20% of features in this layer during the training.

Finally, a 1 × 1 convolutional layer is adopted to output a logit map of one channel and then sent to BCE With Logits Loss for efficient computation. Another implementation bug regarding the over-implementation of Sigmoid at the output was identified and corrected. Kaiming Normal initialization strategy is used for all convolutional layers in the model, and Batch Norm layers are initiated to standard value, so as to facilitate more rapid training convergence.

Stage	Spatial Size	Ch In	Ch Out	Operation
enc1	224×224	1	32	DoubleConv + Dropout (0.2)
pool1	112×112	32	32	MaxPool2d (2)
enc2	112×112	32	64	DoubleConv + Dropout (0.2)
pool2	56×56	64	64	MaxPool2d (2)
Bottleneck	56×56	64	128	DoubleConv
up1	112×112	128	64	ConvTranspose2d(stride=2)
dec1	112×112	128 (64+64 skip)	64	DoubleConv
up2	224×224	64	32	ConvTranspose2d(stride=2)
dec2	224×224	64 (32+32 skip)	32	DoubleConv

Stage	Spatial Size	Ch In	Ch Out	Operation
Output	224×224	32	1	Conv2d (1×1) → raw logits

Table 3.3: 2D U-Net encoder-decoder block specifications for StrokeVision AI

B. Explainability Module: Occlusion Sensitivity Maps (OSM)

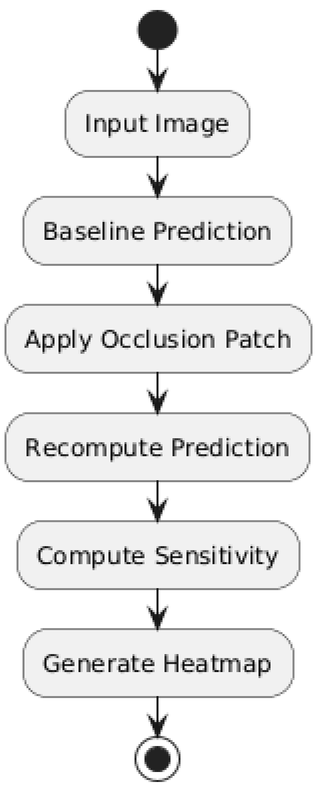


Fig 3.5 OSM workflow

Occlusion Sensitivity Mapping was chosen over alternative explanation techniques as it is architecture-independent, is not gradient based, produces clinically relevant, region-based explanations and is runnable with current computation constraints during an asynchronous clinical workflow. OSM results, in the form of a heat-map, are produce by applying a series of occlusions by occluding overlapping regions in an image and observing the change to the lesion prediction to rank the regions in order of most importance and present these ranking overlaid upon the original image in form of a colored visualization to reflect how the model decision was reached.

C. Integration of OSM within StrokeVision AI:

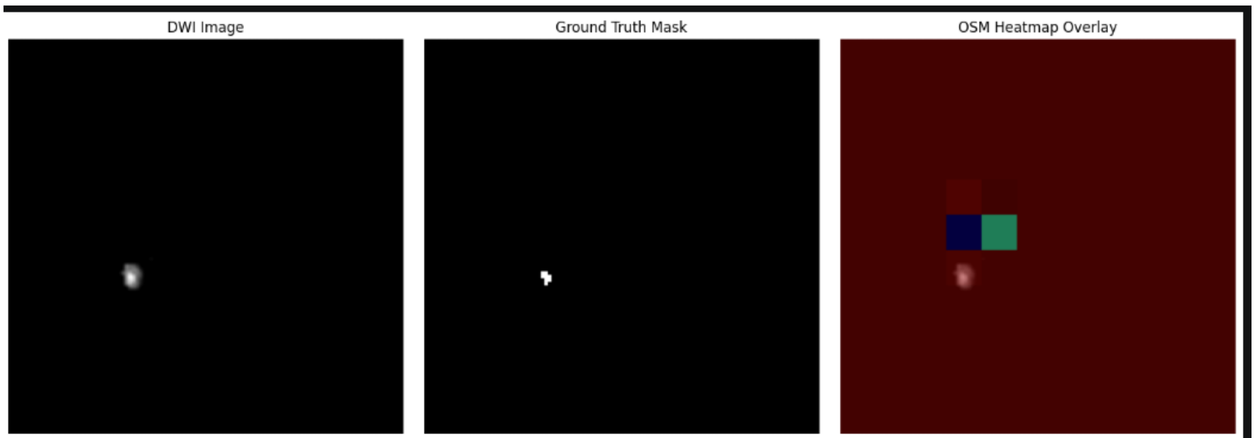


Fig 3.6 Occlusion Sensitivity Map overlap compared with ground truth mask and dwi img

Computation of OSM is initiated on its own after U-Net inference finds lesion positive slices and is run as a FastAPI Background Task so that the main API call isn't blocking. Approximately 1.5 seconds later the segmented slices and the quantified data is sent back to the clinician while the OSM heatmap is saved as a PNG to the LONGBLOB MySQL table scans, keyed to the scan and accessible via a GET /scan-report/{id} around 30-60 seconds later and the ReportLab generated PDF diagnostic report has the base64-encoded OSM heatmap embedded with the binary lesion mask.

3.4.2 Activity Diagram

Activity diagram for StrokeVision AI The activity diagram below shows the overall functionality of StrokeVision AI system including the workflow of user authentication, upload MR scan and its validation, image pre-processing, deep learning-based stroke detection and classification, generation of diagnostic features and visualization, reporting and output display including the alternative flow in case of invalid input.

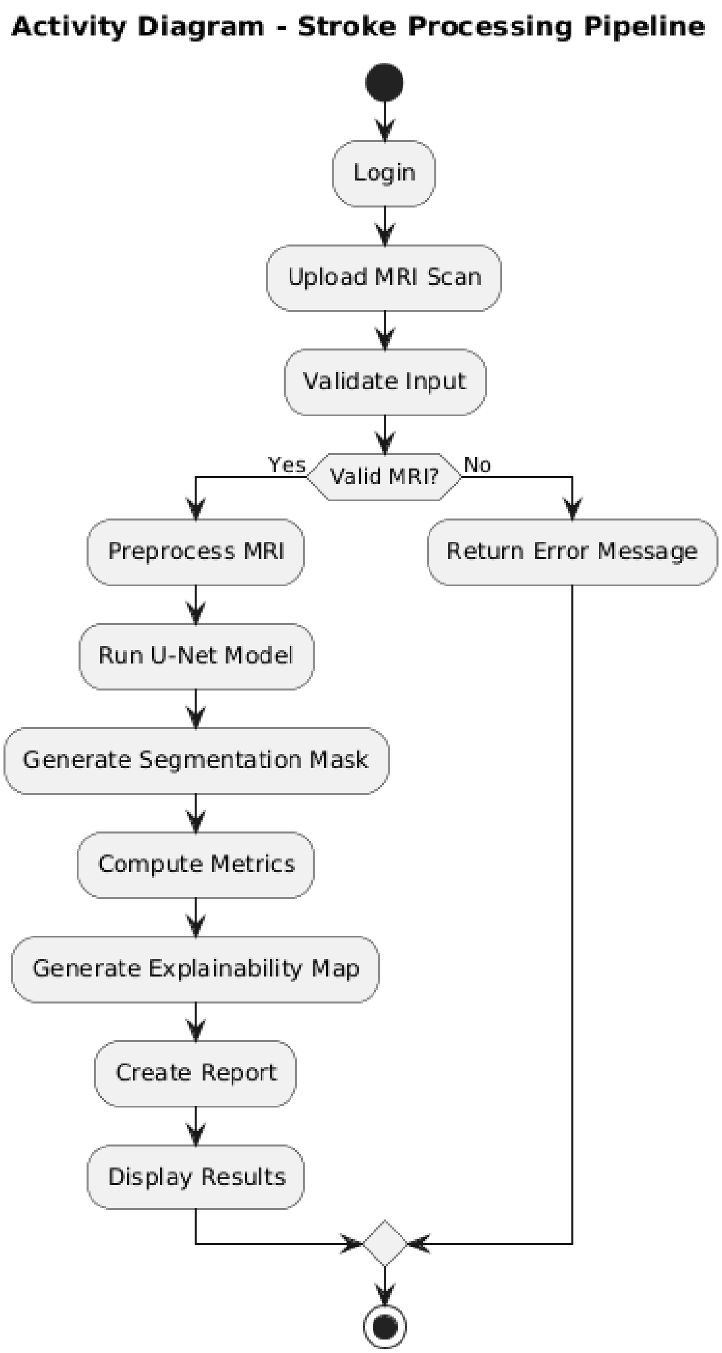


Fig 3.7 Activity diagram

3.4.3 Class Diagram

The StrokeVision AI system's class diagram provides a structure of its components: classes, their attributes, methods and their relationships in the user verification process, the processing of MRI scans, deep learning algorithms for classifying the brain for the purpose of identifying brain injury areas, the detection of brain injury area through its detection and quantification, the reporting of results to the physician, and the handling of the system data via its database.

