International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 15, Issue 01, January 2026)

Multi-Tasking Brain Imaging Analysis with Integrated CT & MRI: Survey of Models, Flows, and Frameworks

Dr. T. Suvarna Kumari[1], Rishitha Konidena[2]

[1]*Asst. Professor, Dept. of Computer Science and Engineering, Chaitanya Bharathi Institute of Technology (A), Hyderabad, India*

[2]*Dept. of Computer Science and Engineering, Chaitanya Bharathi Institute of Technology (A), Hyderabad, India*

***Abstract*— The rapid growth of brain imaging data from CT and MRI scans has created a crucial need for automated frameworks that go beyond single task AI models. This survey takes a closer look at recent advances in multi task learning and multimodal architectures designed for brain image analysis. We concentrate on how current systems integrate segmentation, classification, severity scoring, and into cohesive pipelines that manage data from MRIs and CT scans. The reviewed studies are grouped by their model architectures, workflow designs, vision language strategies, and approaches for explainability. Notable progress includes the use of parallel encoder–decoder setups, transformer-based captioning, and GradCAM-driven visual explanations. However, most studies still concentrate on particular tasks or imaging modalities. Variations in datasets, limited model interpretability, a lack of foundation in report generation, and practical challenges in clinical adoption are still major obstacles. This survey highlights gaps including the absence of fully integrated CT and MRI systems, minimal attention to longitudinal data, and the lack of consistent evaluation methods. Finally, we provide recommendations for creating more dependable, transparent, and scalable AI frameworks for neuroimaging.**

***Keywords*—Brain imaging, multi-task learning, CT, MRI, Segmentation, Classification, Explainable AI, Report Generation, Deep Learning, BraTS, RSNA**

I. Introduction

Modern healthcare diagnostic workflows have faced a significant transformation due to the rapid growth in neuroimaging data from CT and MRI. The need for quick and accurate interpretation has surpassed what radiologists can do by hand due to the increasing number of multislice and multimodal scans. Deep learning frameworks that can automatically perform classification, segmentation, and clinical report generation across a variety of imaging modalities have been prompted by these challenges.

Early studies on automated brain image captioning which we explored has hybrid designs that combined convolutional feature extractors with language models.

Sequential captioning techniques using pre-trained classifiers integrated with LSTM or GPT like decoders showed the potential of generating natural language interpretations of CT images [1]– [4]. The introduction of generative pre-trained transformers for medical image captioning further improved fluency and contextual accuracy [2], [5]– [7]. However, many of these systems lacked domain grounding and factual precision, which led to the rise of vision language transformers and attention-based mechanisms [8], [9]

Despite these advancements, current literature remains fragmented across separate tasks and modalities. Few studies achieve end-to-end integration of CT and MRI analysis. Moreover, explainability, factual accuracy in generated text, and computational efficiency in 3D transformer models continue to limit clinical adoption [5], [10], [11]. There is also a pressing need for harmonized evaluation metrics and standardized datasets to support consistent benchmarking.

This survey reviews 18 representative works from 2019–2025 exploring multi-task and multi-modal brain imaging. It categorizes developments in architecture design, workflow optimization, attention-based vision language methods, and explainable AI frameworks to identify gaps and future directions for clinically interpretable CT and MRI analysis systems.

II. Literature Survey

A. CT Hemorrhage Classification

Early works in CT hemorrhage detection applied convolutional and attention-based classifiers to the RSNA Intracranial Hemorrhage dataset. Selivanov et al. [1] proposed a sequential captioning system combining pre trained CNN classifiers and an LSTM language model for CT interpretation, improving interpretability though limited to single task settings.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 15, Issue 01, January 2026)

Khan and Ali [6] demonstrated the efficiency of generic deep learning models such as ResNet50 and DenseNet121 for subtype detection, achieving an average AUC of 0.91 across hemorrhage classes. Rahman and Kim [10] extended multi-task learning to COVID CT classification, illustrating generalizable architectures for multi output imaging tasks. Despite their robustness, these models remain CT only, focusing on detection without integrating segmentation or language understanding components.

