International Journal of Applied Smart Interdisciplinary Technologies (IJASIT)

Home Page: https://ijasit.org **ISSN: 3139-759X**

https://doi.org/10.68104/ijasit.v1.i3.32

Brain Tumor Progression Prediction and 3D Reconstruction Using ViT–LSTM

Kohilambal R[1], Yukthi G[2,*], N Architha[3], Panya Shree M[4], Apoorva S[5]

[1 2 3 4 5] Department of Computer Science and Engineering (Data Science), Vemana Institute of Technology, Bengaluru, India.

*** Corresponding author.**
E-mail: *yukthi70@gmail.com*

Article History:
Received Date: 09 August 2026
Revised Date: 20 August 2026
Accepted Date: 24 August 2026
Published Date: 24 September 2026

Keywords:

Brain Tumor Detection; MRI; Vision Transformer; Long Short-Term Memory; 3D Reconstruction; Medical Image Analysis; Tumor Progression

ABSTRACT

Brain tumors are serious neurological disorders in which abnormal tissue growth can impair memory, coordination, mobility, and cognition. Magnetic resonance imaging (MRI) provides high-resolution structural information for diagnosis and follow-up, but manual analysis is time-consuming, difficult to scale, and dependent on expert interpretation. This research addresses the limited integration of segmentation, longitudinal progression analysis, volumetric measurement, three-dimensional (3D) visualization, and explainability in existing single-scan systems. Its purpose is to provide a unified workflow that supports monitoring of tumor change across clinical encounters. The method applies MRI validation and preprocessing, U-Net-based tumor segmentation, Vision Transformer (ViT) spatial feature extraction, Long Short-Term Memory (LSTM) temporal modelling, voxel-based volume estimation, Marching Cubes reconstruction with Three.js/Plotly visualization, attention/Grad-CAM explanation, and structured report generation. Segmentation evaluation achieved a Dice score of 95.8% and an intersection-over-union score of 91.5%. The ViT–LSTM module was integrated for longitudinal tumorprogression prediction, while the system generated patient-specific volumetric measurements, interactive 3D meshes, and explanatory visualizations. The integrated workflow therefore combines spatial, temporal, quantitative, visual, and explanatory outputs to support—not replace—clinical interpretation.

1. Introduction

1.1 Background and Clinical Motivation

A brain tumor is an abnormal tissue mass formed by cells growing within the brain and affecting normal function. Such masses may be cancerous or non-cancerous and can impair memory, coordination, mobility, and reasoning. Early detection and continued monitoring support treatment planning and assessment of tumor behavior. MRI is commonly used because its T1, T1-contrast-enhanced (T1ce), T2, and fluid-attenuated inversion recovery (FLAIR) sequences provide complementary information about tissue differences, edema, enhancement, tumor location, size, and structure [8, 10]. Manual interpretation and boundary delineation remain time-consuming, difficult to scale, and dependent on radiological expertise. Deep-learning methods can automate repeatable image-processing tasks, while U-Net-family architectures provide voxel-level masks and transformer architectures capture longer-range spatial relationships [5, 6, 11, 22]. Longitudinal analysis adds a

further requirement: the system must preserve the chronological relation between repeated examinations rather than treating every scan as an independent image.

1.2 Problem, Need, and Research Gap

Existing brain-tumor methods predominantly identify, classify, or segment tumor regions in individual scans. They provide limited integrated support for structural comparison across time, growth-rate estimation, volume tracking, interactive 3D interpretation, and explanation of model output. Independent processing of each scan can obscure changes in direction, expansion, and bound-ary shape. A unified workflow is therefore needed to connect preprocessing, segmentation, spatial feature learning, temporal prediction, volumetric measurement, reconstruction, explainability, and reporting.

The reviewed literature contains strong individual solutions for segmentation, classification, transformer-based contextual modelling, and explainability. However, the research gap addressed here is the absence, in the reviewed systems, of a single pipeline that combines U-Net segmentation, ViT spatial encoding, LSTM analysis of ordered scans, voxel-based volume estimation, 3D reconstruction, and Grad-CAM output.

1.3 Objectives, Scope, and Contributions

The objectives are to develop automated MRI tumor segmentation for tumor boundaries and regions and to generate 3D tumor reconstruction from segmented slices for improved spatial visualization and surgical-planning support. The implemented workflow also integrates ViT feature extraction, LSTM progression modelling, volumetric estimation, explainability, and report generation.