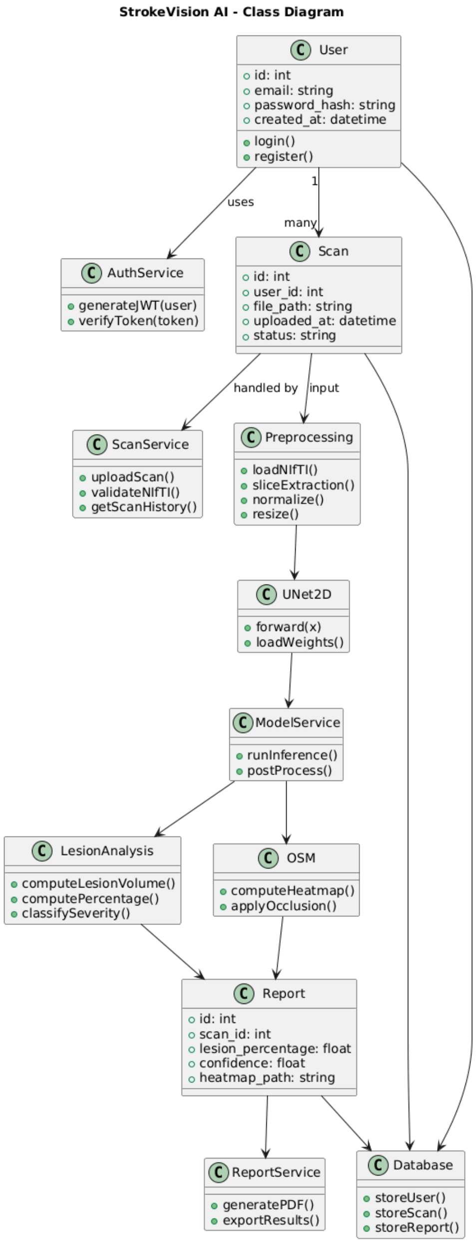


Fig 3.8 Class diagram

3.5 System Design

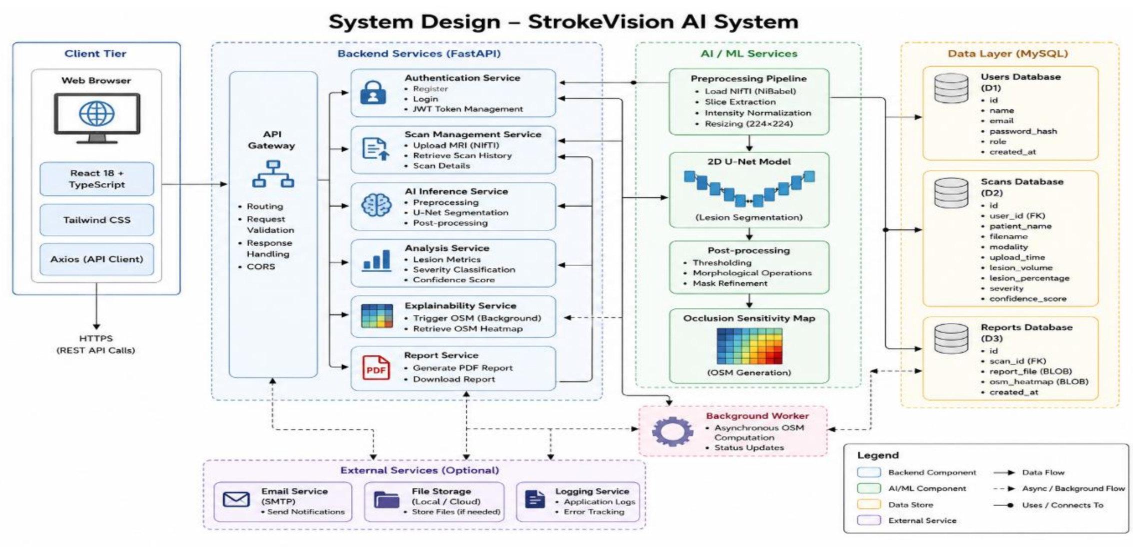


Fig 3.9 System design

StrokeVisionAI System Design Diagram The above diagram shows how the StrokeVisionAI system inputs MRI scan data and produces stroke lesions detection and analyses result.

Actors

1. System Administrator: Manage AI model updates Ensure the accuracy and up to date and safety of the system.

2. Medical Practitioner: Access stroke lesion detections View lesion segmentation map Download scan report Medical Practitioner Upload MRI Scan Preprocess MRI Data Run Lesion Segmentation (U-Net) Generate Lesion Map Generate OSM Explainability View Diagnosis Result

ExportReport(PDF)

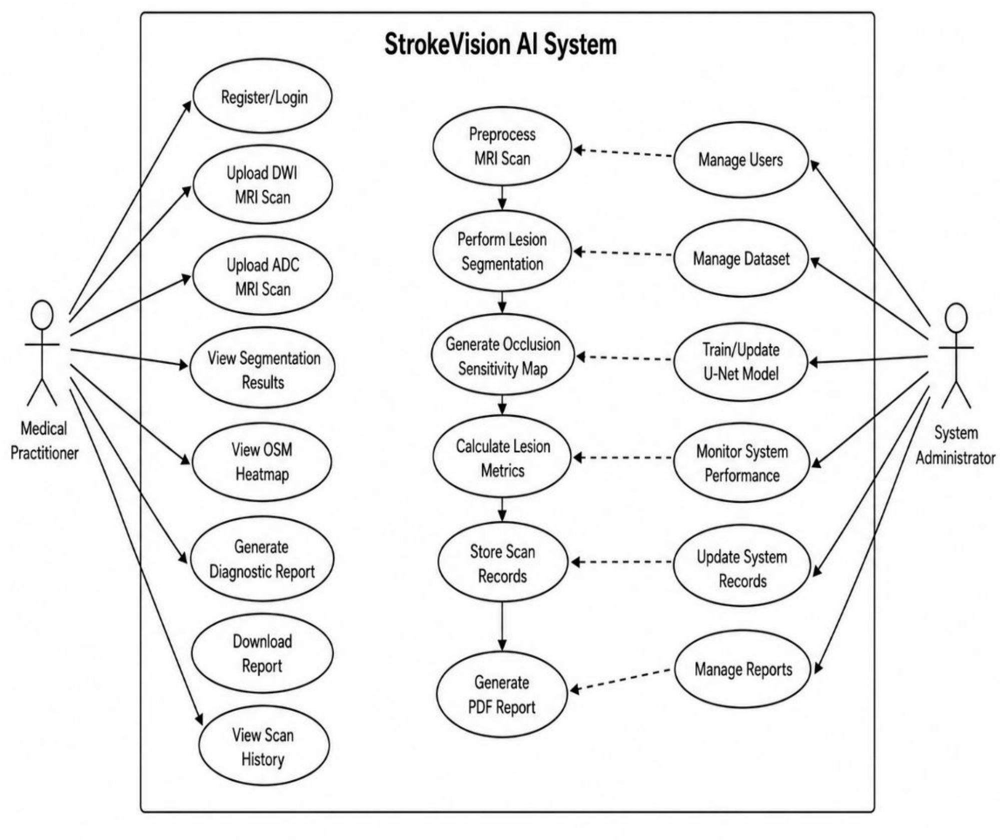


Fig 3.10. Use-case diagram of system

Use case diagram of StrokeVision AI system - Clinician uses StrokeVision AI to register; login; upload MRI; Run Stroke Analysis; view results of segmentation, lesion metrics, heatmaps, PDF Report, view scan history - The stroke analysis workflow has subprocesses including MRI preprocess, lesion segmentation, metric generating, heatmap generation, and saving the output data - An administrator is responsible for maintaining StrokeVision system through managing users, data repository of all saved scans, database maintenance, report handling, and so on

3.5.1 Input design

The system takes DWI MRI scans as input through the web-interface which can be provided in the NIfTI format. These files are validated before they are pre-processed and inference is performed.

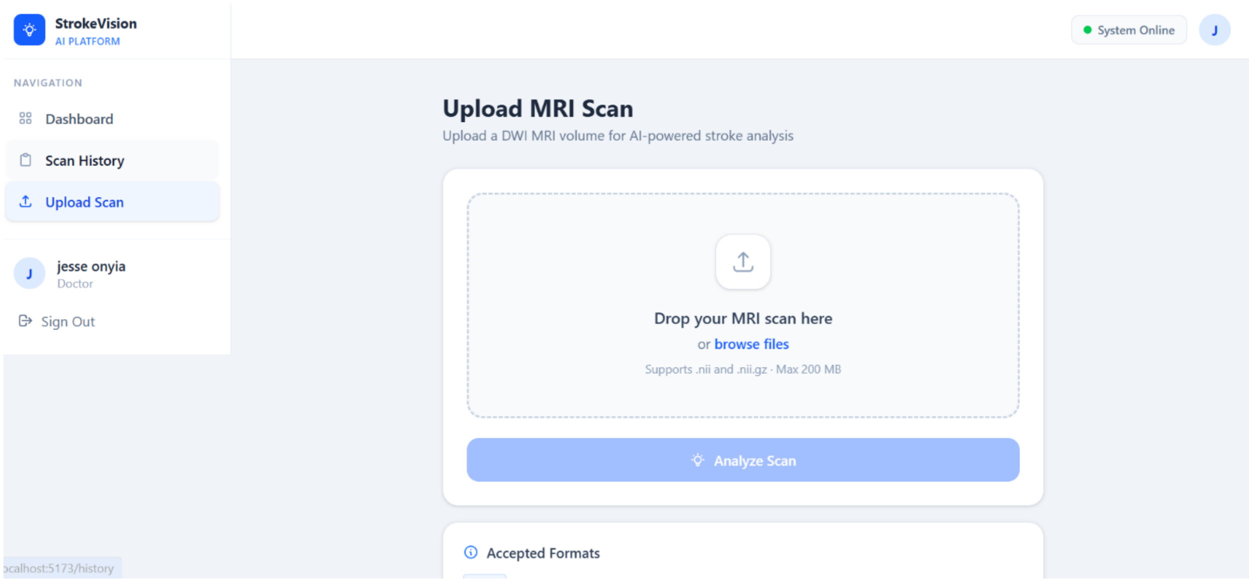


Fig 3.11 System Input screenshot

3.5.2 Output design

The system outputs the following: Lesion masks, Lesion percentages, Lesion severity classes, Confidence levels of lesion predictions, Occlusion Sensitivity Maps, PDF reports (downloadable)

