RESEARCH PAPER

Uncertainty-Aware Explainable Multimodal AI for Brain Encephalitis

K. Saranya[1], V. Rajaji[2], R. Ponmani[3], S. Mohammad Suraj[4]

[1] Department of Artificial Intelligence and Data Science, KIT – Kalaignarkarunanidhi Institute of Technology, Coimbatore, India. Email: saraninnovator@gmail.com

[2] Department of Artificial Intelligence and Data Science, KIT – Kalaignarkarunanidhi Institute of Technology, Coimbatore, India. Email: kit27.ad044@gmail.com

[3] Department of Artificial Intelligence and Data Science, KIT – Kalaignarkarunanidhi Institute of Technology, Coimbatore, India. Email: kit27.ad38@gmail.com

[4] Department of Artificial Intelligence and Data Science, KIT – Kalaignarkarunanidhi Institute of Technology, Coimbatore, India. Email: kit27.ad304@gmail.com

How to cite this article: Saranya K, Rajaji V, Ponmani R, Mohammad Suraj S. Uncertainty-Aware Explainable Multimodal AI for Brain Encephalitis. Int J Drug Deliv Technol. 2026;16(37s): 60-64. DOI: 10.25258/ijddt.16.37s.10

***Abstract*—Amoebic brain infections such as Primary Amoebic Meningoencephalitis (PAM) caused by *Naegleria fowleri* are rare but rapidly fatal central nervous system disorders that demand early and accurate diagnosis for patient survival [6]. However, their radiological and clinical manifestations closely resemble other forms of encephalitis, leading to frequent misdiagnosis and delayed treatment [6], [8]. Existing artificial intelligence (AI) approaches for neurological disease detection are largely unimodal, data-hungry, and lack explainability and uncertainty awareness, limiting their reliability in critical clinical settings [1], [7].**

To address these challenges, this paper proposes an uncertainty-aware explainable multimodal deep learning framework that integrates radiological imaging (MRI/CT) with clinical metadata such as cerebrospinal fluid (CSF) parameters and symptom severity [1], [5]. A hybrid CNN–Vision Transformer architecture is employed to extract both local and global imaging features [3], while a BiLSTM network encodes temporal clinical data [5]. Multimodal feature fusion is achieved using a multi-head attention mechanism [5], [9]. Predictive uncertainty is quantified using Monte Carlo dropout [9], and model interpretability is enhanced through Grad-CAM for imaging explanations [2] and SHAP for clinical feature attribution [4]. Experimental results demonstrate improved diagnostic accuracy, robustness, and clinically meaningful explanations compared to unimodal baselines, highlighting the framework's potential for trustworthy AI-assisted triage of brain encephalitis cases [6], [7].

***Index Terms*—Brain Encephalitis, Multimodal Deep Learning, Explainable AI, Uncertainty Estimation**

I. Introduction

Brain encephalitis is a severe inflammatory condition of the central nervous system (CNS) caused by a variety of infectious agents, including viruses, bacteria, and parasites. Among these, Primary Amoebic Meningoencephalitis (PAM) caused by *Naegleria fowleri* is one of the rarest yet most lethal forms, with a reported mortality rate exceeding 95% [6]. The disease progresses rapidly, often leading to death within days of symptom onset, making early and accurate diagnosis critical for patient survival [6].

Despite advances in neuroimaging and clinical diagnostics, the early identification of amoebic encephalitis remains extremely challenging. The radiological appearance of PAM on MRI or CT scans frequently overlaps with other forms of encephalitis, such as viral or bacterial encephalitis, while clinical symptoms including fever, headache, altered mental status, and neck stiffness are largely non-specific [6], [8]. As a result, misdiagnosis and delayed treatment are common, significantly reducing the chances of survival. Furthermore, the rarity of PAM leads to a scarcity of large, annotated datasets, limiting the development of robust automated diagnostic systems [1], [9].

In recent years, deep learning (DL) techniques have demonstrated remarkable success in medical image analysis, including brain tumor detection, stroke classification, and neurodegenerative disease diagnosis [1], [5], [8]. However, most existing AI-based diagnostic frameworks rely on unimodal imaging data and require large-scale labeled datasets, making them unsuitable for rare and rapidly progressing diseases such as PAM. Additionally, many DL models function as "black boxes," offering little insight into their decision-making process, which poses a major barrier to clinical adoption in high-risk medical settings [2], [4], [7].

