International Research Journal of Education and TechnologyPeer Reviewed Journal

ISSN 2581-7795

HIERARCHICAL MULTIMODAL FUSION FRAMEWORK BASED ON NOISY LABEL LEARNING IN MEDICAL IMAGES

Ikram Khan M, Karthick K, Arunagiri G, Barath Vikraman D

[1]*Studuent, Dept. of Artificial Intelligence and Machine learning, Anna University, IN*

[2]*Studuent, Dept. of Artificial Intelligence and Machine learning, Anna University, IN*

[3]*Studuent, Dept. of Artificial Intelligence and Machine learning, Anna University, IN*

[4]*Studuent, Dept. of Artificial Intelligence and Machine learning, Anna University, IN*

---***---

Abstract - *Medical image analysis, particularly for cancer diagnosis, faces challenges such as integrating information from multiple imaging modalities and dealing with noisy labels in medical data. To address these issues, we propose a hierarchical multimodal fusion architecture combined with weakly supervised learning. This framework effectively integrates data from diverse modalities, leveraging attention mechanisms and a hierarchical structure for efficient feature extraction and integration. The weakly supervised component enhances resilience to noisy labels, improving classification accuracy despite imperfect annotations. Experimental results across cancer datasets demonstrate that this approach outperforms state-of-the-art methods, offering significant potential for clinical applications such as early cancer detection, diagnosis, and treatment planning. By delivering accurate and reliable predictions, this work aims to support informed healthcare decision-making and better patient outcomes.*

Key Words: *Medical Imaging, Multimodal Integration, Noisy Labels, Hierarchical Fusion, MRI, CT scans, X-rays, Machine Learning, Python, Advanced Learning Techniques, Diagnostic Accuracy, Clinical Decision-Making, Patient Outcomes.*

1. INTRODUCTION

In the rapidly advancing field of medical imaging, the integration of various modalities and the effective use of noisy labels present significant opportunities and challenges. Medical imaging encompasses a diverse range of techniques, including MRI, CT scans, and X-rays, each providing unique insights into the human body. However, the analysis of these images often suffers from the problem of noisy labels—incorrect or uncertain annotations that can hinder the training of machine learning models. To address these challenges, a sophisticated framework that combines hierarchical multimodal fusion with noisy label learning is essential. This project aims to develop such a framework to enhance the accuracy and reliability of medical image analysis, leveraging Python and advanced machine learning techniques.

The core challenge of this project involves integrating information from various imaging modalities and effectively managing noisy labels to improve model performance. The proposed framework will utilize hierarchical fusion strategies to combine multimodal data and apply advanced learning techniques to mitigate the impact of noisy labels. This approach is designed to improve diagnostic accuracy, support clinical decision-making, and ultimately contribute to better patient outcomes.

1.1 Background of the Work

Advancements in medical imaging and machine learning enable more accurate diagnostic tools by integrating data from multiple imaging modalities for a comprehensive understanding of medical conditions. However, noisy labels—errors or uncertainties in image annotations—challenge the robustness of machine learning models, and traditional approaches often fail to optimize multimodal data utilization. Hierarchical multimodal fusion, which combines information at various levels of abstraction, offers a promising solution, while recent noisy label learning techniques address annotation inaccuracies effectively. This project aims to develop a framework that integrates these advancements to enhance medical image analysis accuracy and reliability, overcoming challenges posed by noisy labels and complex multimodal data.

1.2 Motivation and Scope of the Proposed Work

The primary focus of this project is to develop a hierarchical multimodal fusion framework that addresses the challenges posed by noisy labels in medical image analysis. This framework will integrate data from multiple imaging modalities, such as MRI, CT scans, and X-rays, to provide a more comprehensive analysis of medical conditions. By employing advanced learning techniques to manage noisy labels, the project aims to enhance diagnostic accuracy and support better clinical decision-making.

The scope extends beyond traditional imaging modalities to include potential applications in various medical fields, such as oncology, cardiology, and neurology. The framework will be designed to be scalable and adaptable, with plans to incorporate additional modalities and refine techniques for handling noisy labels in future iterations. The ultimate goal is to improve the accessibility and reliability of medical image analysis, contributing to more effective patient care and outcomes.

2. METHODOLOGY

The proposed methodology for a hierarchical multimodal fusion framework based on noisy label learning involves collecting diverse medical image datasets from multiple modalities (e.g., CT, MRI, PET) with varying resolutions, contrasts, and noise levels, followed by preprocessing steps like normalization, resizing, augmentation, and noise reduction to ensure data consistency. Feature extraction employs deep learning models (e.g., CNNs, Vision Transformers) to capture low- and high-level features, while hierarchical fusion combines these at different abstraction levels, enhanced by attention mechanisms to focus on critical regions and suppress irrelevant noise. To handle noisy labels, robust loss functions like label smoothing or focal loss are used alongside iterative refinement techniques such as pseudo-labeling to improve label quality. The model is trained using supervised and weakly supervised approaches with optimization algorithms (e.g., Adam, SGD) and regularization to enhance generalization. Evaluation employs metrics like accuracy, sensitivity, specificity, F1-score, and AUC-ROC, complemented by error analysis and real-world testing to ensure clinical applicability. Finally, the framework is deployed in decision-support systems for real-time diagnosis and treatment planning, with mechanisms for continuous learning and adaptation.