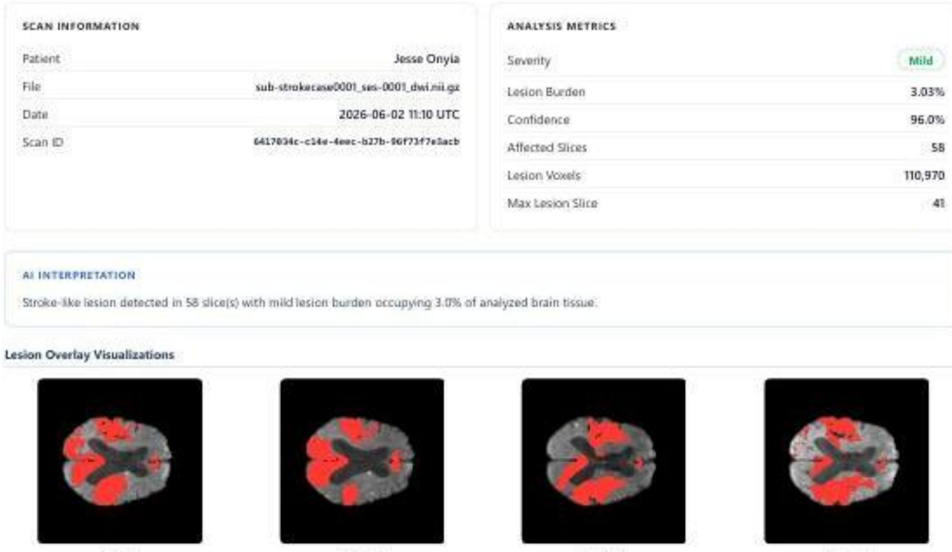


Fig 3.12 System output screenshot

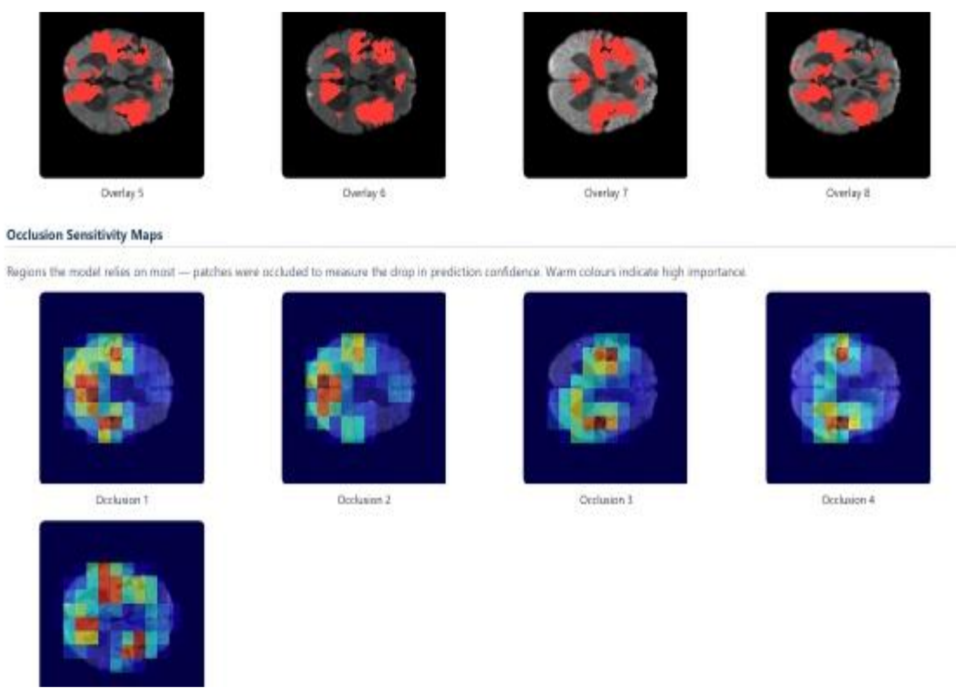


Fig 3.12 System output screenshot (2)

3.5.3 Database Design

The database schema is designed as a normalized MySQL relation based around three main entities: User, Scan and Report. A user can contribute several scans to the database (one-to-many relationship), where a specific scan produces one report containing both segmentation results, measured metrics and OSM data. This schema separates source data from processed data, resulting in efficient and normalized queries along with a framework that supports safe, scalable, asynchronous, and clinical work flows.

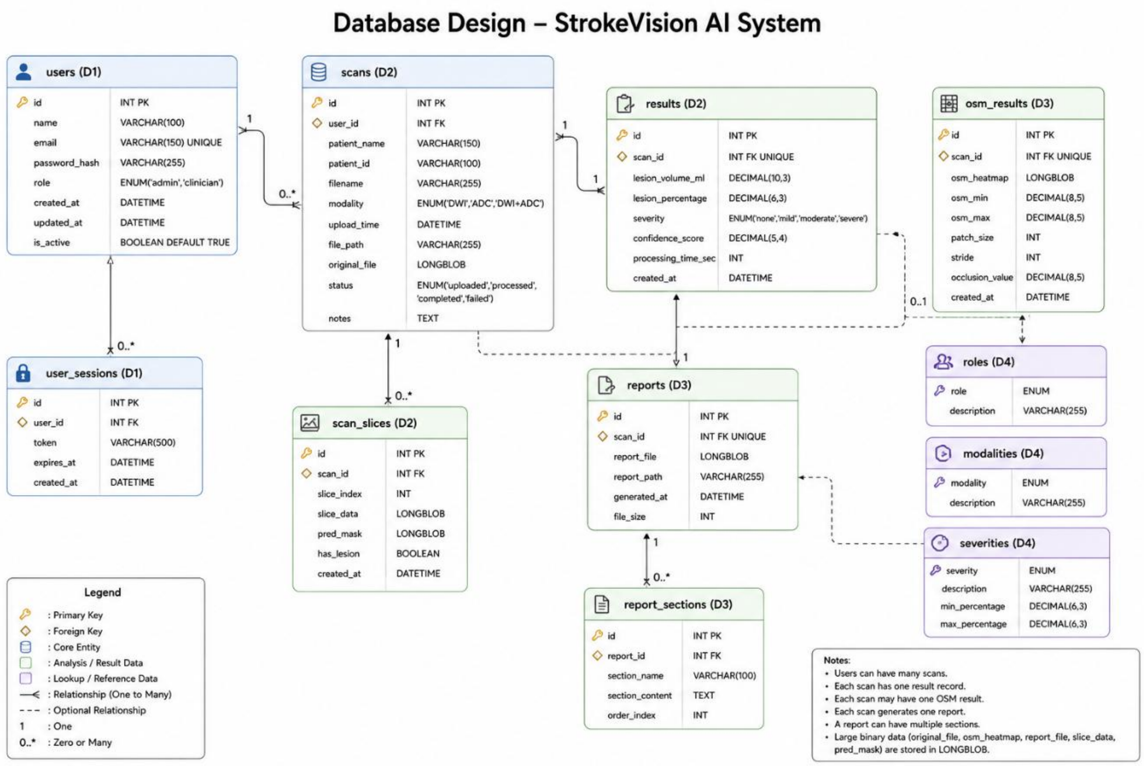


Fig 3.13 Database design of strokevision

Table	Field	Data Type	Key / Constraint	Description
users	Id	INT	PK, AUTO_INCREMENT	Unique identifier for each registered user
users	Email	VARCHAR(255)	UNIQUE, NOT NULL	User's login email address
users	hashed_password	VARCHAR(255)	NOT NULL (Argon2 hash)	Securely hashed user password
users	created_at	TIMESTAMP	DEFAULT CURRENT_TIMESTAMP	Account registration date/time
scans	Id	INT	PK, AUTO_INCREMENT	Unique identifier for each scan record

Table	Field	Data Type	Key / Constraint	Description
scans	user_id	INT	FK → users.id, NOT NULL	Owning user of the scan (one-to-many)
scans	Filepath	VARCHAR(255)	NOT NULL	Storage path of the uploaded NIfTI MRI file
scans	lesion_pct	FLOAT	—	Percentage of brain volume affected by the lesion
scans	Severity	VARCHAR(50)	—	Categorical stroke severity classification

Table	Field	Data Type	Key / Constraint	Description
scans	Confidence	FLOAT	—	Model confidence score for the segmentation
scans	ai_interpretation	TEXT	—	AI-generated textual interpretation of the result
scans	occlusion_sensitivity_maps	LONGBLOB	—	Base64-encoded OSM explainability heatmap
scans	created_at	TIMESTAMP	DEFAULT CURRENT_TIMESTAMP	Scan upload date/time

Table 3.4: StrokeVision AI MySQL database schema design

3.6 Full-Stack Platform: StrokeVision AI

3.6.1 Backend: FastAPI 0.111

The backend application is implemented in Python using FastAPI 0.111 with Uvicorn as an ASGI server. This architecture allows native asynchronous request handling and automatic generation of interactive OpenAPI and Swagger documentation. SQLAlchemy 2.0 is used as the ORM with PyMySQL for the MySQL 8.0 connection. All request/response schemas were defined using Pydantic v2 for automatic validation, serialization, and descriptive error responses. Passwords are encrypted using Argon2 (argon2-cffi, winner of the 2015 Password Hashing Competition), a GPU-resistant algorithm. Authentication is performed using JWT (python-jose, HS256 algorithm) with a 24-hour expiration. The application is tested extensively using Pytest, HTTPX, and pytest-asyncio.