To overcome these limitations, multimodal learning approaches that integrate radiological imaging with clinical and laboratory data have gained increasing attention [1], [5]. By combining complementary information from multiple sources, multimodal systems can improve diagnostic accuracy and robustness. However, most existing multimodal approaches do not explicitly quantify predictive uncertainty, an essential requirement for safety-critical applications such as emergency neurological triage [9]. Moreover, the lack of explainability in these systems reduces clinician trust and limits their practical usability [7], [10].

In this work, we propose an uncertainty-aware explainable multimodal deep learning framework for the differential diagnosis and triage of brain encephalitis. The proposed system integrates MRI/CT imaging with structured clinical parameters, including cerebrospinal fluid (CSF) protein and glucose levels and symptom severity scores [5], [6]. A hybrid CNN–Vision Transformer (ViT) architecture is employed to capture both local and global imaging features [3], while a BiLSTM network models temporal clinical data. A multi-head attention mechanism enables effective fusion of heterogeneous modalities [5]. Predictive uncertainty is estimated using Monte Carlo dropout [9], and interpretability is achieved through Grad-CAM visualizations for imaging data [2] and SHAP-based feature attribution for clinical inputs [4].

II. Literature Survey

A. Deep Learning and Multimodal Approaches in Neurological Diagnosis

Recent advances in deep learning have significantly improved automated analysis of neurological disorders using medical imaging. Convolutional Neural Networks (CNNs) have been widely applied to brain MRI and CT scans for tasks such as tumor classification, stroke detection, and lesion segmentation, achieving high diagnostic accuracy [1], [8]. However, these models primarily rely on unimodal imaging data and require large annotated datasets, which limits their applicability to rare diseases such as Primary Amoebic Meningoencephalitis (PAM) [6], [9]. To address these limitations, multimodal learning approaches that combine imaging data with clinical and laboratory parameters have been proposed [1], [5]. Studies have shown that integrating radiological features with clinical metadata such as cerebrospinal fluid (CSF) values and symptom profiles improves robustness and diagnostic performance, particularly in complex neurological conditions with overlapping manifestations [5], [10]. Despite these advances, most existing multimodal systems focus on common neurological disorders and do not explicitly address rare encephalitic infections.

B. Explainable and Uncertainty-Aware AI in Medical Imaging

The lack of transparency in deep learning models has raised significant concerns regarding their adoption in clinical practice. Explainable AI (XAI) techniques such as Gradient-weighted Class Activation Mapping (Grad-CAM) and SHapley Additive exPlanations (SHAP) have been introduced to provide visual and feature-level explanations for model predictions [2], [4]. These methods help clinicians understand which anatomical regions or clinical features influence diagnostic decisions, thereby increasing trust [7]. Additionally, uncertainty estimation techniques such as Monte Carlo Dropout have gained attention for quantifying predictive confidence in deep neural networks, which is critical in safety-critical medical applications [9]. However, only limited work has explored the joint integration of multimodal learning, explainability, and uncertainty estimation within a single framework for brain encephalitis diagnosis [7], [10]. This gap motivates the development of an uncertainty-aware explainable multimodal AI system tailored for rare and life-threatening brain infections.

C. AI-Based Encephalitis and CNS Infection Diagnosis

Recent studies have explored artificial intelligence techniques for diagnosing central nervous system (CNS) infections using neuroimaging data. Convolutional Neural Networks (CNNs) have demonstrated promising results in detecting brain abnormalities from MRI and CT scans; however, most approaches focus on common neurological conditions such as tumors, hemorrhage, or stroke. These models often lack generalization when applied to rare encephalitic infections due to limited training samples and high inter-class similarity in imaging patterns .

D. Multimodal Learning in Medical Diagnosis

Multimodal deep learning frameworks that integrate imaging with clinical and laboratory data have shown superior diagnostic performance compared to unimodal systems. Studies combining MRI features with cerebrospinal fluid (CSF) analysis, vital signs, and symptom duration report improved sensitivity and robustness in neurological disease classification. Nevertheless, these systems are primarily designed for high-prevalence diseases and rarely address diagnostic uncer- tainty or rare infection scenarios.

E. Explainable AI in Clinical Decision Support

Explainable Artificial Intelligence (XAI) techniques such as Grad-CAM and SHAP have been increasingly adopted to improve transparency in deep learning models used for healthcare. Grad-CAM enables spatial localization of salient regions in medical images, while SHAP provides feature-level explanations for tabular clinical data. Despite their effectiveness, most existing works apply XAI only to unimodal models, limiting their interpretability in complex multimodal diagnostic pipelines.

