INFORMATICS INSTITUTE OF TECHNOLOGY

In Collaboration with

ROBERT GORDON UNIVERSITY ABERDEEN

Multimodal Brain Tumor Detection System

Group 21 Literature Review by

Chanura Karunanayake - 20240955

Inzaman Careem - 20241672

Arunasalam Kabilash - 20241793

Lahiru Bashitha - 20232524

Supervised by

Ms. Suvetha Suvendran

October 2025

Table of Contents

Introduction

Brain tumor detection remains one of the most complex challenges in the medical field, traditionally dependent on MRI or CT scan images. However, textual clinical notes written by doctors often contain valuable contextual information, such as previous test results, observations, and symptoms, which are not directly visible in medical images. To address these limitations, this project proposes a Multimodal Brain Tumor Detection and Classification System that integrates both textual and visual modalities. The main purpose will be to enhance diagnostic accuracy by identifying patterns that exist in both images and clinical data.

The proposed system separately runs 4 models, where the Clinical Text Feature Extraction Model generates semantic embeddings and capturing key medical meanings from patient clinic data, then the Image Feature Extraction Model extract visual features from brain scanned images. After multimodal fusion, concatenation of textual and visual features produces a unified representation. Finally, fused vector and deep learning classifier detects and categorizes the tumors. This connected workflow improves the diagnostic accuracy.

Problem Domain

The proposed multimodal framework serves as a step toward an intelligent, explainable, and comprehensive diagnostic system. Mainly mirroring how real doctors reason from both patient records and imaging data.

Previous works

Clinical Text Feature Extraction Model

In 2025, Lee *et al.* studied utilized sources such as PubMed, MIMIC IV, and diagnostic codes. This model introduced new designs and improved methods that enhanced the model's scalability and performance. Clinical ModernBERT effectively categorizes electronic health records and retrieves them from PMC. It also delivers better computational efficiency than BioClinicalBERT. Clinical ModerBERT has a top-1 accuracy of 63.31%.

In 2025, Obeidat *et al.* studied uses of the BIO tagging scheme on five datasets to compare LLMs (Mistral, Llama) with encoder models (Bert, BiomedBERT, DeBERTa-v3). LLMs outperform encoders by 2-8% in most cases, with DeBERTa-v3-Large performing best and BiomedBERT being much faster. Accuracy: DeBERTa-v3-Larga reached a macro F1-score of 92.74%, while BiomedBERT was lower but 220x faster.

In 2025, Guleria *et al.* proposed a software framework that modifies traditional models into modern ones, using ML classifiers along with pre-trained BERT and LSTM models for analyzing medical transcripts. Results show LSTM and BERT outperform other ML classifiers in text Classification tasks. Accuracy: - LSTM reached a macro accuracy of 0.94, outperforming SVM (0.65) and CNN (0.66).

In 2025, Ben Abdennour *et al.* introduced using BERT for feature extraction, which assists in managing difficult medical languages. This enhances classification accuracy in comparison to conventional CNN and LSTM models that depend on typical text representations. Accuracy: - The BERT-CNN model reached a macro accuracy of 96.66%, outperforming CNN (94.99%) and LSTM (94.72%).

In 2025, Saeyeon Cheon *et al.* introduces a domain-specific NER dataset from infectious disease surveillance reports and presents Survice-BERT, a BERT-based model fine-tuned for pandemic-related information extraction. It supports early warning and forecasting of communicable diseases with strong extraction performance. Accuracy: - Survice-BERT achieved an average F1-Score of 0.99.

Image Feature Extraction Model

In 2025, S. Saranaya and G. Radhika conducted a research paper specifically using the VGG16 architecture for CNN architecture. VGG16 algorithm operates successfully with medical images, especially in brain tumor classification and network design includes minimal filters that facilitate improved feature detection and extraction capabilities. This model achieves Ten-fold cross-validation with approximately 99% high classification accuracy when successfully differentiating between tumor and no-tumor cases, identifying glioma and meningioma, and identifying meningioma and pituitary tumors.