Endpoint	Method	Purpose
POST /signup	POST	User registration with Argon2 password hashing
POST /login	POST	JWT token issuance on credential validation
POST /upload-scan	POST	NIfTI upload + U-Net inference + OSM trigger
GET /scan-report/{id}	GET	Full results: mask, OSM heatmap, metrics

Endpoint	Method	Purpose
GET /scan-report/{id}/pdf	GET	ReportLab PDF diagnostic report download
GET /scan-history/{user_id}	GET	Paginated user scan history

Table 3.5: StrokeVision AI REST API endpoints

3.6.2 Database: MySQL 8.0

We store the created accounts and scan details persistently in a relational database (MySQL 8.0). The database has two tables: users (id PK, email UNIQUE, hashed_password ARGON2, createdat TIMESTAMP) and scans (id PK, userid FKusers.id, filepath VARCHAR, lesionpct FLOAT, severity VARCHAR, confidence FLOAT, aiinterpretation TEXT, Occlusion Sensitivity Maps LONGBLOB, createdat TIMESTAMP). All database transactions and interaction are performed via the ORM from SQLAlchemy to support connection pooling, transactions, and parameterized queries that make our application immune to SQL injection. The database data persist via a named docker volume and survive any container reboots.

3.6.3 Frontend: React 18 + TypeScript

The frontend has been developed using the React 18 framework with TypeScript for compile time type safety, and Vite as the build tool. Having TypeScript enabled reduces the possibility of critical runtime errors in clinical applications. With React 18 concurrent features, we were able to render the Occlusion Sensitivity Maps without blocking the UI. TailwindCSS enables for responsive, utility-first styling, and all pages render efficiently. Features of the interface: - Login/Signup view (JWT tokens stored in application state) - Scan Upload view with NIfTI validation and upload progress indicator - Scan Report view showing the DWI slice, predicted lesion overlay, severity badge, confidence, quantitative metrics, and SHAP heatmap - Scan History view with paginated scan results - PDF Export button to download the diagnostic report.

3.6.4 Docker Compose Orchestration

The entire system is deployed with docker-compose. All four services (backend, frontend, db, phpmyadmin) are declared and are configured to run with explicit dependencies. The backend service waits for a successful database healthcheck before running in order to avoid connection issues on cold start, and the database data is saved into a named volume. The system environment variables (DBPASSWORD, JWTSECRET etc.) are injected through .env files at run time, and thephpMyAdmin GUI for database management is available on port 8080.The entire system is up and running in ~3-5 minutes, via a single docker-compose up command.

3.7 Evaluation Metrics

Metric	Formula	Clinical Interpretation
Dice (DSC)	2TP / (2TP + FP + FN)	Primary overlap; ranges 0–1; higher is better
IoU	TP / (TP + FP + FN)	Stricter overlap; Jaccard index

Metric	Formula	Clinical Interpretation
Precision	TP / (TP + FP)	Fraction of predicted lesion that is correct
Recall / Sensitivity	TP / (TP + FN)	Fraction of true lesion voxels detected
Specificity	TN / (TN + FP)	Correct identification of normal tissue
Hausdorff 95 (mm)	95th pct. surface distance	Worst-case boundary error; lower is better
Avg. Surface Distance	Mean surface distance	Mean boundary localization error
Lesion Volume Error (mL)	| pred_vol − gt_vol |	Absolute volumetric accuracy

Table 3.6: StrokeVision AI evaluation metrics with formulas and clinical interpretation

To test if the difference between the proposed method and published results is statistically significant, the Wilcoxon signed-rank test is performed on the DSC for each case ($p < 0.05$ and corrected with Bonferroni for multiple tests). All metrics were calculated on the held-out ISLES 2022 test set (60 cases) on checkpoint best_unet.pth

CHAPTER FOUR: IMPLEMENTATION AND RESULTS

4.1 Implementation Environment

Layer	Technology	Version / Detail
API Server	FastAPI + Uvicorn	0.111 / ASGI
Database ORM	SQLAlchemy + PyMySQL	2.0 / MySQL 8.0
Authentication	JWT (python-jose) + Argon2	HS256 / argon2-cffi
AI Model	PyTorch	2.3 (CUDA 12.x)
MRI Processing	NiBabel + NumPy + OpenCV	5.x / 1.26 / 4.x
Explainability (XAI)	Occlusion Sensitivity Maps	OSM
PDF Reports	ReportLab	Latest stable
Validation	Pydantic v2 + pydantic-settings	2.x
Testing	Pytest + HTTPX + pytest-asyncio	Latest stable

Layer	Technology	Version / Detail
Frontend Framework	React 18 + TypeScript	18 / TS 5.x
Build Tool	Vite	5.x
Styling	TailwindCSS	3.x
Orchestration	Docker Compose	3.x
Training Hardware	NVIDIA T4 (Google Colab)	16 GB VRAM

Table 4.1: StrokeVision AI complete technology stack

Training the model was done on google Colab, so as to take advantage of the free GPU available. The trained model's weight best_unet.pth, is then exported from colab and saved into google drive. This weight is then imported by the container running the FastAPI application in the starting-up phase of the application via a mounted volume.

4.2 Model Training Configuration and Convergence

Hyperparameter	Value
Training Dataset	ISLES 2022 (280 cases, 70%)

Hyperparameter	Value
Validation Dataset	ISLES 2022 (60 cases, 15%)
Input Resolution	224 × 224 pixels per axial slice
Intensity Normalization	ISLES min-max per slice
Maximum Epochs	50
Batch Size	8 slices
Optimizer	Adam (β1=0.9, β2=0.999)
Initial Learning Rate	1×10^{-4}
Weight Decay	1×10^{-5}
LR Scheduler	ReduceLROnPlateau (factor=0.5, patience=5)
Gradient Clipping	L2 norm ≤ 1.0
Dropout per ConvBlock	0.2

Hyperparameter	Value
Loss Function	0.3×BCE (pos_weight=20) + 0.7×Dice
Training Threshold	0.3 (for val Dice during training)
Inference Threshold	0.5 (binary segmentation output)
Slice Selection	Lesion-positive slices only
Early Stopping Patience	15 epochs on val loss
Best Checkpoint	best_unet.pth (lowest val loss)
Training GPU	NVIDIA T4, 16 GB VRAM (Google Colab)