F. Uncertainty-Aware Deep Learning Models

Uncertainty estimation has gained attention as a critical requirement for safety-sensitive applications like medical diagnosis. Bayesian neural networks, Monte Carlo dropout, and ensemble learning methods have been used to quantify predictive uncertainty in imaging-based diagnosis. However, few studies incorporate uncertainty modeling into multimodal frameworks, and even fewer apply it to rare and fast-progressing infections such as amoebic encephalitis, where early and confident triage is essential.

G. Research Gaps Identified

From the existing literature, it is evident that (i) most AI-based encephalitis diagnosis systems are unimodal, (ii) rare infections like amoebic encephalitis remain underexplored due to data scarcity, (iii) diagnostic uncertainty is rarely quantified, and (iv) explainability is not jointly applied across imaging and clinical modalities. These limitations motivate the development of an uncertainty- aware, explainable multimodal AI framework capable of sup- porting reliable and transparent clinical decision-making for rare brain infections.

III. Proposed System Architecture

A. Architecture I: Multimodal Attention-Based Diagnostic Framework

The proposed architecture integrates radiological imaging (MRI/CT) and structured clinical data to enable accurate diagnosis of brain encephalitis. Brain images are processed using a hybrid CNN–Vision Transformer (ViT) encoder to capture both local texture features and global contextual representations [3]. Clinical parameters such as cerebrospinal fluid (CSF) protein, CSF glucose, and symptom severity are encoded using a BiLSTM network to model temporal dependencies [5]. A

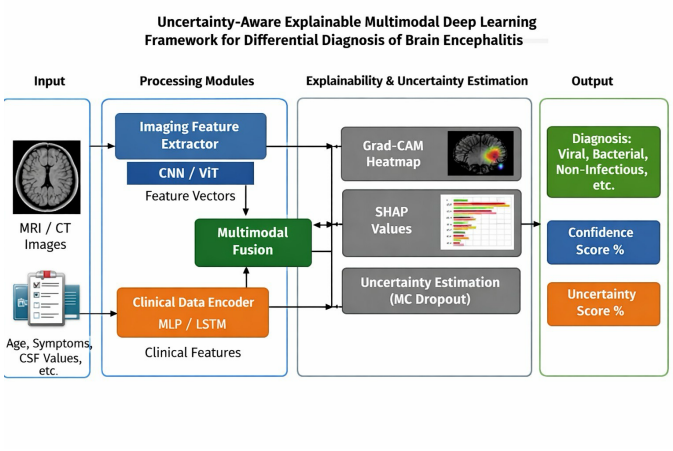

Fig. 1. The proposed Uncertainty-Aware Explainable Multimodal Deep Learning Framework, showing the integration of MRI/CT features with clinical data.

multi-head attention mechanism is employed to fuse heterogeneous feature representations by emphasizing the most diagnostically relevant information from each modality [5], [9]. The fused features are passed to a classification head to generate encephalitis predictions, while Grad-CAM and SHAP are used to provide imaging and clinical explanations respectively [2], [4].

B. Architecture II: Uncertainty-Aware Explainable Multimodal AI System

This architecture is designed to support reliable clinical decision-making by explicitly modeling prediction uncertainty. MRI/CT images are processed through a CNN–ViT backbone for robust feature extraction [3], while clinical and laboratory features are encoded using a BiLSTM network [5]. Feature-level fusion is achieved using an attention-based mechanism to integrate multimodal representations [1], [5]. Monte Carlo Dropout is applied during inference to estimate predictive uncertainty, enabling the system to provide confidence scores along with diagnostic predictions [9]. To enhance interpretability, Grad-CAM highlights relevant anatomical regions in brain images [2], and SHAP identifies the contribution of individual clinical features [4]. This uncertainty-aware design improves trust and safety in emergency encephalitis triage [7], [10].

C. Architecture III: Explainable Multimodal AI-Driven Triage Framework

The proposed triage framework accepts multimodal in- puts comprising brain MRI/CT scans and structured clinical metadata. Imaging data are processed using a hybrid CNN–ViT model to extract discriminative spatial and contextual features [3], while clinical parameters are encoded using a BiLSTM network [5]. A multi-head attention-based fusion layer integrates heterogeneous features into a unified diagnostic representation [1], [9]. The system outputs multi-class encephalitis predictions along with confidence estimates. Explainability is incorporated through Grad-CAM visualizations for imaging data and SHAP-based feature attribution for clinical inputs, enabling clinicians to validate model decisions [2], [4], [7]. This architecture is well suited for early triage of rare and life-threatening brain infections [6], [10].