The scope covers MRI preprocessing, tumor seg-mentation, progression forecasting, volume estimation, 3D reconstruction, and explanatory visualization. The system is intended for medical-imaging research, tumor-progression studies, clinical visualization, and AI-assisted decision support. It does not independently diagnose patients or replace clinical judgment.

The proposed sequence is MRI acquisition and validation; noise reduction, resizing, intensity normalization, and skull stripping; U-Net segmentation; ViT patch-based feature extraction; LSTM modelling of ordered scans; voxel-based volume estimation; Marching Cubes reconstruction with Three.js/Plotly visualization; Grad-CAM explanation; and report generation. Section 2 reviews prior work, Section 3 presents system specifi-cations, Section 4 describes the methodology, Section 5 reports the results, Section 6 discusses applications, and Section 7 presents the conclusion and future work.

2. Literature Review

2.1 Related Work

ZainEldin et al. combined deep learning with sine-cosine fitness grey-wolf optimization for MRI brain-tumor detection and classification [1]. Jia and Shu developed BiTr-Unet, a CNN–Transformer network for multimodal MRI brain-tumor segmentation, illustrating how convolutional locality and transformer context can be combined [2]. Tjoa and Guan surveyed explainable AI with particular attention to medical applications and the need for accountable model interpretation [3].

Attention U-Net introduced attention gates that allow a segmentation network to emphasize relevant structures while suppressing irrelevant responses [23]. The Swin Transformer uses shifted-window attention to construct hierarchical visual representations [4], while U-Net uses an encoder–decoder structure with skip connections for biomedical segmentation [5]. nnU-Net extends this line of work by automatically configuring preprocessing, architecture, training, and post-processing for a new segmentation task [6].

TransBTS combines a 3D CNN encoder–decoder with transformer-based global feature modelling for multimodal brain-tumor segmentation [24]. Myronenko used a variational-autoencoder branch to regularize a 3D brain-tumor segmentation network through input reconstruction [25]. Krishnan et al. developed a rotation-invariant Vision Transformer for brain-tumor classification, demonstrating the relevance of transformer-based spatial learning to MRI analysis [7].

2.2 Existing Techniques

CNN, U-Net, residual, and encoder–decoder models learn successive local and multiscale MRI features. V-Net and 3D U-Net extend segmentation to volumetric data [15, 16]. Vision transformers treat image patches as tokens and use self-attention to model contextual relationships [11]. Swin Transformer introduces hierarchical shifted windows [4], and UNETR uses a transformer encoder for 3D medical-image segmentation [14]. These

works support the distinction used in the proposed architecture: U-Net generates the tumor mask, whereas ViT generates spatial representations for later temporal modelling.

LSTM networks are designed to retain and update information across ordered sequences [12]. In the present system, the ordered sequence is formed from patient-specific follow-up examinations. Grad-CAM provides spatial saliency maps from gradients and feature activations [13]; the proposed workflow uses these maps as explanatory output rather than as a segmentation substitute.

MRI standardization is also important. Skull strip-ping can be performed through brain-extraction methods, intensity nonuniformity can be corrected where needed, and interoperable neuroimaging formats support volumetric processing [18, 17, 19]. For 3D display, segmented surfaces can be triangulated using established isosurface extraction methods and displayed with medical-visualization or browser-based tools [21, 20].

2.3 Comparative Analysis and Research Gaps

BraTS established a common multimodal benchmark for brain-tumor segmentation, while TCIA provides a public infrastructure for cancer-imaging collections [8, 9]. Later BraTS work broadened standardized evaluation and radiogenomic tasks [10]. These resources support reproducible research, but a dataset repository does not itself define how a particular study divided subjects, constructed longitudinal pairs, or prevented patient overlap across data splits.

The reviewed methods improve classification, CNN–Transformer segmentation, attention-based modelling, self-configuring segmentation, volumetric processing, and explanation. The combined gap addressed by this work comprises:

(i) chronological MRI analysis,

(ii) ViT spatial encoding connected to LSTM temporal learning,

(iii) quantitative tumor-volume estimation,

(iv) interactive 3D reconstruction, and (v) integrated visualization, explanation, and reporting.

2.4 Limitations of Existing Systems

Single-scan processing cannot directly quantify change across encounters. Two-dimensional masks do not automatically provide physical volume unless voxel spacing and slice geometry are incorporated. Classification or segmentation output without explanatory localization can be difficult to interpret, and computational cost may limit routine use. The proposed system addresses these limitations through an integrated pipeline, but clinical utility still depends on documented external validation and expert assessment.