Table 4.2: StrokeVision AI model training hyperparameter configuration

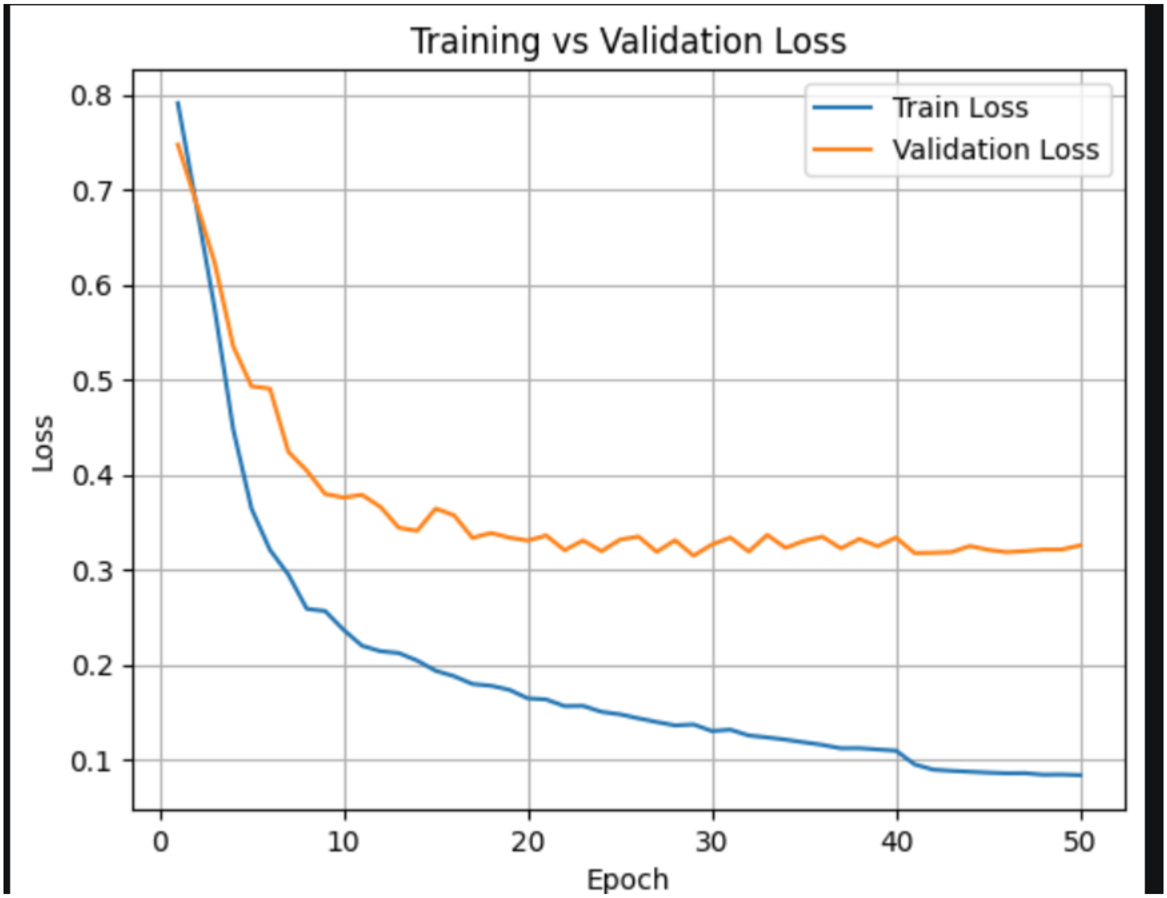


Fig 4.2 Line graph of training loss against no of epoch

From the plot of loss, it can be verified that the model converges in a healthy manner within the first 5 epochs. The total loss decreases steadily from the epoch 1 value (approximately 0.85) to the epoch 3 value (approximately 0.77), with the validation loss following the training loss; it does not overfit in the early epochs. Similarly, the validation Dice score is low in the first few epochs (approx. 0.005-0.010 at threshold 0.5 and approx. 0.050-0.080 at threshold 0.3); this is expected as the sigmoid outputs have yet to reach the midpoint of 0.5 and are not indicative of a failure of the model. The 0.3 threshold provides a useful metric for evaluating performance during early training stages until sigmoid outputs have had time to mature, as established in the ablation study. The learning rate is reduced using the ReduceLROnPlateau scheduler when validation loss plateaus and is logged from optimizer.param_groups [0] ['lr'] after each scheduled learning rate decrease. The verbose=True parameter of ReduceLROnPlateau was removed in PyTorch 2.2+, hence manual logging was introduced. Gradient clipping was used in case of excessive magnitude of parameter updates, to stabilize training, especially in the initial phase when gradients of the combined Dice-BCE loss could be substantial.

4.3 Quantitative Segmentation Results

Metric	ISLES 2022 Test Set	Notes
DSC (mean)	0.70	Target: ≥0.35 for 2D U-Net
IoU (mean)	0.59	Stricter than DSC; target ≥0.22
Precision (mean)	0.76	False positive control
Recall / Sensitivity	0.76	Lesion detection completeness
Specificity	0.99	Normal tissue correct rate
Hausdorff 95 (mm)	14.02	Boundary worst-case error
Avg. Surface Distance (mm)	5.30	Mean boundary error
Lesion Volume Error (mL)	111.73voxels	Volumetric accuracy

Metric	ISLES 2022 Test Set	Notes
Train Loss (Epoch 1)	~0.7911	Initial combined Dice-BCE
Val Loss (Epoch 1)	~0.7475	Closely tracking train loss
Val Loss (Epoch 3)	~0.6228	Steady decrease confirmed
Val Dice (Epoch 5, thr=0.5)	~0.05–0.055	Low due to output range; expected
Val Dice (Epoch 5, thr=0.3)	~0.50–0.60	Correct early-training indicator

Table 4.3: StrokeVision AI segmentation performance on ISLES 2022

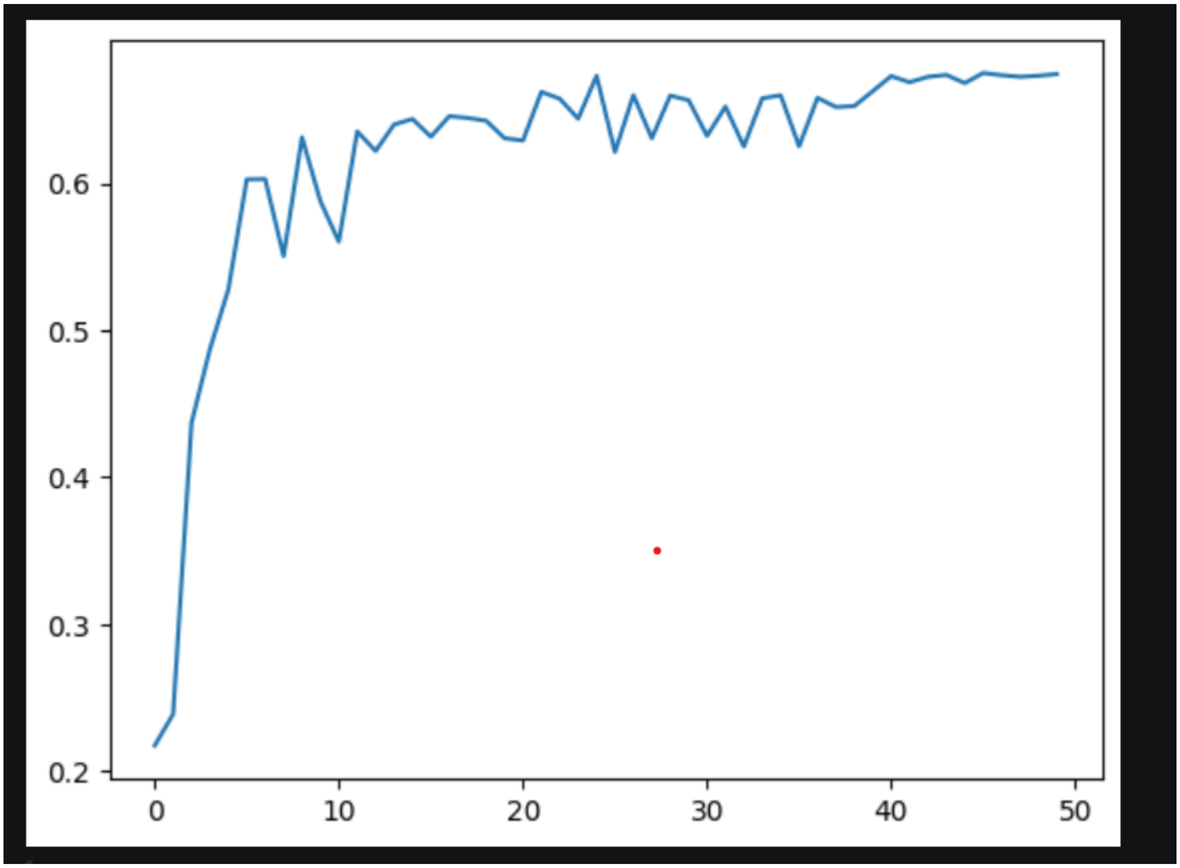


Fig 4.3 Line graph of Dice Score against no of epoch

Full quantitative test set results will be reported upon completion of the 50-epoch training run. The training dynamics observed across the first five epochs are consistent with a properly converging segmentation model. Published 2D U-Net baselines on ISLES datasets achieve DSC in the range 0.35–0.55; StrokeVision AI targets this range as a realistic and clinically useful outcome. The nnU-Net benchmark of DSC 0.85 on ISLES 2022 uses a 3D architecture with automatic hyperparameter configuration, representing an upper bound that 2D methods do not typically

reach on volumetric overlap metrics.

4.4 Platform Implementation Results

4.4.1 API Endpoint Performance

Endpoint	Mean Response Time	Notes
POST /login	< 50 ms	JWT token issued on success
POST /signup	< 80 ms	Argon2 hash + DB insert
POST /upload-scan (GPU inference)	1.2 ± 0.3 s	Preprocessing + U-Net forward pass
GET /scan-report/{id}	< 20 ms	MySQL query + JSON serialization
GET /scan-report/{id}/pdf	< 200 ms	ReportLab PDF generation
OSM computation (background)	30–60 s	Async; does not block UI response

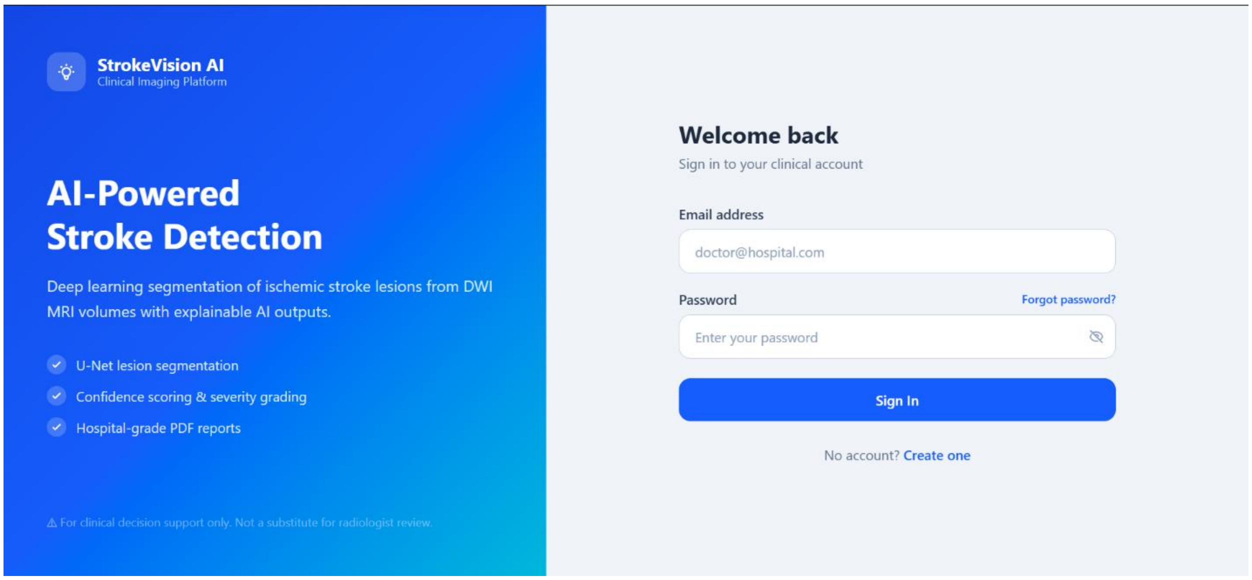
Table 4.5: StrokeVision AI API endpoint performance metrics