B. MRI Tumor Segmentation and Analysis

MRI segmentation remains a core benchmark for 3D medical vision. Kumar and Reddy [12] employed Grad-CAM on ResNet50 based MRI classifiers to enhance explainability for tumor detection. Singh and Sharma [13] developed an attention-based CNN for Alzheimer's detection, highlighting the promise of spatial attention mechanisms in neurodegenerative disease analysis. Park and Choi [14] moved beyond static segmentation by linking MRI segmentation masks to LLMbased textual explanation, demonstrating the first step toward image to text interpretation in volumetric scans. Joh and Baik [11] created a web deployed explainable brain tumor platform integrating 3D segmentation and visualization. While effective on BraTS datasets, most MRI systems are computationally intensive and rarely cross learn from CT modalities, leaving multi-modal learning largely unexplored.

C. Medical Image Captioning and Automated Report Generation

Medical image captioning has evolved from CNN–RNN hybrids to transformer-based architectures. Dylov and Fedulova [2] introduced GPT based models for generating clinically relevant captions, whereas Deep Image Captioning by Karpathy et al. [4] surveyed end-to-end captioning pipelines, motivating later medical adaptations. Hierarchical frameworks such as that of Patel et al. [3] integrated classification outputs with captioning modules, offering structured report generation. Varol-Arisoy et al. [8] proposed DualPrompt-MedCap, introducing prompt tuning to improve domain adaptation for medical text generation.

Arisoy and Uysal's Vision AttentionDriven Language Framework [9] and Cherukuri et al.'s GCSM3VLT transformer [15] has further improved factual reliability through guided self-attention mechanisms. Similarly, DS@BioMed's 2024 Image CLEF submission [16] demonstrated that incorporating concept detection modules enhances both fluency and content relevance in generated medical captions.

Explainability remains very essential for the development of trustworthy clinical AI solutions. Gupta and Singh [5] reviewed techniques such as Grad-CAM, LIME, and SHAP in deep medical imaging models, stressing their significance for clinical interpretability. Rahman and Kim [10] showed that explainable multitask architectures can boost diagnostic confidence in CT image analysis. Joh and Baik [11] built a cloud based explainable platform for MRI tumor detection, providing transparent and traceable predictions. Cherukuri et al. [15] explored multimodal transformers that integrate both vision and language streams, while Lee et al. [17] proposed a retrieval augmented approach that is combined with text domain expansion to improve factual accuracy in MRI report generation. Although these studies mark significant steps toward multimodal explainability, a completely unified CT and MRI framework is yet to be achieved.

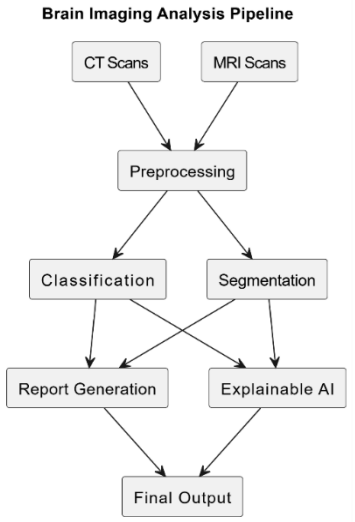

Fig. 1. Generic brain imaging analysis pipeline synthesizing major literature on CT and MRI tasks

D. Summary and Gaps

Across all surveyed literature, four gaps persist: 1) Lack of unified CT and MRI fusion: existing systems operate on single modalities. 2) Limited multitask coupling: classification, segmentation, and captioning are treated as isolated stages. 3) Explainability not standardized: visualization quality and clinical interpretability vary widely. 4) Deployment bottlenecks: most studies stop at offline validation without web scale or patient friendly deployment.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 15, Issue 01, January 2026)

Future research including the proposed FYP framework should emphasize integrated CT with MRI fusion, end-to-end explainable pipelines, and report generation understandable to both clinicians and patients.

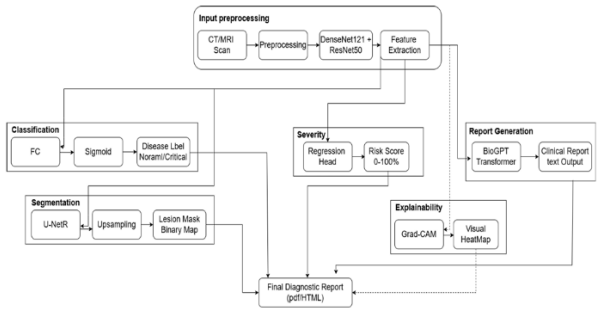

Fig. 2. Proposed unified pipeline for brain imaging analysis.