3. System Specification

3.1 Hardware Requirements

The hardware configuration comprises a 10th-generation Intel Core i7 or AMD Ryzen R7 processor; an NVIDIA RTX 3060 with at least 8 GB of GPU memory, or a cloud-equivalent Tesla T4; at least 16 GB DDR4 RAM, with 32 GB recommended for multimodal batches; at least 500 GB SSD storage for NIfTI volumes, tensors, and checkpoints; and Windows 10/11 with WSL2 support. These requirements describe the implementation environment rather than minimum clinical-deployment specifications.

3.2 Software Requirements

Python 3.10 is the principal development environment. PyTorch 2.x supports ViTLSTM architecture definition, training, inference, and Grad-CAM hooks. TensorFlow 2.x supports data-pipeline construction and model deployment; it is not used as a second training framework for the same ViTLSTM model. OpenCV, SimpleITK, and NiBabel support image preprocessing and NIfTI handling; NumPy, SciPy, and Pandas support numerical processing; Matplotlib and Plotly support analysis; Three.js supports browser-based 3D rendering; and scikit-learn supports evaluation. VS Code is identified as the main development environment, with PyCharm and Jupyter Notebook used for project management and rapid experimentation.

Table 1 Summarizes the libraries before their use in the subsequent methodology.

Library/Framework	Purpose
OpenCV, SimpleITK, NiBabel	MRI preprocessing, skull stripping support, contrast handling, resizing, nor-malization, resampling, and NIfTI input/output.
TensorFlow 2.x	Data-pipeline construction and model-deployment support.
PyTorch 2.x	ViT–LSTM definition, training, inference, and Grad-CAM hooks.
NumPy, SciPy, Pandas	Tensor and tabular operations, numerical processing, and voxel-based vol-ume estimation.
Matplotlib, Plotly, Three.js	Static graphs, tumor-growth trajectories, and interactive 3D rendering.
Scikit-learn	Dice, IoU, accuracy, confusion-matrix, and related evaluation utilities.
React, TypeScript, HTML/CSS, Tail-wind	Web interface for upload, monitoring, visualization, and results.
Flask	Backend inference, data exchange, and report-generation services.

4. Proposed System and Methodology

4.1 System Architecture

The system studies tumor progression across multiple clinical encounters rather than relying only on one MRI snapshot. As shown in Fig. 1, the user interface communicates with a Python backend and data layer. The deep-learning layer separates the roles of preprocessing, U-Net segmentation, ViT feature extraction, LSTM progression prediction, Grad-CAM explanation, volu-metric analysis, and Marching Cubes reconstruction with Three.js/Plotly visualization. The backend coordinates upload, model execution, stored results, and report generation through RESTful services. Authentication, storage, notifications, and model-update services are deployment components represented in the original architecture.

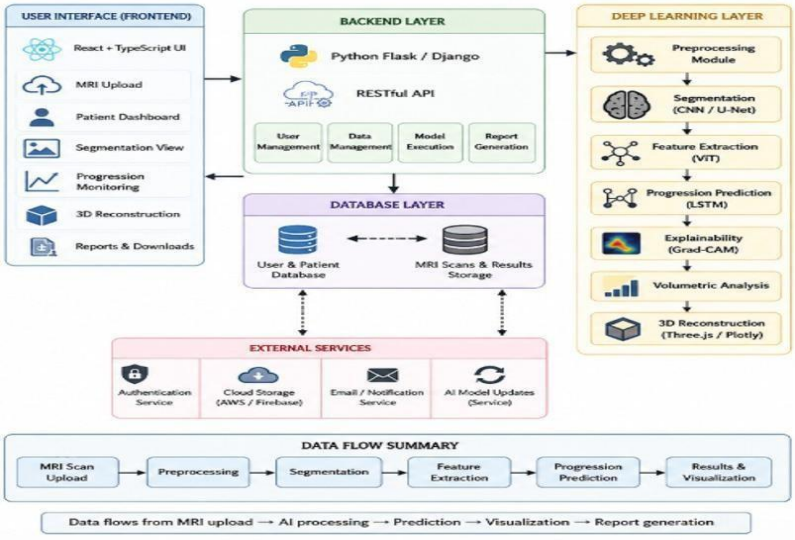


Fig. 1. System Architecture

4.3 Data Acquisition and Dataset Description

The MRI-analysis workflow uses the BraTS, TCIA, and Kaggle datasets/resources described in the project im-plementation. BraTS provides multimodal glioma data, including T1, T1ce, T2, and FLAIR volumes with tumor-subregion annotations [8, 10]. TCIA provides cancer-imaging collections relevant to longitudinal MRI analysis [9], while Kaggle provides additional brain-tumor MRI examples.