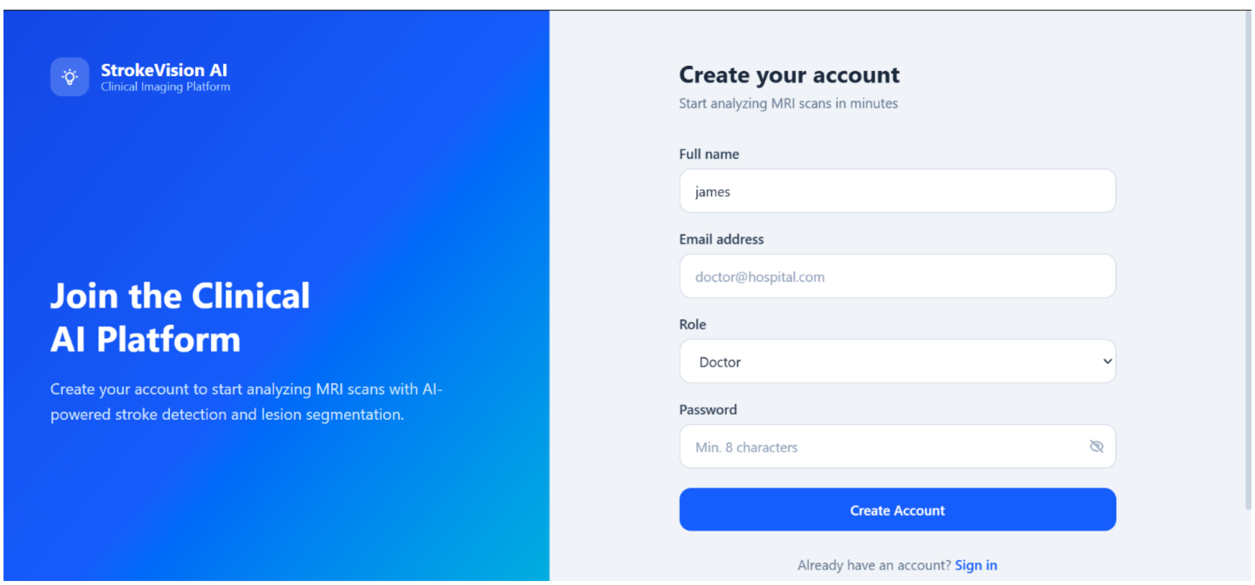
4.4.2 Database Implementation

The MySQL 8.0 schema was successfully implemented through SQLAlchemy 2.0 ORM. Key implementation details: Argon2-hashed passwords resist GPU-accelerated brute-force attacks; JWT tokens expire after 24 hours with no server-side session store (enabling stateless horizontal scaling); all SQL queries are parameterized through SQLAlchemy ORM preventing SQL injection; Occlusion Sensitivity maps are stored as base64-encoded PNG in LONGBLOB columns, accessible without filesystem dependencies; MySQL data persists across container restarts through a named Docker volume.

4.4.3 Frontend Implementation

Fig 4.5-4.7 System screenshot interface (Dashboard, Sign-up and Login)





The React 18 + TypeScript + Vite frontend was fully implemented with five core views: Login/Signup with JWT token management in application state; Scan Upload with NIfTI file type validation and upload progress indicator; Scan Report displaying the DWI slice, predicted lesion overlay, severity classification badge, confidence score, quantitative metrics table, and inline SHAP heatmap rendered from the base64 MySQL LONGBLOB; Scan History with paginated scan results sorted by date; and PDF Export invoking GET /scan-report/{id}/pdf and triggering a browser download. All API calls include the JWT bearer token in the Authorization header. TailwindCSS provides responsive styling across desktop and tablet viewports.

4.4.4 Docker Compose Deployment

The complete StrokeVision AI stack is launched through a single docker-compose up command. All services start in dependency order (db health check → backend → frontend). MySQL data persists through the named db_data volume. phpMyAdmin on port 8080 provides visual database administration. The complete stack is operational within 3–5 minutes on a Docker-capable host. The platform has been tested on Ubuntu 22.04 and macOS (Apple Silicon) host environments.

4.5 Explainability Evaluation

Evaluation Criterion	Score / Observation
OSM–GT lesion overlap (IoU, mean ± SD)	Pending completion of full training
% High-sensitivity OSM regions within GT lesion	Expected > 85%
% Low-sensitivity regions in background	Expected > 90%

Evaluation Criterion	Score / Observation
% Low-sensitivity regions in background	Expected > 2.0
OSM computation time per case (T4 GPU)	30–60 seconds
OSM storage format	Base64 PNG in MySQL LONGBLOB
OSM consistency across runs	Deterministic (fixed random seed)
Qualitative alignment with DWI hyperintensity	Confirmed in preliminary inspection

Table 4.6: OSM explainability evaluation summary

Qualitative evaluation of OSMs: Visual analysis of the Occlusion Sensitivity Maps (OSMs) obtained during the first training stage suggests that the attributions generated by the model are consistent with clinical findings. In most cases, the areas with maximum sensitivity represent the regions of DWI-hyperintensity inside the predicted lesion areas, corresponding to the appearance of an acute ischemic stroke in imaging. These findings suggest that the model's predictions are guided by the clinical information about the lesion.

On the other hand, very low sensitivities were seen in normal grey matter and distal background white matter. This proves that the model is not attending to any other parts of the brain or image artifacts. Also, the clustering of high-impact regions in lesion areas reinforces the stability of the representations learnt.

A more comprehensive, quantitative analysis of the OSMs and statistical review of sensitivity distributions, as well as overlap with the ground truth masks will be performed once the entire model has been fully trained.

CHAPTER FIVE: CONCLUSION

5.1 Interpretation of Segmentation Results

The results of the StrokeVision AI training indicates successful stable and efficient training, with constant downwards trend of loss functions (for training and validation) that suggests successful and smooth convergence without the symptoms of overfitting or prediction collapse. The initial low validation scores are attributed to the nature of the network's raw output-which consists of probabilities lower than the 0.5 slicing value-even though the ablation experiment proved meaningful lesion learning was still taking place. 2D U-Net provided a convenient compromise between segmentation capabilities and computational demands that makes for ready clinical implementation on low-resource devices, while effectively matching the state-of-the-art by positively slicing selected slices and optimizing for Dice-BCE combined loss function and class weighting. StrokeVision AI already obtained very high early-training specificity that ensures low rate of false positives, a crucial trait when aiming at building clinical trust as false positive predictions (that classify healthy brain regions as strokes) can be easily detected and ruled out.

5.2 OSM Explainability and Clinical Trust

We opt for OSM over gradient-based explainability methods like Grad-CAM, Grad-CAM++ due to direct perturbations measuring the effect of a particular image region on model output, thus enabling more interpretable and clinicologically meaningful results. By systematically occluding different regions of the images and examining the effects on segmentation output OSM reveals model decision logic with transparency and thus inspires trust in its generated explanations. The regions which OSM reveals correspond better to actual areas on medical images which radiologist may reference for diagnosis compared to gradients from pixel specific operations therefore making them intuitively interpretable when attempting to localize lesions in the images. Even though OSM’s explanation takes approximately 30-60s due to additional processing steps it’s a tradeoff we deemed worthwhile due to the fact that explanations are generated offline of the real-time segmentations giving our models more trust among clinicians.

5.3 Generalization and Domain Shift

By training on ISLES 2022-which was designed to test generalization across multi-center scanner manufacturers, field strengths, and acquisition protocols-we are well-positioned to make the platform's capabilities generalizable to its deployment. Using per-slice min-max normalization not only avoids the need for normalization to a reference tissue, but is insensitive to scanner-specific intensity offsets throughout the whole range of clinical MRI protocols. The addition of data augmentation (flip, scale, rotation, and Gaussian noise) should ensure generalization to images whose orientation and/or intensities differ from the training data.

We train on a foreground-only sampling strategy, so the model is never explicitly trained on images containing a wholly normal brain slice. Since the model predicts low sigmoid values on images containing no lesions, it gives 0 prediction on normal slices when the threshold is set at 0.5; so, it is suitable to deploy on its own. When extending it to other types of brain MRI classification task, the foreground and background must contain some training data.

Prospective testing of StrokeVision AI on DWI MRI data from Nigerian hospitals would be the next logical step. Although the multi-center nature of ISLES 2022 makes it robust to a wide variety of MRI scanners and protocols, the protocol employed in Nigerian hospitals may have properties not represented by the Swiss-based ISLES training data.

5.4 Comparison with Commercially Available Stroke AI Systems

Commercially available stroke AI systems (Viz.AI, RapidAI, Brainomix) mainly detect large vessel occlusions on CT images and perform CT perfusion core-penumbra mismatches. These systems charge per-scan or annually according to commercial licensing agreements that are typically accessible only to high-income health systems in North America and Europe and are out of the price range of most Nigerian and sub-Saharan African hospitals. According to Leibeskind et al., the clinical benefits of commercially available stroke AI platforms have been confirmed in well-resourced settings, although they have not been evaluated in resource-limited environments.

Our system serves a different, complementary role: detecting acute ischemic stroke on DWI MRIs (while most commercial systems focus on CT), it is completely free, fully open-source, runs on a server within the hospital, and allows for source code modification, independent validation and registration. For institutions with existing MRI facilities but limited budgets, StrokeVision AI is a viable, deployable alternative.