Another paper, in 2023 by Rasheed, Z. introduces a CNN algorithm to classify three brain tumor types from MRI images, achieving 98.04% accuracy. Compared to models like VGG16, ResNet50, and InceptionV3, the algorithm shows strong potential for assisting doctors in accurately identifying tumor types. From the pre-trained models, ResNet50 achieved a greater 98.04% accuracy and 98% precision, recall, F1-scores.

In 2025, Ahammed, F. strictly tests the Brain Tumor MRI dataset to achieve very accurate tumor classification using deep learning models. The DenseNet169, EfficentNetB3 and VGG16 neural networks are used to analyze the data separately and make predictions. Within these models DenseNet169 and EfficientNetB3 both give 99.58% and VGG16 shows 99.41% accuracy.

In 2025, Limbani, N. conducted NeuroNet model for brain tumor categorization and segmentation. NeuroNet model combines a U-Net architecture with modifications for segmentation and a VGG16 inspired CNN for categorization and as well as XAI used for interpretability and reached 96.5% accuracy, 96.5% Precision ,95.5% Recall and 96% F1-Score.

In 2025**,** Thinger, C. researches a paper by allowing CNNs and CNN-SVM hybrids models for automatic, accurate and efficient of challenging datasets, deep learning-especially CNN has revolutionized medical imaging and reached CNN and CNN-SVM, Accuracy of 92.3% and 94.7%, Precision of 91.5% and 93.6% and F1 of Score 91.1% and 93.9. Hybrid models combine Support Vector Machines classification power with CNN feature extracting ability. Out of two methods, CNN-SVM model shown better generalizing capacity and handling of class imbalance.

Multimodal Fusion

Multimodal Fusion is the process of merging multiple types of modalities (text, audio, image, etc.) to get the most optimum output, minimizing the errors, compared to single-source approaches. The use of Multi Modal Fusion in medical diagnosis has shown incredible precision and accuracy (Jing Ru Teoh, 2024)**.**

Research in 2025 by (Rukmani, 2025) using Multimodal fusion of CT-MRI images showed higher accuracy and superior to models that incorporate only a single modality of data. This study used a feature fusion (concatenation-based) (early fusion) to merge the features extracted from CT images and MRI images to detect Brain Tumors.

Another Research conducted in 2025 (Lakshman Kumar Jamili, 2025), which specifically used MRI scan images and clinical notes (text data) for Tumor segmentation. They used an Attention-based fusion method in an intermediate stage to merge the 2 modalities. The research processed the image, using CNN-based Imaging Model, and the clinical Notes, using transformer-based large language model, and used attention-based fusion layer to combine them before segmentation. This method resulted in a higher precision: 0.82 and F1 score: 0.84 than the baseline model.

A Review done in 2020 (Shih-Cheng Huang, 2020), discusses effectiveness of numerous multimodal fusions systems for multiple medical fields which included “Prediction of survival time for brain tumor patient”, (Nie, D. *et al.*), which used early fusion with CNN extracted features of MRI images and patient data. This gave a result of: Fusion - 90.66% accuracy, MRI - 81.04% accuracy, Demographics and tumor features - 62.96% accuracy. Which gave an overall increased accuracy of 1.2 - 27.7% and AUROC of 0.02 - 0.16 over modern single modality models.

The Review paper by Sarathambekai, Ajith, Harish Krishna, Sailesh Kumar, Abijith (Sarathambekai, 2025) compares and reviews a process which is executed using Adaptive Multi-Scale and 3D Fusion Networks, which specialize in integrating information across different modalities. the paper also states fusion efficiency and interpretability are boosted by Attention Mechanisms, which dynamically weigh the importance of features during the combination step.