The proposed Integrated Multi-Task AI Framework for Brain Image Analysis and Reporting include automated interpretation of brain CT and MRI scans by a single deep-learning pipeline, which performs classification, segmentation, report generation, and explanation. The system is modular and clinically interpretable, hence making it easy to integrate reliably into existing hospital workflows.

TABLE I:
Observations on Different Research Papers

S.No.	Title	Year	Methodology	Observed Features	Limitations
1	Sequential Brain CT Image Captioning Based on the PreTrained Classifiers and a Language Model [1]	2023	CNN + LSTM	• Generates captions from CT scans • Uses pre-trained classifiers • Multi-label prediction	• Works only on CT • Captions miss clinical details • Limited interpretability
2	Medical Image Captioning via Generative Pretrained Transformers [2]	2023	GPT-2 Transformer	• Uses transformers for captioning • Outputs fluent text • Works on multiple image types	• Needs large training data • Factual accuracy sometimes low • Requires domain tuning
3	Automatic Interpretation of Brain Medical Images Using Hierarchical Classification and Image Captioning [3]	2024	Hierarchical CNN-RNN	• Performs hierarchical classification • Combines with image captioning • Works for CT and MRI	• Captions lack clinical accuracy • Not robust for rare cases • Still single-task focused
4	Deep Image Captioning: A Review [4]	2022	Literature Review	• Comprehensive review of methods • Shows main trends • Outlines future challenges	• No experimental validation • Doesn't compare performance • Lacks clinical discussion
5	Explainable AI (XAI) in Deep Learning-Based Medical Image Analysis [5]	2023	Grad-CAM, LIME, SHAP Review	• Covers explainability methods • Focuses on Grad-CAM and saliency • Highlights clinical need for XAI	• Few workflows apply XAI • Limited real clinical use • No performance benchmarks
6	Medical Image Analysis Using Deep Learning Algorithms [6]	2021	CNN and GAN models	• Reviews CNN applications • Works across multiple diseases • Demonstrates strong versatility	• Models can overfit • Focused mainly on classification • No explainability module
7	Intra- and Inter-Head Orthogonal Attention for Image Captioning [7]	2025	Transformer with Orthogonal Attention	• Uses attention head efficiently • Improves image-text alignment • Enhances captioning quality	• Needs clinical datasets • Not tested for multi-tasking • Limited to captioning

8	Dual Prompt-MedCap:Dual-Prompt Approach for Medical Captioning [8]	2024	Vision-Language Transformer	• Dual prompts boost caption quality • Adds context to medical reports • Based on transformer backbone	• Reports still basic • No multimodality tested • Needs fact-checking module
9	Vision Attention Driven Language Framework for Medical Report Generation [9]	2025	Vision Transformer + Text Decoder	• Attention links image and text • Generates structured reports • Supports multiple image sources	• Factuality sometimes low • Needs clinician validation • Poor on rare pathologies
10	DS@ BioMed at Image CLEF medicalCaption 2024 [16]	2024	CNN + Concept Detection	• Concept detection boosts captions • Competes on open benchmark • Uses attention mechanisms	• Limited dataset size • Not a deployable model • Captions may lack depth
11	GCS-M3VLT: Guided Context Self-Attention Multi-Modal Transformer [15]	2025	Multi-Modal Transformer	• Fuses vision and language • Applies self-attention context • Handles multimodal input	• High computation cost • Requires labeled datasets • Not clinically validated
12	Improving Factuality of 3D Brain MRI Report Generation [17]	2025	Retrieval + Transformer	• Produces factual MRI reports • Uses paired retrieval	• Needs larger MRI dataset • Works on MRI only • Lacks full integration
13	SKIP Net: Enhanced Brain Tumor Classification [18]	2024	CNN with Skip Attention	• Detects tumors precisely • Tested on BraTS dataset	• Only for tumor classification • No segmentation link • Not multimodal
14	Attention-BasedCNN for Alzheimer' s Classification [13]	2023	Attention CNN	• Uses attention visualization • Detects Alzheimer's accurately • Employs biomedical signals	• Only for Alzheimer's • No segmentation/reporting • Needs broader validation
15	Enhancing Brain Tumor Detection Using Grad-CAM with ResNet50 [12]	2023	ResNet50 + Grad-CAM	• Provides visual heatmaps • Improves explainability • Good for MRI tumors	• Heatmaps need expert review • MRI-only dataset • No text output
16	Segmentation to Explanation: MRI Reports via LLMs [14]	2024	U Net + LLM	• Segments MRI and writes reports • Uses large language models • End-to-end interpretability	• High computing cost • MRI-only domain • Reports need factual checks
17	Explainable multi-Task COVID CT Framework [10]	2022	CNN + Grad-CAM	• Multi-task on COVID CT data • Adds explainable outputs • Severity scoring included	• Only for COVID CT • No MRI data • Limited generalization
18	Web-DeployedExplainable AI Brain Tumor Platform [11]	2025	Web-Based Explainable Model	• Online platform for AI scans • Integrates XAI visualization • Designed for tumor diagnosis	• Only for tumors • Not multi-task pipeline • Needs clinical scalability