4.3 Data Security and Access Control

Sensitive stored data are protected with AES-256 encryp-tion, network communication uses TLS 1.3, and access is governed through secure authentication, encrypted cloud storage, and role-based controls. These safeguards protect data storage and transmission but do not consti-tute a claim of regulatory or clinical compliance.

4.4 MRI Upload and Validation

The interface accepts local MRI data and supported sources, then verifies the file format and dimensional consistency before processing. Figure 2 shows the MRI upload interface. Imaging data are organized by session and converted to NIfTI where required; NIfTI conver-sion and metadata handling preserve voxel spacing and orientation for later volume calculation [19].

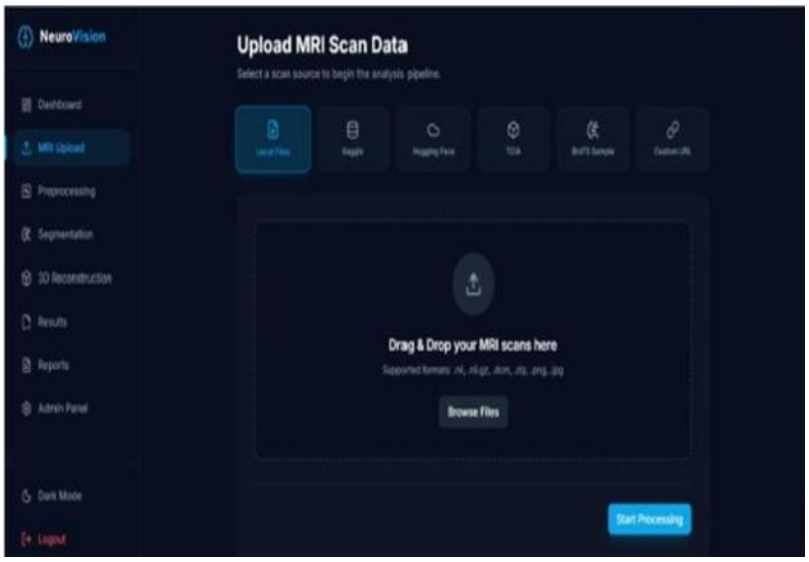


Fig. 2. MRI Scan Upload Interface

4.5 Preprocessing

Preprocessing applies skull stripping, Z-score intensity normalization, Gaussian denoising, and spatial resam-pling to a 256 × 256 × 256 voxel grid. Skull stripping removes non-brain structures, normalization reduces intensity-scale variation, denoising suppresses noise, and resampling supplies a common spatial grid. Fig-ure 3 shows the preprocessing interface, while Fig. 4 shows an active preprocessing run. Established brain-extraction and bias-field-correction methods provide context for MRI standardization [18, 17].

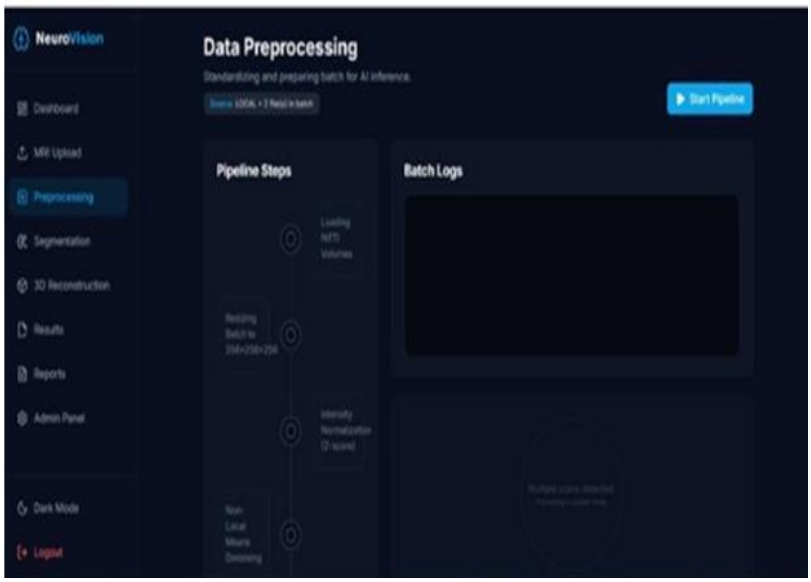


Fig. 3. MRI Data Preprocessing Interface



Fig. 4. MRI Preprocessing Pipeline Interface

4.6 Tumor Segmentation

The U-Net encoder–decoder produces a voxel-level tu-mor mask. Its role is specifically boundary and region segmentation; it is not the temporal predictor. Skip-connected encoder–decoder designs are established for biomedical segmentation [5], and volumetric variants provide related 3D context [15, 16]. Figure 5 displays the retained AI segmentation interface.