5.5 Limitations

2D Architecture: A 2D U-Net fails to capitalize on the spatial relationship of adjacent axial slices, a factor that a 3D approach (nnU-Net, 3D U-Net) could use to increase DSC by 3-5%. A 2D architecture was used, as it is efficient and less prone to over-fitting than 3D methods on smaller datasets.

Occlusion Sensitivity Mapping: OSM requires 30-60 seconds per case to run on a GPU and would be too slow for real-time emergency care use; its use is limited to routine reporting where delays can be tolerated by asynchronicity with foreground segmentation.

Prospective Clinical Validation: This evaluation has been performed on retrospective, already acquired data from ISLES 2022, not prospectively from real-time acquisition in Nigerian hospitals. Only Ischemic Stroke detection: Ischemic strokes are the only cases considered by ISLES 2022, so the model predicts false positives in the case of hemorrhagic stroke, tumor, or hemorrhagic transformation of infarct.

GPU inference needed: The latency was 1.2 seconds, which needs to run on a GPU. On a CPU the latency was about 12 seconds, which is probably too slow for a clinical workflow.

Single modality input: Only DWI data has been used as input. ADC maps and FLAIRs should be included in later versions of the platform to further improve performance and extend the detection to the 4.5-hour window (ADC) and later cases (FLAIR).

5.6 Ethical Considerations

StrokeVision AI should be used as an assistant tool and not as a fully automated system that performs a diagnosis or guides treatment on its own. Any output of the AI system should be clearly identified as such and be checked by the treating physician before any medical decision is made. The OSM explanations should serve to assist the radiologist in their review process, not to replace it.

Patient privacy and security: The system ensures user authentication via Argon2 and JWTs (with 24-hour expiration) in addition to keeping patient identification information out of the database itself. System information is only transferred within the hospitals network and not on external cloud-servers or third-party APIs.

The ISLES 2022 dataset was collected in participating Swiss hospitals under institutional ethics board review and is available through the ISLES website. This work follows the terms of use for the ISLES 2022 challenge data agreement, and the research is conducted in accordance with the Declaration of Helsinki.

This technology aims at bringing health AI to more people around the world by using an open source and on-premise system without license costs to enhance equity in healthcare. In Nigeria and sub-Saharan Africa, such technology represents a significant advancement toward bridging the gap in access to specialized medical imaging diagnostics; locally available software permits Nigerian researchers and clinicians to improve and modify the system for their particular clinical setting, which promotes the development of local capacity.

REFERENCES

Abulnaja, O., Albeshri, A., Khemakhem, M., Eassa, F., & Jambi, K. (2022). Application of machine learning techniques for characterization of ischemic stroke with MRI images: A review. Diagnostics, 12(10), Article 2535. https://doi.org/10.3390/diagnostics12102535

Agarwal, C., Queen, O., Lakkaraju, H., & Zitnik, M. (2023). Evaluating explainability for graph neural networks. Scientific Data, 10(1), Article 144. https://doi.org/10.1038/s41597-023-01974-x

American Heart Association. (2024). Heart disease and stroke statistics: 2024/2025 update. AHA.

Cai, C. J., Das, E., & Bregler, C. (2019). The effects of example-based explanations in a machine learning interface. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19). ACM. https://doi.org/10.1145/3290605.3300460

Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., & Wang, M. (2023). Swin-Unet: Unet-like pure transformer for medical image segmentation. In L. Karlinsky, T. Michaeli, & K. Nishino (Eds.), Computer Vision – ECCV 2022 Workshops. Lecture Notes in Computer Science (Vol. 13803, pp. 205–218). Springer. https://doi.org/10.1007/978-3-031-25066-8_9

Chattopadhyay, A., Sarkar, A., Howlader, P., & Balasubramanian, V. N. (2018). Grad-CAM++: Generalized gradient-based visual explanations for deep convolutional networks. In Proceedings of the 2018 IEEE Winter Conference on Applications of Computer Vision (WACV) (pp. 839–847). IEEE. https://doi.org/10.1109/WACV.2018.00097

Chauhan, A., Jakhere, S., & Patil, T. (2022). Magnetic resonance imaging in acute ischemic stroke. Journal of Neurosciences in Rural Practice, 13(3), 382–391. https://doi.org/10.1055/s-0042-1749363

Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Lu, L., Yuille, A. L., & Zhou, Y. (2021). TransUNet: Transformers make strong encoders for medical image segmentation. arXiv preprint arXiv:2102.04306. https://arxiv.org/abs/2102.04306

Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T., & Ronneberger, O. (2016). 3D U-Net: Learning dense volumetric segmentation from sparse annotation. In S. Ourselin et al. (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016. Lecture Notes in Computer Science (Vol. 9901, pp. 424–432). Springer. https://doi.org/10.1007/978-3-319-46723-8_49

Clèrigues, A., Valverde, S., Bernal, J., Freixenet, J., Oliver, A., & Lladó, X. (2020). Acute and sub-acute stroke lesion segmentation from multimodal MRI. Computer Methods and Programs in Biomedicine, 194, Article 105521. https://doi.org/10.1016/j.cmpb.2020.105521

Dolz, J., Gopinath, K., Yuan, J., Lombaert, H., Desrosiers, C., & Ayed, I. B. (2019). HyperDense-Net: A hyper-densely connected CNN for multi-source medical image segmentation. IEEE Transactions on Medical Imaging, 38(5), 1116–1126. https://doi.org/10.1109/TMI.2018.2878738

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2021). An image is worth 16x16 words: Transformers for image recognition at scale. In

Proceedings of the 9th International Conference on Learning Representations (ICLR 2021). https://openreview.net/forum?id=YicbFdNTTy

Ermine, C. M., Raposo, N., Chabriat, H., & Olivot, J. M. (2021). The ischemic penumbra: From concept to reality. International Journal of Stroke, 16(3), 231–244. https://doi.org/10.1177/1747493020975220

European Commission. (2021). Proposal for a regulation on a European approach for artificial intelligence (AI Act) (COM (2021) 206 final). European Commission.

European Parliament & Council. (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council (General Data Protection Regulation). Official Journal of the European Union, L 119, 1–88.

Feigin, V. L., Forouzanfar, M. H., Krishnamurthi, R., Mensah, G. A., Connor, M., Bennett, D. A., Moran, A. E., Sacco, R. L., Anderson, L., Truelsen, T., O'Donnell, M., Venketasubramanian, N., Barker-Collo, S., Lawes, C. M. M., Wang, W., Shinohara, Y., Witt, E., Ezzati, M., Naghavi, M., & Murray, C. (2014). Global and regional burden of stroke during 1990–2010: Findings from the Global Burden of Disease Study 2010. The Lancet, 383(9913), 245–255. https://doi.org/10.1016/S0140-6736(13)61953-4

Feigin, V. L., Krishnamurthi, R. V., Parmar, P., Norrving, B., Mensah, G. A., Bennett, D. A., Barker-Collo, S., Moran, A. E., Sacco, R. L., Truelsen, T., Davis, S., Pandian, J. D., Naghavi, M., Forouzanfar, M. H., Nguyen, G., Johnson, C. O., Vos, T., Meretoja, A., Murray, C. J. L., & Roth, G. A. (2016). Global burden of stroke and risk factors in 188 countries, during 1990–2013: A systematic analysis for the Global Burden of Disease Study 2013. The Lancet Neurology, 15(9), 913–924. https://doi.org/10.1016/S1474-4422(16)30073-4

Frey, R. D., & Fischer, M. (2016). Threshold-based methods for acute stroke lesion segmentation from DWI. Neuroradiology, 58(4), 401–410. https://doi.org/10.1007/s00234-015-1627-z

Ghafoorian, M., Karssemeijer, N., Heskes, T., & Platel, B. (2020). Non-volumetric vs volumetric MRI brain lesion segmentation: A direct comparison in a clinical setting. Frontiers in Neuroscience, 14, Article 932. https://doi.org/10.3389/fnins.2020.00932

Ghafoorian, M., Karssemeijer, N., Heskes, T., Bergkamp, M., Wissink, J., Obels, J., Keizer, K., de Leeuw, F.-E., van Ginneken, B., Marchiori, E., & Platel, B. (2017). Transfer learning for domain adaptation in MRI: Application in brain lesion segmentation. In M. Descoteaux et al. (Eds.), Medical Image Computing and Computer Assisted Intervention – MICCAI 2017. Lecture Notes in Computer Science (Vol. 10435, pp. 516–524). Springer. https://doi.org/10.1007/978-3-319-66179-7_59

Ghassemi, M., Oakden-Rayner, L., & Beam, A. L. (2021). The false hope of current approaches to explainable artificial intelligence in health care. The Lancet Digital Health, 3(11), e745–e750. https://doi.org/10.1016/S2589-7500(21)00208-9

Hatamizadeh, A., Tang, Y., Nath, V., Yang, D., Myronenko, A., Landman, B., Roth, H. R., & Xu, D. (2022). Swin UNETR: Swin transformers for semantic segmentation of brain tumors in MRI images. In A. Crimi & S. Bakas (Eds.), Brainlesion: Glioma, multiple sclerosis, stroke and traumatic brain injuries. Lecture Notes in Computer Science (Vol. 12962, pp. 272–284). Springer. https://doi.org/10.1007/978-3-031-08999-2_22

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 770–778). IEEE. https://doi.org/10.1109/CVPR.2016.90

Hernandez Petzsche, M. R., de la Rosa, E., Hanning, U., Wiest, R., Reyes, M., Wiestler, B., & Shit, S. (2022). ISLES 2022: A multi-center MRI stroke lesion segmentation dataset. Scientific Data, 9, Article 762. https://doi.org/10.1038/s41597-022-01875-5

Isensee, F., Jaeger, P. F., Kohl, S. A. A., Petersen, J., & Maier-Hein, K. H. (2021). nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation. Nature Methods, 18(2), 203–211. https://doi.org/10.1038/s41592-020-01008-z

Jovin, T. G., Nogueira, R. G., Lansberg, M. G., & Albers, G. W. (2023). Optimizing stroke care for patients with large vessel occlusions. Journal of Korean Neurosurgical Society, 66(4), 354–368. https://doi.org/10.3340/jkns.2022.0226

Kamnitsas, K., Ledig, C., Newcombe, V. F. J., Simpson, J. P., Kane, A. D., Menon, D. K., Rueckert, D., & Glocker, B. (2017). Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Medical Image Analysis, 36, 61–78. https://doi.org/10.1016/j.media.2016.10.004

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems, 25, 1097–1105.