Classification & Diagnose Model

In 2024, Park, S. and Kim, J. compared CNN and Transformer-based pre trained models for brain tumor detection from MRI scans. Four models, VGG-16, ResNet-50, EfficientNetV2L, and ViT-B/16 - were trained using transfer learning with data augmentation and hyperparameter tuning. Using XAI methods (LIME, SHAP, Grad-CAM), they visualized decision regions. The results showed negligible gains from augmentation, while hyperparameter tuning improved performance. CNNs, such as VGG-16 and ResNet-50 achieved the highest accuracy and most interpretable visualizations.

In 2021, Dai, Y., Gao, Y. and Liu, F. proposed TransMed, a hybrid model combining CNN and Transformer for multi-modal medical image classification. TransMed efficiently extracts low-level image features and cross-modality high-level information, establishing long-range dependencies between modalities. The model was trained/ tested on parotid gland tumor and knee injury classification datasets. It achieved 10.1% and 1.9% average accuracy improvements over modern CNN-based models and fusion methods. The results demonstrate promising potential for broader medical image analysis applications.

In 2022, Zhang, Y. *et al.* proposes a Multimodal Medical Transformer (mmFormer) that influences hybrid modality-specific encoders and a modality-correlated encoder, and connects the Transformer and CNN to build the distant needs, both in and through different modalities. They validated mmFormer with BraTS 2018 dataset, etc. The proposed method outperforms the modern methods and demonstrates greater robustness to incomplete multimodal learning of brain tumor segmentation.

In 2022, Yeung, M. *et al.* introduced the Unified Focal Loss, a ranked framework that generalizes Dice and cross-entropy-based losses to effectively address class imbalance in medical image segmentation. They evaluated it against six related loss functions across 2D, 3D binary, and 3D multiclass segmentation tasks, and it proved robust to class imbalance and constantly outperformed other loss functions in segmentation quality. Their study highlights the critical role of loss function selection in achieving high-performance segmentation on imbalanced datasets.

In 2024, Durairaj, V. and Uthirapathy, P. proposed Trans-IMSM, a directed filter-based interactive multi-scale and multi-modal transformer for brain tumor detection. This approach generates high-quality fused CT-MRI images using CT and MRI brain scan datasets. Extensive testing evaluated the efficacy of the Trans-IMSM method. Results demonstrated its robustness and effectiveness, achieving 98.64% accuracy and an SSIM of 0.94. These findings highlight the method's superior fusion quality and improved diagnostic performance for brain tumor detection.

Comparing Previous work

Research	**Author**	**Year**	**Dataset**	**Model**	**Metric**
Clinical Text Feature Extraction Model					
Clinical modernbert: An efficient and long-context encoder for biomedical text	Lee, S., Wu, S. and Chiang, J.	2025	PubMed Abstracts, MIMIC-IV Clinical Notes, Structured Medical Ontologies	ModernBERT, BioClinicalBERT	ModernBERT accuracy 63.31%
Do LLMs Surpass Encoders for Biomedical NER?	Obeidat, S., Nahian, M. and Kavuluru, R.	2025	JNLPBA, BioRED, ChemProt, BC5CDR, Reddit-Impacts	Bert, BiomedBERT, DeBERTa-v3	DeBERTa-v3 F1-score of 92.74%.
NLP-based clinical text classification and sentiment	Guleria, P.	2024	Medical transcriptions for various	BERT, LSTM, SVM, CNN	Accuracy: LSTM - 94%.