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347-6435(Online) Volume 15, Issue 01, January 2026)

TABLE II:
Comparative Insight: Current State vs. Identified Gaps in Surveyed Domains

S. No.	Domain / Research Focus	Current State (Surveyed Works)	Identified Gaps / Limitations	Future Research Direction
1	CT Hemorrhage Classification	• CNN and attention models achieve 90%+ accuracy. • Models trained on RSNA dataset for subtype detection.	• No integration with report generation. • Works only on CT scans, not cross modality.	• Develop CT–MRI unified classification pipelines. • Add explainable, language linked outputs.
2	MRI Tumor Segmentation	• U-Net and 3D CNNs perform well on BraTS data. • Explainable Grad-CAM models aid visualization.	• Segmentation and classification handled separately. • No text reporting or multi-task extension.	• Fuse segmentation with report generation. • Introduce multimodal (CT+MRI) learning.
3	Medical Image Captioning	• Transformer-based and GPT-based models produce fluent captions. • DualPrompt and Vision-Language Transformers improve contextual relevance.	• Lacks factual grounding and medical accuracy. • No multi-task or cross-modality integration.	• Use medical ontology for factuality. • Merge captioning with diagnosis and segmentation.
4	Explainable AI (XAI) and Visualization	• Grad-CAM and SHAP explain decisions visually. • Web-based tools enable user interpretability.	• Explainability remains post-hoc. • Few models link visual explanation to text.	• Build real-time, integrated XAI pipelines. • Use multimodal attention maps for joint reasoning.
5	Multi-Task / Multi-Modal Integration	• Transformers and CNN hybrids show early fusion attempts. • Some frameworks merge classification + captioning.	• CT–MRI fusion rarely attempted. • Absence of common standards.	• Create complete CT–MRI multi-task pipelines. • Develop standardized evaluation metrics.

III. Discussion

The reviewed literature across CT and MRI domains reveals a strong evolution from task specific deep learning toward more unified, interpretable, and multimodal frameworks. Early models mainly focused on single task objectives such as classification [6], [13], [18], segmentation [12], [14], or captioning [1]– [3], with limited cross domain interaction. Although each achieved notable accuracy and interpretability improvements, they often lacked the contextual and clinical reasoning necessary for end-to-end automation. Vision Attention Driven Language models [9], and GCS-M3VLT [15] represent a shift toward multimodal integration. These models show how attention-based transformers can more easily understand medical context by improving alignment between image features and textual outputs.

Explainability has emerged as a parallel research thread [5], [10]– [12]. Although Grad-CAM, LIME, and SHAP provide better interpretability, there is still little integration of these technologies into generative or decision support systems. Studies like Joh and Baik's web deployed explainable AI platform [11] show practical promise but still function as standalone prototypes rather than components of multitask clinical systems.

Across surveyed works, three major patterns are evident:

1) *Task Fragmentation:* Classification, segmentation, and captioning are typically executed in isolation rather than as coordinated tasks.
2) *Data Silos:* Most datasets (e.g., RSNA for CT, BraTS for MRI) remain domain specific, preventing models from learning cross-modality relationships.

3) *Limited Clinical Integration:* While transformer based captioning systems generate readable outputs, their factual and diagnostic accuracy remains low compared to radiologist reports.

Finally, the integration of generative pretrained models (e.g., GPT-based [2]) with medical imaging architectures can be done but underexplored section. To achieve this, future frameworks must focus on unified CT and MRI representation learning, multimodal fusion transformers, and interpretable language vision reasoning models.