For an automated mask P and reference mask G, the reported overlap metrics are defined by Eq. (1) and Eq. (2):

$$Dice\,(P, G) = \frac{2|P \cap G|}{|P| + |G|} \quad (1)$$

$$IoU\,(P, G) = \frac{|P \cap G|}{|P \cup G|} \quad (2)$$

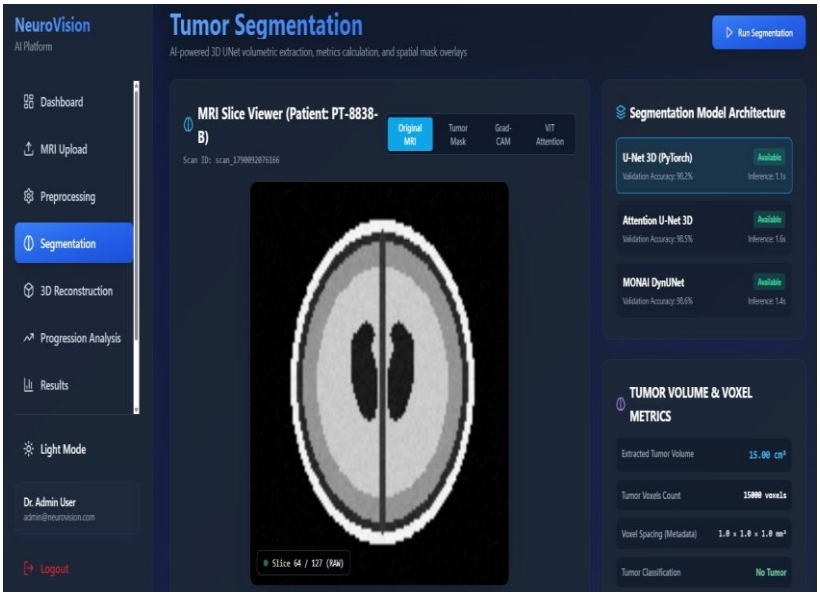


Fig. 5. AI Tumor Segmentation Interface

4.7 ViT Spatial Feature Extraction

Each segmented MRI slice is divided into fixed-size patches and processed by a ViT encoder with multi-head self-attention. The ViT module captures patch-level context, irregular tumor boundaries, lesion texture, and long-range spatial relation-ships [11]. This module produces spatial feature vectors; it does not replace the U-Net tumor mask.

4.8 LSTM Temporal Modelling

Patient-specific ViT feature vectors are arranged in clinical time order and passed to the LSTM. The recurrent module learns sequential change and produces progression output for increasing, decreasing, or stable tumors. LSTM is used because its gated state updates are designed for ordered dependencies [12].

4.9 Tumor Volume Estimation

The physical tumor volume is obtained from the number of mask voxels and the voxel spacing. If NT is the tumor-voxel count and sx, sy, and sz are voxel dimensions in millimetres, volume in cubic centimetres is calculated using Eq. (3):

$$V_{\mathrm{cm}3} = \frac{N_T S_x S_y S_z}{1000} \qquad (3)$$

This calculation distinguishes physical volume from a raw pixel or voxel count.

4.10 3D Reconstruction

Mask contours are converted into a patient-specific triangulated surface using the Marching Cubes algorithm [21] and rendered as an interactive 3D model. The reconstructed output includes tumor volume, surface area, and maximum diameter. Three.js/Plotly provides interactive visualization, while established platforms such as 3D Slicer illustrate the broader use of 3D visualization in medical-image computing [20]. Figure 6 retains the developed reconstruction interface.