analyses of complex medical transcripts using transformer model and machine learning classifiers			medical specialties.		
Hybrid BERT-CNN Approach for Medical Text Classification	Abdennour, G., Gasmi, K. and Ejbali, R.	2025	cancer-related documents data.	BERT-CNN, CNN, LSTM	BERT-CNN accuracy 96.66%.
Survice-BERT: A BERT Model for Named Entity Recognition in Infectious Disease Surveillance Reports	Cheon, S. and Ahn, I.	2025	-	Survice-BERT	Achieved a F1-Score of 0.99.
Image Feature Extraction Model					
Brain Tumor Classification and Stage Detection Using Deep Learning	Shanmugarj, G., *et al.*	2025	High-resolution MRI images	VGG16	Ten-fold cross-validation: 99%
Automated Classification of Brain Tumors from Magnetic Resonance Imaging Using Deep Learning	Rasheed, Z. *et al.*	2023	Contranst-enhaced MR images	VGG16, ResNet50, InceptionV3	Accuracy: ResNet50 - 98.04%.
Automated Detection and Classification Method of Brain Tumors Using CNNs and Deep Transfer Learning	Ahmmed, F. *et al.*	2025	Brain tumor MRI dataset	DenseNet169, EfficentNetB3, VGG16	Accuracy: DenseNet169 - 99.58%. EfficientNetB3 - 99.58%.
Brain MRI Classification Using Deep Learning Algorithm	Thinger, C. *et al.*	2025	Brain tumor MRI dataset	CNNs, CNN-SVM hybrids	CNN-SVM accuracy 94.7%.
NeuroNet: Enhanced CNN with FPN and	Limbani, N., Rastogi, S. and D, R.	2025	Brain tumor MRI dataset	NeuroNet	Accuracy: 96.5%, Precision:

Attention Mechanisms for Brain Tumor Classification and Segmentation Using XAI					96.5%, Recall: 95.5%, F1-Score: 96%.
Multimodal Fusion					
Advancing healthcare through multimodal data fusion: a comprehensive review of techniques and applications	Teoh, J. *et al.*	2024	Diagnosis data	Late Fusion	Diagnostic Accuracy: 94.82%
An Explainable Machine Learning Model for Early Detection of Brain Tumors: Integrating Multi-Modal Medical Imaging and Intelligent Feature Fusion	Devi, S. *et al.*	2025	The cancer imaging Archive (TCIA), TCGA-GBM, Br35H,	Feature Extraction	Accuracy - 97.3%, Precision - 96.4%, Recall - 96.0%, F1-score - 96.2%
Deep Learning Models For Automated Tumor Segmentation: Integrating Clinical Notes And Imaging Data With Llms	Jamili, L., Kulkarni, S. and Jain, U.	2025	Public datasets (e.g., BraTS), Curated and deidentified patient records	Attention-based fusion	Precision - 0.82, F1 score - 0.84
Fusion of medical imaging and electronic health records using deep learning: a systematic review and	Huang, S. *et al.*	2020	Glioma Image Database, Independent dataset	Early fusion - concantation method	Accuracy: Fusion: 90.66%, MRI: 81.04%

implementation guidelines					
Deep Learning based Multimodal Image Fusion for Brain Tumor	S, S. *et al.*	2025	MIMIC-CXR	Feature Fusion, Multi-scale Fusion	Accuracy-97.3%
Classification & Diagnose Model					
Explainability of Deep Neural Networks for Brain Tumor Detection	Park, S., Kim, J.	2024	MRI datasets	VGG-16, ResNet-50, EfficientNetV2 L, and ViT-B/1	VGG-16, ResNet-50, achieved the highest accuracy.
TransMed: Transformers Advance Multi-Modal Medical Image Classification	Dai, Y., Gao, Y. and Liu, F.	2021	Evaluated on parotid gland tumor and knee injury classification datasets.	CNN and Transformer.	Achieved 10.1% and 1.9% average accuracy
Multimodal Medical Transformer for Incomplete Multimodal Learning of Brain Tumor Segmentation	Zhang, Y. *et al.*	2022	BraTS 2018 dataset.	CNN and Transformer. Modalities: T1, T1c, T2, and FLAIR	Outperforms modern methods and demonstrates superior robustness
Advancing Brain Tumor Detection Through Trans-IMSM Model	Durairaj, V., Uthirapathy, P.	2024	Used CT and MRI brain scan datasets.	Trans-IMSM	Achieving a 98.64% accuracy and a SSIM of 0.94
Generalizing Dice and cross entropy-based losses to handle class imbalanced medical image segmentation	Yeung, M. *et al.*	2022	KiTS19, BraTS20, CVC-ClinicDB, DRIVE, BUS2017 datasets.	Unified Focal Loss, CNNs.	Dice and cross-entropy-based losses.