IV. Conclusion

This survey has systematically reviewed eighteen research works on multitask and multimodal brain image analysis across both CT and MRI domains. The combined results show that although individual advancements in caption generation, classification, and segmentation have demonstrated strong performance, they are still mainly dispersed across individual tasks or modalities. Vision language transformers [8], [9], [15] and explainable AI models [5], [11], [12] have actually contributed significantly toward interpretability and contextual reasoning, still a clinically deployable, an end-to-end CT and MRI integrated framework is still absent.

The comparison of existing methodologies indicates three persistent gaps: (1) lack of unified CT and MRI representation learning, (2) weak factual grounding in medical captioning, and (3) limited clinical validation for explainable models.

A successful future framework will be one that not only predicts accurately but also validates its reasoning an ultimate fostering trust, transparency, and clinical usability in AI driven brain imaging.

REFERENCES

[1] A. Selivanov, O. Y. Rogov, and D. Chesakov, "Sequential brain ct image captioning based on the pre-trained classifiers and a language model," IEEE Access, pp. 1–8, 2023.

[2] D. Dylov and I. Fedulova, "Medical image captioning via generative pretrained transformers," Nature Scientific Reports, pp. 1–10, 2023.

[3] W. S. Mayzura, R. Sarno, and N. Setiawan, "Automatic interpretation of brain medical images using hierarchical classification and image captioning model," in IEEE International Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2024, pp. 120– 128.

[4] X. Zhang, A. Jia, and J. Ji, "Deep image captioning: A review of methods, trends and future challenges," Pattern Recognition Letters, pp. 102–115, 2022.

[5] A. Gupta and M. Singh, "Explainable artificial intelligence (xai) in deep learning-based medical image analysis," Frontiers in Artificial Intelligence, pp. 1–12, 2023.

[6] T. Khan and S. Ali, "Medical image analysis using deep learning algorithms," Computers in Biology and Medicine, pp. 78–90, 2021.

[7] X. Zhang and D. Lee, "Intra- and inter-head orthogonal attention for image captioning," IEEE Transactions on Neural Networks and Learning Systems, pp. 1–9, 2025.

[8] M. Varol-Arisoy, A. Arisoy, and I. Uysal, "Dualprompt-medcap: A dual-prompt enhanced approach for medical image captioning," Expert Systems with Applications, pp. 100–111, 2024.

[9] A. Arisoy and I. Uysal, "A vision attention driven language framework for medical report generation," Medical Image Analysis, pp. 1–12, 2025.

[10] H. Rahman and J. Kim, "Explainable multi-task covid ct framework," Computers in Biology and Medicine, pp. 30–41, 2022.

[11] H. K. Joh and S. H. Baik, "Web-deployed explainable ai brain tumor platform," IEEE Access, pp. 1–10, 2025.

[12] S. Kumar and P. Reddy, "Enhancing brain tumor detection in mri images through explainable ai using grad-cam with resnet50," IEEE Access, pp. 1–9, 2023.

[13] R. Singh and P. Sharma, "An attention-based cnn architecture for alzheimer's classification and detection," Biomedical Signal Processing and Control, pp. 56–66, 2023.

[14] S. Park and J. Choi, "From segmentation to explanation: Generating textual reports from mri with llms," Frontiers in Radiology, pp. 200–212, 2024.

[15] T. K. Cherukuri, N. S. Shaik, and D. H. Ye, "Gcs-m3vlt: Guided context self-attention based multi-modal medical vision language transformer for retinal image captioning," IEEE Journal of Biomedical and Health Informatics, pp. 1–10, 2025.

[16] N. Nguyen, H. Tu, and P. Nguyen, "Ds@biomed at imageclefmedical caption 2024: Enhanced attention mechanisms in medical caption generation through concept detection integration," in ImageCLEF Medical Workshop, 2024, pp. 34–42.

[17] J. Lee, Y. Oh, and D. Lee, "Improving factuality of 3d brain mri report generation with paired image-domain retrieval and text-domain augmentation," Medical Image Analysis, pp. 88–100, 2025.

[18] R. Patel and A. Kumar, "Skipnet: Spatial attention skip connections for enhanced brain tumor classification," Neurocomputing, pp. 45–54, 2024.