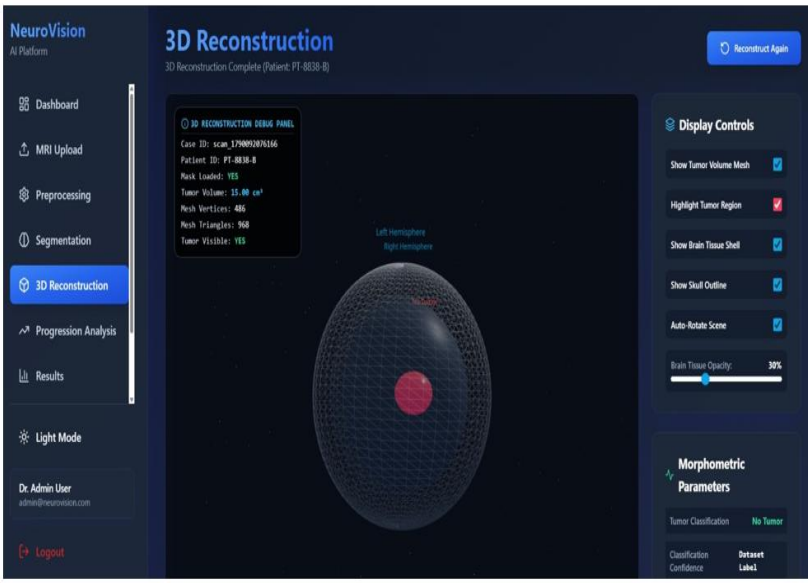


Fig. 6. 3D Volumetric Reconstruction Interface

4.11 Explainable AI and Reporting

The system generates Vision Transformer self-attention maps and Grad-CAM visualizations that highlight regions con-tributing to the model output [13]. These explanatory views accompany, rather than replace, segmentation masks and quantitative measurements. The report combines masks, volume values, progression graphs, 3D views, and explanation maps. Medical-XAI literature emphasizes that such explanations require careful interpretation and validation [3].

4.12 Data Processing Workflow

Figure 7 summarizes the implemented order. The written workflow matches the diagram: dataset assembly → upload and validation → preprocessing → U-Net segmentation → ViT spatial feature extraction →

chronological data prepara-tion → LSTM progression prediction → volumetric analysis → 3D reconstruction → Grad-CAM explanation → result visualization and report generation. Clinical interpretation remains outside the automated model and is performed by qualified users.