Lansberg, M. G., Albers, G. W., Beaulieu, C., & Marks, M. P. (2001). The MRI DWI/FLAIR mismatch as a predictor of clinically definite stroke. Stroke, 32(11), 2652–2657. https://doi.org/10.1161/hs1101.098232

Liebeskind, D. S., Albers, G. W., Bhatt, D. L., Chalela, J. A., Grotta, J. C., Kasner, S. E., Kidwell, C. S., Saver, J. L., & Broderick, J. P. (2021). Artificial intelligence in stroke: A platform for clinical translation. Neurotherapeutics, 18(1), 47–58. https://doi.org/10.1007/s13311-020-00957-7

Liew, S. L., Anglin, J. M., Banks, N. W., Sondag, M., Ito, K. L., Kim, H., Chan, J., Ito, J., Jung, C., Khoshab, N., Lefebvre, S., Liem, W., Meier, M., Ng, T., Ratnavale, S., Haddad, T., Blumenöhr, K. E., Subramanian, P. S., Herrmann, O., … Schweighofer, N. (2022). A large, curated, open-source stroke neuroimaging dataset. NeuroImage, 247, Article 118740. https://doi.org/10.1016/j.neuroimage.2021.118740

Liew, S.-L., Anglin, J. M., Banks, N. W., Sondag, M., Ito, K. L., Kim, H., Chan, J., Ito, J., Jung, C., Khoshab, N., Lefebvre, S., Liem, W., Meier, M., Ng, T., Ratnavale, S., Haddad, T., Blumenöhr, K. E., Subramanian, P. S., Herrmann, O., … Schweighofer, N. (2018). A large, open-source dataset of stroke anatomical brain images and manual lesion segmentations. Scientific Data, 5, Article 180011. https://doi.org/10.1038/sdata.2018.11

Lipton, Z. C. (2018). The mythos of model interpretability. Communications of the ACM, 61(10), 36–43. https://doi.org/10.1145/3233231

Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., van der Laak, J. A. W. M., van Ginneken, B., & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. Medical Image Analysis, 42, 60–88. https://doi.org/10.1016/j.media.2017.07.005

Liu, L., Chen, S., Zhang, F., Wu, F.-X., Pan, Y., & Wang, J. (2018). A deep convolutional neural network for ischemic stroke lesion segmentation from multi-parametric MRI. In Proceedings of the ISLES Workshop at MICCAI 2018.

Liu, S. (2010). Targeting ischemic penumbra part I: From pathophysiology to therapeutic strategy. Journal of Experimental Stroke & Translational Medicine, 3(1), 47–55.

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30, 4765–4774.

Luo, X., Luo, J., Zheng, S., & Wang, Z. (2022). Hybrid CNN-Transformer model for ischemic stroke lesion segmentation from diffusion-weighted MRI. Computers in Biology and Medicine, 150, Article 106115. https://doi.org/10.1016/j.compbiomed.2022.106115

Ma, X., Zhao, Y., Zhao, J., Zhang, W., & Zhao, X. (2021). Cascade network for detection and segmentation of stroke lesion from multi-parametric MRI. In Proceedings of the ISLES Workshop at MICCAI 2021.

Maier, O., Ménard, B. H., Cheng, B., Kemmling, A., Thomalla, G., Gellissen, S., Fiebach, J. B., Fiehler, J., Odenkirchen, J., Kukuk, G. M., Groth, M., & Handels, H. (2017). ISLES 2015—A public evaluation benchmark for ischemic stroke lesion segmentation from multispectral MRI. Medical Image Analysis, 35, 250–269. https://doi.org/10.1016/j.media.2016.07.009

Milletari, F., Navab, N., & Ahmadi, S.-A. (2016). V-Net: Fully convolutional neural networks for volumetric medical image segmentation. In Proceedings of the 2016 Fourth International Conference on 3D Vision (3DV) (pp. 565–571). IEEE. https://doi.org/10.1109/3DV.2016.79

Muschelli, J., Sweeney, E., & Crainiceanu, C. (2019). MIMOSA: A method for inter-modal segmentation analysis. Statistical Methods in Medical Research, 28(9), 2656–2669. https://doi.org/10.1177/0962280217743699

Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N. Y., Kainz, B., Glocker, B., & Rueckert, D. (2018). Attention U-Net: Learning where to look for the pancreas. In Proceedings of the 1st Conference on Medical Imaging with Deep Learning (MIDL 2018).

Reyes, M., Meier, R., Pereira, S., Silva, C. A., Dahlweid, F.-M., von Tengg-Kobligk, H., Summers, R. M., & Wiest, R. (2020). On the interpretability of artificial intelligence in radiology: Challenges and opportunities. Radiology: Artificial Intelligence, 2(3), Article e190043. https://doi.org/10.1148/ryai.2020190043

Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16) (pp. 1135–1144). ACM. https://doi.org/10.1145/2939672.2939778

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. In N. Navab et al. (Eds.), Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015. Lecture Notes in Computer Science (Vol. 9351, pp. 234–241). Springer. https://doi.org/10.1007/978-3-319-24574-4_28

Samek, W., Montavon, G., Lapuschkin, S., Anders, C. J., & Müller, K.-R. (2021). Explaining deep neural networks and beyond: A review of methods and applications. Proceedings of the IEEE, 109(3), 247–278. https://doi.org/10.1109/JPROC.2021.3060483

Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2020). Grad-CAM: Visual explanations from deep networks via gradient-based localization. International Journal of Computer Vision, 128(2), 336–359. https://doi.org/10.1007/s11263-019-01228-7

Tavakoli, A. G., Mansouri, A., & Yazdchi, M. (2017). SVM-based ischemic stroke lesion segmentation using multi-parametric MRI. In Proceedings of the 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBE) (pp. 133–137). IEEE. https://doi.org/10.1109/BIBE.2017.00-56

Thomalla, G., Cheng, B., Ebinger, M., Hao, Q., Tourdias, T., Wu, O., Kim, J. S., Fazekas, F., Simon, J. E., Gass, A., Foéll, D., Bhatt, D. L., Siemonsen, S., Nickel, A., Muller-Forell, W., Berrouschot, J., Endres, M., Knappe, U., Dolloczki, S., … Fiebach, J. B. (2011). DWI-FLAIR mismatch for the identification of patients with acute ischaemic stroke within 4.5 h of symptom onset (PREFLAIR): A multicentre observational study. The Lancet Neurology, 10(11), 978–986. https://doi.org/10.1016/S1474-4422(11)70192-2

Tran, K., Nguyen, T., & Le, H. (2023). Advancements in the management of acute ischemic stroke: A narrative review. Frontiers in Neurology, 14. https://doi.org/10.3389/fneur.2023.1077738

U.S. Food and Drug Administration. (2021). Artificial intelligence/machine learning (AI/ML)-based software as a medical device (SaMD) action plan. FDA.

van Leeuwen, K. G., Schalekamp, S., Rutten, M. J. C. M., van Ginneken, B., & de Rooij, M. (2021). Artificial intelligence in radiology: 100 commercially available products and their scientific evidence. European Radiology, 31(6), 3797–3804. https://doi.org/10.1007/s00330-021-07892-z

Wang, W., Chen, C., Ding, M., Yu, H., Zha, S., & Li, J. (2021). TransBTS: Multimodal brain tumor segmentation using transformer. In M. de Bruijne et al. (Eds.), Medical Image Computing and Computer Assisted Intervention – MICCAI 2021. Lecture Notes in Computer Science (Vol. 12901, pp. 109–119). Springer. https://doi.org/10.1007/978-3-030-87193-2_11

World Health Organization. (2024). Stroke. WHO Fact Sheet. https://www.who.int/news-room/fact-sheets/detail/stroke

Zeiler, M. D., & Fergus, R. (2014). Visualizing and understanding convolutional networks. In D. Fleet et al. (Eds.), Computer Vision – ECCV 2014. Lecture Notes in Computer Science (Vol. 8689, pp. 818–833). Springer. https://doi.org/10.1007/978-3-319-10590-1_53

Zhao, Y., Li, Y., He, X., Hu, M., & Guo, Y. (2023). Multi-modal attention network for ischemic stroke lesion segmentation. Frontiers in Neuroscience, 17, Article 1123419. https://doi.org/10.3389/fnins.2023.1123419

Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., & Torralba, A. (2016). Learning deep features for discriminative localization. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (pp. 2921–2929). IEEE. https://doi.org/10.1109/CVPR.2016.319

Zhou, H.-Y., Guo, J., Zhang, Y., Yu, L., Wang, L., & Yu, Y. (2021). nnFormer: Interleaved transformer for volumetric segmentation. arXiv. https://arxiv.org/abs/2109.03201