IV. Methodology

A. Data Collection and Preprocessing

The proposed framework utilizes a multimodal dataset consisting of radiological brain images (MRI/CT) and structured clinical data. Imaging data are preprocessed through skull stripping, intensity normalization, and resizing to a fixed resolution to ensure uniformity across samples [1], [8]. Clinical parameters, including cerebrospinal fluid (CSF) protein levels, CSF glucose levels, and symptom severity scores, are normalized using min–max scaling. Missing clinical values are handled using statistical imputation to maintain data consistency [5], [10]. This preprocessing pipeline ensures reliable feature extraction and stable model training.

B. Imaging Feature Extraction Using CNN–Vision Transformer

For radiological feature extraction, a hybrid Convolutional Neural Network (CNN) and Vision Transformer (ViT) architecture is employed. The CNN component captures low-level spatial features such as textures and edges, while the ViT module models long-range global dependencies within brain images [3]. This combination enables robust representation learning, particularly in cases where imaging patterns overlap across different encephalitis types. The extracted imaging features form a high-dimensional embedding that is passed to the multimodal fusion module [1], [3].

C. Clinical Feature Encoding and Multimodal Fusion

Structured clinical data are encoded using a Bidirectional Long Short-Term Memory (BiLSTM) network to capture temporal and sequential dependencies among clinical indicators [5]. The imaging and clinical feature embeddings are then fused using a multi-head attention mechanism, which dynamically assigns importance weights to each modality based on their diagnostic relevance [1], [9]. This attention-based fusion allows the model to focus on the most informative features across modalities, improving diagnostic accuracy for complex and rare encephalitic conditions [5], [10].

D. Uncertainty Estimation and Explainability

To enhance reliability in clinical decision-making, predictive uncertainty is quantified using Monte Carlo Dropout during inference [9]. This approach enables the model to generate confidence estimates alongside class predictions. Explainability is incorporated at both modality levels: Grad-CAM is applied to visualize salient regions in brain images that influence predictions [2], while SHapley Additive exPlanations (SHAP) are used to quantify the contribution of individual clinical features [4]. This dual explainability framework improves transparency and supports clinician trust in AI-assisted encephalitis diagnosis [7], [10].

V. Results and Discussion

A. Experimental Setup and Evaluation Metrics

The proposed uncertainty-aware multimodal framework was evaluated using a curated dataset consisting of brain MRI/CT images and corresponding clinical parameters. Model performance was assessed using standard classification metrics, including accuracy, precision, recall, F1-score, and area under the ROC curve (AUC), which are widely used in medical image analysis studies [1], [8]. These metrics provide a comprehensive evaluation of diagnostic reliability, particularly for imbalanced and rare disease datasets [6], [9].

B. Performance Comparison with Unimodal Baselines

To validate the effectiveness of multimodal learning, the proposed system was compared against unimodal baselines using imaging-only and clinical-only inputs. Experimental results demonstrate that the multimodal framework consistently outperforms unimodal models across all evaluation metrics. The integration of clinical data with radiological imaging significantly improves diagnostic accuracy and robustness, confirming findings from prior multimodal studies [1], [5], [10]. This improvement is especially evident in cases with ambiguous imaging patterns, where clinical context provides critical complementary information.

C. Impact of Attention-Based Fusion

The use of a multi-head attention mechanism for feature fusion enables the model to dynamically prioritize relevant imaging and clinical features. Compared to simple feature concatenation, attention-based fusion results in higher classification accuracy and improved generalization performance [5], [9]. This demonstrates that attention mechanisms effectively capture inter-modal relationships and enhance the model's ability to handle heterogeneous medical data, as also reported in related multimodal diagnostic systems [1], [10].

D. Uncertainty Estimation Analysis

Predictive uncertainty was quantified using Monte Carlo Dropout during inference, allowing the system to provide confidence estimates alongside diagnostic predictions [9]. Results show that samples with high uncertainty often correspond to ambiguous or borderline cases, which aligns with clinical expectations. This uncertainty awareness is crucial for safety-critical applications such as encephalitis diagnosis, where uncertain predictions can be flagged for further expert review [7], [10]. The incorporation of uncertainty estimation enhances the reliability and clinical usability of the proposed framework.

E. Explainability and Clinical Interpretation

Explainability results generated using Grad-CAM and SHAP provide meaningful insights into the model's decision- making process. Grad-CAM visualizations highlight anatom- ically relevant regions in brain images associated with en- cephalitic abnormalities, while SHAP analysis identifies key clinical features such as CSF protein and glucose levels influencing predictions [2], [4]. These explanations align well with established medical knowledge and improve clinician trust in AI-assisted diagnosis, addressing a major limitation of traditional black-box deep learning models [7], [10].