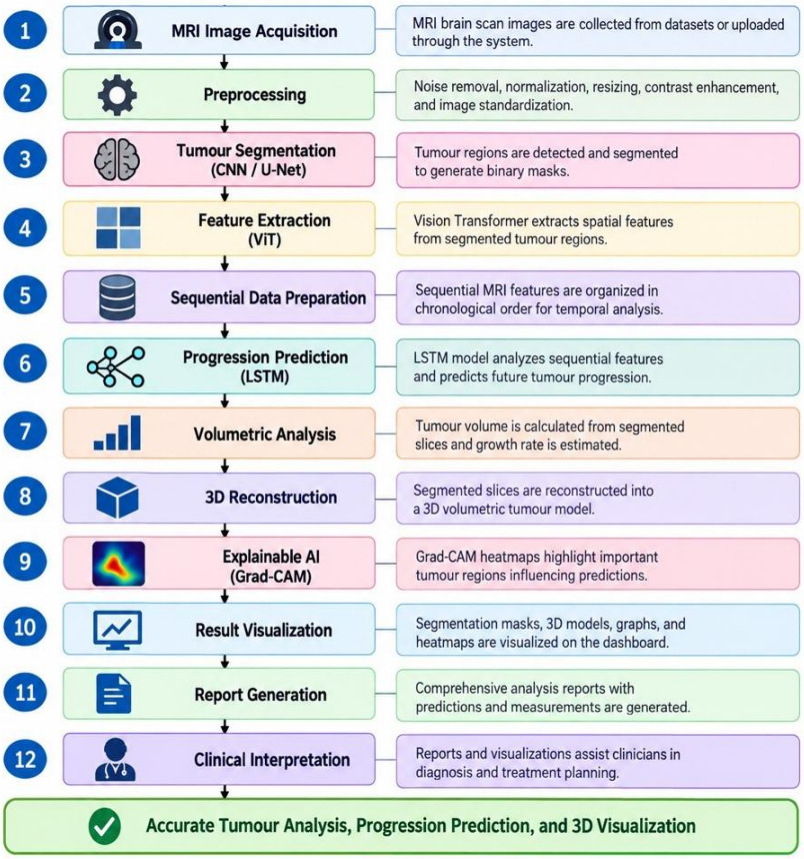


Fig. 7. Data Processing Workflow of the Proposed System

4.13 Functional Testing

Table 2 presents module-level functional tests covering upload, preprocessing, segmentation, progression prediction, volumetric analysis, and dashboard performance. Demonstration cases include TC-MRI-01 for MRI upload, TC-PRE-01 for preprocessing, TC-SEG-01 for segmentation, TC-3DR-01 and TC-3DR-02 for 3D reconstruction and incomplete-mask handling, TC-VIT-01 for ViT–LSTM progression prediction, TC-XAI-01 for explainable AI, TC-PROG-01 for tumor-progression analysis, and TC-RES-01 for results and clinical-summary generation. All primary functional test cases passed, while the incomplete-segmentation edge case (TC-3DR-02) failed because insufficient segmentation data pre-vented complete 3D reconstruction. These tests demonstrate interface and pipeline execution and are distinct from clini-cal validation.

Table 2 Functional Test Cases

Module	Total	Pass	Fail	Rate
Upload/validation	5	5	0	100%
Preprocessing	4	4	0	100%
Segmentation	4	4	0	100%
Progression	4	4	0	100%
Volumetric analysis	4	3	1	75%
Dashboard	3	3	0	100%

5. Results and Performance Evaluation

5.1 Verified Segmentation Results

The U-Net segmentation module achieved a Dice Similarity Coefficient of 95.8% and an IoU of 91.5% in the demonstrated evaluation. The Dice score measures overlap between the predicted and reference masks, while IoU measures their intersection relative to their union, as defined by Eq. (1) and Eq. (2).

5.2 Demonstrated Test-Case Outputs

In a representative progression test case, the ViT–LSTM module predicted a future tumor volume of 18.90 cm3 with 85.0% prediction confidence. This confidence is specific to the demonstrated case and is not presented as overall model accuracy. A representative 3D-reconstruction output produced a patient-specific tumor mesh with a volume of 13.80 cm3 and a surface area of 28.4 cm2. The reconstruction interface also reports maximum diameter. These values are individual test-case outputs rather than aggregate performance measures.

5.3 Qualitative System Capabilities

The ViT–LSTM module was integrated for longitudinal tumor-progression prediction. Preprocessing performs resizing, normalization, denoising, and MRI preparation for downstream analysis. The system generates interactive 3D tumor reconstructions using Marching Cubes and provides Vision Transformer self-attention and Grad-CAM visualizations that highlight regions contributing to model output.

5.4 Benefits and Improvements

The integrated pipeline provides progression and volumetric-change information in addition to individual masks. Each prediction is accompanied by attention and Grad-CAM visualizations. Automation connects NIfTI ingestion, preprocess-ing, segmentation, progression prediction, 3D rendering, and report assembly. Unlike a static single-scan tool, the LSTM processes ViT spatial signatures across chronologically ordered visits. Generalization across scanners and demographics remains a subject for future external validation.

6. Applications

6.1 Hospital Radiology Departments

The system can prepare segmentation output and report drafts for review in high-volume radiology workflows, poten-tially reducing repetitive boundary-marking work while leaving final interpretation to qualified clinicians.

6.2 Oncology Treatment Monitoring

Tumor-volume comparison across chemotherapy, radiation therapy, and follow-up visits can present measurable ev-idence of shrinkage, stability, or growth to support treatment-response assessment. The system does not determine treatment independently.

6.3 Neurosurgical Planning

Interactive 3D reconstruction can provide a consolidated view of tumor shape, depth, and spread to supplement conven-tional two-dimensional slices during multidisciplinary planning.

6.4 Medical Imaging Research

The pipeline provides a reference workflow for segmentation, longitudinal prediction, volumetric analysis, reconstruc-tion, and explainable AI using public imaging resources such as BraTS and TCIA.

6.5 Clinical Decision Support

The system supports, rather than replaces, clinical judgment by presenting confidence information, masks, volume esti-mates, progression output, and Grad-CAM explanations together for review and discussion.

7. Conclusion and Future Work

The proposed workflow integrates MRI validation and preprocessing, U-Net-based tumor segmentation, ViT spatial feature extraction, chronological LSTM temporal modelling, voxel-based volumetric analysis, 3D reconstruction using Marching Cubes, Explainable AI, and an integrated web dashboard for visualization and structured reporting. Seg-mentation evaluation achieved a Dice score of 95.8% and an IoU of 91.5%. The workflow combines spatial, temporal, quantitative, visual, and explanatory outputs, while the modular structure permits segmentation, prediction, volumetric analysis, reconstruction, and visualization components to be evaluated and upgraded separately. The ViT–LSTM framework, volumetric estimation, and Grad-CAM integration are completed components. Future work will evaluate the system on larger longitudinal datasets, extend multiclass progression analysis, improve temporal and volumetric forecasting, assess generalization across scanners and protocols, include additional tumor types, and undertake formal clinical validation and integration.