Technological Review

This chapter evaluates various brain tumor detection approaches, segmentation, multimodal fusion, and methodologies presented in the literature. The evaluation focuses on predictive performance and standard metrics, robustness and generalization, interpretability and explainability, computational cost and clinical applicability.

Comparison of Methods

Model	Dataset	Performance	Strengths	Limitations
BERT-CNN	Cancer-related documents data	Accuracy of 96.66%.	Contextual understanding	Computational cost
Survice-BERT	-	F1-Score of 99%.	Robustness	Computational cost
DenseNet169, EfficentNetB3	Brain tumor MRI dataset	DenseNet169 accuracy 99.58%. EfficientNetB3 accuracy 99.58%.	Good generalization	High computational cost, complex model
NeuroNet		Accuracy: 96.5%, Precision: 96.5%, Recall :95.5%, F1-Score: 96%.	Good performance	Harder to interpret
Feature Extraction	The cancer imaging Archive (TCIA), TCGA-GBM, Br35H	Accuracy - 97.3%, Precision - 96.4%, Recall - 96.0%, F1-score - 96.2%	Good generalization	Bias to dataset
Feature Fusion, Multi-scale Fusion	MIMIC-CXR	Accuracy- 97.3%	Multi scale learning and better robustness	High model complexity, difficult optimization
EfficientNetV2L / ViT-B/16 (Transformer)	MRI	Slightly lower than CNNs	Explores transformer explainability	Higher computational demand
TransMed	Multi-modal (parotid, knee)	+10.1% / +1.9% improvement over CNN baselines	Strong cross-modality fusion	-
mmFormer	BraTS 2018	Outperformed previous methods	Robust segmentation	High computation

VGG16, ResNet50	High-resolution/ Contrast-enhanced MRI images and MRI scans	Ten-fold cross-validation: 99%, High accuracy (98.04%).	Reliable transfer learning	Gains from augmentation are negligible
Trans-IMSM	CT + MRI	Accuracy - 98.64%, SSIM - 0.94	Robust multimodal detection	Dataset details limited

Summary

This table compares the deep learning models such as CNNs, Transformers, hybrid and multimodal approaches, which is applied to medical imaging and health datasets, which are public datasets. The table gives a summary of the models which has achieved a very high accuracy (95 – 99%), F1 score, precision and recall, which shows a strong potential for medical diagnosis (mostly brain tumor detection) and classification tasks using these set of models, which covers all 4 components, although computational cost and model complexity are challenges to be overcome.

References

Ahammed, F. *et al.* (2025) Automated Detection and Classification Method of brain Tumors Using CNNs and Deep Transfer Learning. Available at: https://doi.org/10.1109/QPAIN66474.2025.11172022 (Accessed: 9 October 2025).

Abdennour, G., Gasmi, K. and Ejbali, R. (2025) *Hybrid BERT-CNN Approach for Medical Text Classification*. Available at: https://doi.org/10.1007/978-3-031-87769-8_24 (Accessed: 9 October 2025).

Cheon, S. and Ahn, I. (2025) *Survice-BERT: A BERT Model for Named Entity Recognition in Infectious Disease Surveillance Reports*. Available at: https://doi.org/10.4167/jbv.2025.55.3.235 (Accessed: 9 October 2025).

Devi, S. *et al.* (2025) An Explainable Machine Learning Model for Early Detection of Brain Tumors: Integrating Multi-Modal Medical Imaging and Intelligent Feature Fusion. Engineering Technology & Applied Science Research, 15(5), pp.26448–26453. Available at: https://doi.org/10.48084/etasr.11539 (Accessed: 9 October 2025).

Durairaj, V. and Uthirapathy, P. (2024) *Interactive Multi-scale Fusion: Advancing Brain Tumor Detection Through Trans-IMSM Model*. Available at: https://doi.org/10.1007/s10278-024-01222-7 (Accessed: 9 October 2025).