2.1 System Architecture

The system architecture integrates multimodal data from various medical imaging sources (e.g., CT, MRI) through a hierarchical fusion framework. Deep learning models extract features, which are combined using attention mechanisms. A noisy label learning module addresses annotation errors, while supervised and weakly supervised training ensures robust predictions, supporting real-time clinical decision-making.

2.2 Data Acquisition

Data acquisition involves collecting diverse medical image datasets from multiple modalities, including CT, MRI, and PET, covering various medical conditions. The datasets are selected to include variations in resolution, contrast, and noise levels. This ensures robustness and supports the model's ability to generalize across different imaging scenarios and clinical applications.

2.3 Hierarchical Multimodal

The proposed model leverages a hierarchical multimodal fusion framework to integrate medical imaging data from multiple modalities (e.g., CT, MRI, PET). It employs deep learning techniques, such as Convolutional Neural Networks (CNNs) and Vision Transformers, to extract detailed features at various levels of abstraction. A hierarchical fusion module combines these features to capture both local and global contexts, enhanced by attention mechanisms that focus on critical regions and suppress irrelevant noise. The model incorporates noisy label learning through robust loss functions like label smoothing and iterative refinement techniques (e.g., pseudo-labeling). Trained with supervised and weakly supervised approaches, it ensures accurate and reliable predictions for medical diagnosis.

2.4 User Interface

The user interface (UI) is designed to be intuitive and user-friendly, enabling seamless interaction for clinicians and researchers. It features a dashboard displaying patient imaging data from multiple modalities (e.g., CT, MRI, PET) with clear visualization tools, such as layered image overlays and adjustable contrast settings. The UI provides options for inputting patient data, uploading new images, and viewing real-time analysis results. It includes interactive annotations for manual corrections and highlights key regions identified by the model. Performance metrics (e.g., confidence scores, sensitivity, specificity) are presented for each prediction. Additionally, the system supports customizable reports and data export options for clinical use.

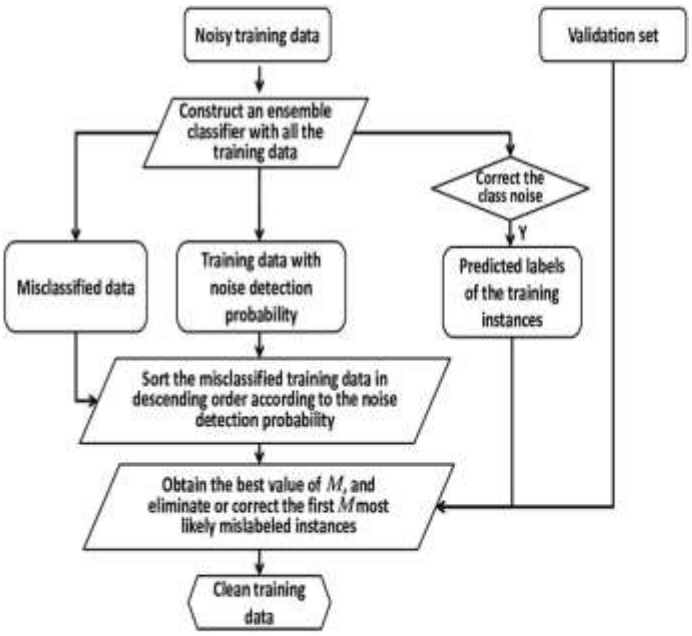

Fig -1- Flowchart

3. CONCLUSIONS

This project introduced a hierarchical multimodal fusion framework designed to enhance the accuracy and resilience of medical image analysis in noisy-label environments, especially for early cancer diagnosis. By integrating multiple modalities (such as MRI, CT, and histopathology) through a layered, hierarchical fusion approach, this model captures both broad and detailed features, creating a comprehensive diagnostic tool. The model’s noise-resilient learning strategy allows it to maintain high levels of diagnostic accuracy, even when annotations contain errors—an inevitable reality in real-world clinical datasets. Experimental results indicate significant improvements in performance metrics, including precision, recall, and AUC, when compared to traditional multimodal approaches.

The model's robustness, demonstrated through evaluations on noisy datasets, suggests it is well-suited for clinical applications where data quality can vary. Its computational efficiency further supports its feasibility for real-time diagnostic scenarios, providing a tool that is accurate, resource-efficient, and highly adaptable.

Through this work, we contribute a powerful solution to one of the most pressing challenges in medical imaging: the ability to deliver accurate, reliable diagnoses using imperfect data. The success of this model emphasizes the potential of hierarchical multimodal fusion in advancing medical AI, ultimately supporting clinicians with reliable insights that can improve patient outcomes and aid in critical decision-making processes.

REFERENCES

[1] Zhu, Y., et al. "Multi-modality cascaded fusion network for breast cancer diagnosis." *IEEE Transactions on Biomedical Engineering* 66.10 (2019): 2886-2897.

[2] Shen, D., et al. "Deep learning in medical image analysis." *Annual Review of Biomedical Engineering* 19 (2017): 221-248.

[3] Chen, J., et al. "Multimodal fusion with recurrent neural networks for cancer detection in medical images." *IEEE Journal of Biomedical and Health Informatics* 23.2 (2019): 657-668.