References

[1] Hanaa ZainEldin, Samah A. Gamel, El-Sayed M. El-Kenawy, Amal H. Alharbi, Doaa Sami Khafaga, Abdelhameed Ibrahim, and Fatma M. Talaat. "Brain Tumor Detection and Classification Using Deep Learning and Sine-Cosine Fitness Grey Wolf Optimization." Bioengineering 10, no. 1 (2023): 18. https://doi.org/10.3390/bioengineering10010018.

[2] Qiran Jia and Hai Shu. "BiTr-Unet: A CNN–Transformer Combined Network for MRI Brain Tumor Segmentation." In Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries, 3–14. Lecture Notes in Computer Science. Cham: Springer, 2022. https://doi.org/10.1007/978-3-031-09002-8_1.

[3] Erico Tjoa and Cuntai Guan. "A Survey on Explainable Artificial Intelligence (XAI): Toward Medical XAI." IEEE Transactions on Neural Networks and Learning Systems 32, no. 11 (2021): 4793–4813. https://doi.org/10.1109/TNNL S.2020.3027314.

[4] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. "Swin Transformer: Hierarchical Vision Transformer Using Shifted Windows." In Proceedings of the IEEE/CVF International Conference on Computer Vision, 10012–10022, 2021. https://doi.org/10.1109/ICCV48922.2021.00986.

[5] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. "U-Net: Convolutional Networks for Biomedical Image Seg-mentation." In Medical Image Computing and Computer-Assisted Intervention– MICCAI 2015, 234–241. Cham: Springer, 2015. https://doi.org/10.1007/978-3-319-24574-4_28.

[6] Fabian Isensee, Paul F. Jaeger, Simon A. A. Kohl, Jens Petersen, and Klaus H. Maier-Hein. "nnU-Net: A Self-Configuring Method for Deep Learning-Based Biomedical Image Segmentation." Nature Methods 18, no. 2 (2021): 203–211. https://doi.org/10.1038/s41592-020-01008-z.

[7] Palani Thanaraj Krishnan, Pradeep Krishnadoss, Mukund Khandelwal, Devansh Gupta, Anupoju Nihaal, and T. Sunil Kumar. "Enhancing Brain Tumor Detection in MRI with a Rotation Invariant Vision Transformer." Frontiers in Neuroinformatics 18 (2024): 1414925. https://doi.org/10.3389/fninf.2024.1414925.

[8] Bjoern H. Menze et al. "The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS)." IEEE Transactions on Medical Imaging 34, no. 10 (2015): 1993–2024. https://doi.org/10.1109/TMI.2014.2377694.

[9] Kenneth Clark et al. "The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repos-itory." Journal of Digital Imaging 26, no. 6 (2013): 1045–1057. https://doi.org/10.1007/s10278-013-9622-7.

[10] Ujjwal Baid et al. "The RSNA-ASNR-MICCAI BraTS 2021 Benchmark on Brain Tumor Segmentation and Radio-genomic Classification." arXiv (2021). https://doi.org/10.48550/arXiv.2107.02314.

[11] Alexey Dosovitskiy et al. "An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale." In Inter-national Conference on Learning Representations, 2021. https://doi.org/10.48550/arXiv.2010.11929.

[12] Sepp Hochreiter and Jürgen Schmidhuber. "Long Short-Term Memory." Neural Computation 9, no. 8 (1997): 1735–1780. https://doi.org/10.1162/neco.1997.9.8.

[13] Ramprasaath R. Selvaraju et al. "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localiza-tion." In Proceedings of the IEEE International Conference on Computer Vision, 618–626, 2017. https://doi.org/10.110 9/ICCV.2017.74.

[14] Ali Hatamizadeh et al. "UNETR: Transformers for 3D Medical Image Segmentation." In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 574–584, 2022. https://doi.org/10.1109/WACV51458.2022.0 0181.

[15] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. "V-Net: Fully Convolutional Neural Networks for Vol-umetric Medical Image Segmentation." In 2016 Fourth International Conference on 3D Vision, 565–571. IEEE, 2016. https://doi.org/10.1109/3DV.2016.79.