Dai, Y., Gao, Y. and Liu, F. (2021) *TransMed: Transformers Advance Multi-Modal Medical Image Classification*. Available at: https://doi.org/10.3390/diagnostics11081384 (Accessed: 9 October 2025).

Guleria, P. (2024) *NLP-based clinical text classification and sentiment analyses of complex medical transcripts using transformer model and machine learning classifiers*. Available at: https://doi.org/10.1007/s00521-024-10482-x (Accessed: 9 October 2025).

Huang, S. *et al.* (2020) Fusion of medical imaging and electronic health records using deep learning: a systematic review and implementation guidelines. npj Digital Medicine, 3(1). Available at: https://doi.org/10.1038/s41746-020-00341-z (Accessed: 9 October 2025).

Jamili, L., Kulkarni, S. and Jain, U. (2025) Deep Learning Models For Automated Tumor Segmentation: Integrating Clinical Notes And Imaging Data With Llms. 13(2), pp. 2320–2882. Available at: https://www.ijcrt.org/papers/IJCRT2502990.pdf (Accessed: 9 October 2025).

Limbani, N., Rastogi, S. and D, R. (2025) NeuroNet: Enhanced CNN with FPN and Attention Mechanisms for Brain Tumor Classification and Segmentation Using Explainable AI (XAI). Available at: https://doi.org/10.1109/ICBMESH66209.2025.11182182 (Accessed: 9 October 2025).

Lee, S., Wu, S. and Chiang, J. (2025) Clinical ModernBERT: An efficient and long context encoder for biomedical text. Available at: https://doi.org/10.48550/arXiv.2504.03964 (Accessed: 9 October 2025)

Obeidat, S., Nahian, M. and Kavuluru, R. (2025) Do LLMs Surpass Encoders for Biomedical NER? Available at: https://doi.org/10.48550/arXiv.2504.00664 (Accessed: 9 October 2025).

Park, S. and Kim, J. (2024) *Explainability of Deep Neural Networks for Brain Tumor Detection*. Available at: https://doi.org/10.48550/arXiv.2410.07613 (Accessed: 9 October 2025).

Rasheed, Z. *et al.* (2023) Automated Classification of Brain Tumors from Magnetic Resonance Imaging Using Deep Learning. Available at: https://doi.org/10.1109/QPAIN66474.2025.11172022 (Accessed: 9 October 2025).

Shanmugaraj, G. *et al.* (2024) Brain Tumor Classification and Stage Detection Using Deep Learning. Available at: https://doi.org/10.1109/ICDICI66477.2025.11135002 (Accessed: 9 October 2025).

S, S. *et al.* (2025) Deep Learning based Multimodal Image Fusion for Brain Tumor Diagnosis: A Comprehensive Review. Available at: https://doi.org/10.5281/zenodo.15301947 (Accessed: 9 October 2025).

Thinger, C. *et al.* (2025) Brain MRI Classification Using Deep Learning Algorithm. Available at: https://doi.org/10.1109/ICCSAI64074.2025.11064631 (Accessed: 9 October 2025).

Teoh, J. *et al.* (2024) Advancing healthcare through multimodal data fusion: a comprehensive review of techniques and applications. Available at: https://doi.org/10.7717/peerj-cs.2298 (Accessed: 9 October 2025).

Yeung, M. *et al.* (2022) *Unified Focal loss: Generalizing Dice and cross entropy-based losses to handle class imbalanced medical image segmentation*. Available at: https://doi.org/10.1016/j.compmedimag.2021.102026 (Accessed: 9 October 2025).

Zhang, Y. *et al.* (2022) *mmFormer: Multimodal Medical Transformer for Incomplete Multimodal Learning of Brain Tumor Segmentation*. Available at: https://doi.org/10.48550/arXiv.2206.02425 (Accessed: 9 October 2025).