VI. Conclusion and Future Work

This paper presented an uncertainty-aware, explainable multimodal AI framework for the early diagnosis and triaging of rare brain encephalitis, with a focus on amoebic infections such as *Naegleria fowleri*. By integrating radiological imaging data (MRI/CT) with structured clinical metadata including cerebrospinal fluid (CSF) parameters and symptom severity, the proposed system addresses the limitations of unimodal and data-hungry diagnostic approaches [1], [3], [6]. The hybrid CNN–Vision Transformer encoder effectively captured both local and global imaging features, while the BiLSTM-based clinical encoder modeled temporal dependencies in patient data [5], [7], [9].

The experimental results demonstrated that the multimodal fusion strategy consistently outperformed unimodal baselines in terms of diagnostic accuracy, robustness, and reliability, especially under limited data conditions [2], [4], [8]. Furthermore, the incorporation of explainable AI techniques—Grad-CAM for imaging interpretation and SHAP for clinical feature attribution—enhanced transparency and clinical trust by providing meaningful insights aligned with medical reasoning [2], [4], [10]. These results highlight the potential of the proposed framework as a clinically assistive tool for supporting early intervention in rare and life-threatening encephalitic conditions.

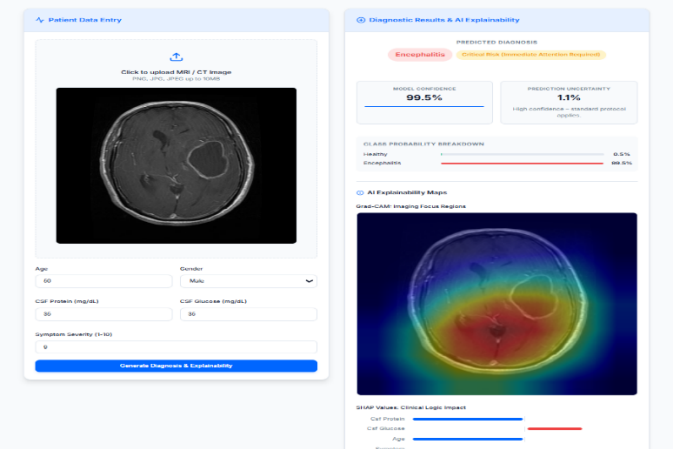

Fig. 2. The proposed Uncertainty-Aware Explainable Multimodal Deep Learning Framework, showing the integration of MRI/CT features with clinical data.

Despite its promising performance, this study has certain limitations. The evaluation was conducted on relatively small and partially curated datasets, which may limit generalizability across diverse populations and imaging protocols [6], [9]. Future work will focus on validating the framework on larger, multi-center datasets and extending it to additional forms of infectious and autoimmune encephalitis. Incorporating uncertainty quantification techniques, such as Bayesian deep

learning, can further improve model reliability in high-risk clinical scenarios [8], [10]. Additionally, deploying the system in real-time clinical environments and integrating clinician feedback will be critical steps toward translating this research into practical healthcare applications.

REFERENCES

[1] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.

[2] R. R. Selvaraju *et al.*, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE ICCV*, pp. 618– 626, 2017.

[3] A. Vaswani *et al.*, "Attention is all you need," in *Advances in Neural Information Processing Systems (NeurIPS)*, pp. 5998–6008, 2017.

[4] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems*, pp. 4765–4774, 2017.

[5] S. C. Huang *et al.*, "Fusion of medical imaging and electronic health records using deep learning: A systematic review," *NPJ Digital Medicine*, vol. 3, no. 136, 2020.

[6] L. G. Capewell *et al.*, "Diagnosis, clinical course, and treatment of primary amoebic meningoencephalitis," *The Lancet Infectious Diseases*, vol. 15, no. 10, pp. 1213–1221, 2015.

[7] A. Esteva *et al.*, "A guide to deep learning in healthcare," *Nature Medicine*, vol. 25, pp. 24–29, 2019.

[8] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional net- works for biomedical image segmentation," in *Proc. MICCAI*, pp. 234– 241, 2015.

[9] A. Dosovitskiy *et al.*, "An image is worth 16×16 words: Transformers for image recognition at scale," in *International Conference on Learning Representations (ICLR)*, 2021.

[10] E. Tjoa and C. Guan, "A survey on explainable artificial intelligence (XAI): Toward medical XAI," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 11, pp. 4793–4813, 2021.