UNIVERSITÄT BONN

University of Bonn

Master’s thesis for obtaining the academic degree
„Master of Science (M.Sc.)“

Detection of Focal Cortical Dysplasia Type II Using Text Descriptions

Author:
Mikhelson German

First Examiner:
Prof. Dr. Very Smart

Second Examiner:
Prof. Dr. Also Smart

Advisor:
Dr. Lange Annalena

Submitted: December 20, 2025

Declaration of Authorship

I declare that the work presented here is original and the result of my own investigations. Formulations and ideas taken from other sources are cited as such. It has not been submitted, either in part or whole, for a degree at this or any other university.

Location, Date

Signature

Abstract

Numerous methods have been developed for the detection of tumors in internal organs such as the lungs, brain, kidneys, and breast [1–4]. However, detecting epileptogenic lesions remains significantly more challenging. Unlike tumors, these lesions do not typically increase in size over time, and there is a severe shortage of publicly available annotated datasets. Researchers often need to contact hospitals and clinical centers directly to obtain even a minimal number of scans. Although recent studies have demonstrated that deep learning models can detect epileptogenic lesions [5], their performance in terms of Intersection over Union (IoU) rarely exceeds 35%, highlighting the ongoing difficulty of this task.

At the same time, recent work on tumor detection using text-guided approaches has shown promising results [6–8]. Notably, [9] demonstrated that training on only 10% of the dataset can yield performance comparable to using the full dataset. Inspired by these advances, this study proposes a new method that combines visual and textual features to improve Focal Cortical Dysplasia detection under limited data conditions. The proposed approach further demonstrates that incorporating textual descriptions significantly enhances segmentation accuracy. We present a comparison of multiple types of textual annotations and analyze their influence on model performance.

Contents

List of Acronyms

FCD Focal Cortical Dysplasia

IoU Intersection over Union

1 Introduction

2 Related Works

This section reviews existing approaches to focal cortical dysplasia (FCD) detection and text-guided medical image segmentation for tumor detection tasks, focusing on their architecture, fusion strategies, and limitations.

Early work on FCD detection focused exclusively on vision-based methods [references]. It is noteworthy that one of the newest models, the MELD Graph [link], represents the surface of the cerebral cortex as a graph with multiple resolutions and applies a GNN-UNet to segment lesions. It provides high sensitivity and specificity by identifying characteristic peaks (more than 20% in saliency) and determining calibration reliability using the expected calibration error. However, it has not been tested on patients with multiple FCDs, it lacks cross-attentional mechanisms for more complete integration of features, and it does not integrate textual clinical information or evaluate zero-shot generalization across FCD subtypes.

A second line of research explores language-guided segmentation by embedding text semantics into the segmentation pipeline. Early methods (Tomar et al., Li et al.) simply tokenized text and merged embeddings with image features via attention, but struggled to capture high-level semantics. More recent approaches (Lee et al., Zhong et al.) employ deep pretrained text encoders (e.g., CXR-BERT) fused with ConvNeXt-Tiny image features. The Target-sensitive Semantic Distance Module (TSDM) computes contrastive distances between segmentation masks to focus on disease-related regions, while the Language-guided Target Enhancement Module (LTEM) uses cross-attention to reinforce critical image areas. Bi-directional contrastive loss (averaged cross-entropy in both image→text and text→image directions) yields more fine-grained guidance and enables models trained on 25% of data to outperform single-modal baselines trained on full datasets.

To address unpaired multi-modal data, MulModSeg [link] conditions text embeddings on imaging modality using frozen CLIP/BioBERT/Med-CLIP encoders combined with medical prompts, and alternates training between vision and text branches (3D-UNet or SwinUNETR). This scheme improves generalization without requiring paired CT/MR scans. Experiments show that varying the CT:MR ratio in training data shifts performance, underscoring the importance of balanced modality representation. However, effectiveness depends critically on prompt template design and alternating-training convergence.

Organ-aware multi-scale segmentation models (OMT-SAM) extend text prompting to specify organs or tissues, overcoming MedSAM's reliance on geometric prompts. Pretrained CLIP encoders produce paired image/text features at multiple ViT layers; a cross-attention fusion yields rich prompt embeddings. Since small targets are underrepresented by BCE loss alone, combining Dice loss and BCE loss better captures fine structures. OMT-SAM reports improvements in DSC, NSD, and HD95 for organ segmentation, but requires sophisticated text prompt engineering and still depends on ViT performance on small lesions.

Finally, weakly-supervised methods such as SimTxtSeg leverage simple text cues—e.g., "lesion in left hemisphere"—to guide segmentation without pixel-level labels. Using cross-entropy, DSC, IoU, PPV, NSD, and HD95 metrics, these pipelines demonstrate that even coarse textual hints can significantly improve mask quality over vision-only baselines. Yet, they often assume the availability of consistent, high-quality text labels and have not been validated on rare pathologies like FCD.

Building on these insights, we draw inspiration from Ariadne's Thread [link], which uses simple text prompts and a lightweight GuideDecoder to segment infected regions in chest X-rays. Remarkably, this approach achieves over a 6% Dice improvement compared to unimodal baselines while using only 10% of the training set, highlighting the power of multimodal prompting in low-data regimes. For our FCD task, we adopt the MELD Graph model [link] as a pretrained backbone, since it was trained on the largest dataset among the methods surveyed, providing a robust basis for feature extraction and downstream adaptation.

Importantly, to the best of our knowledge, no prior study has applied text-guided segmentation methods to the focal cortical dysplasia detection task, underscoring the novelty of our approach. In summary, while graph-based GNNs excel at modeling cortical geometry and text-guided methods enrich segmentation with semantic context, no existing approach jointly addresses surface-space lesion detection, fine-grained clinical narratives, domain shifts across scanners, and zero-shot generalization to new FCD subtypes. Our work aims to fill this gap by integrating rich textual descriptions derived from clinical reports into a multi-resolution GNN framework, with cross-attention fusion and contrastive alignment to improve robustness and detection accuracy across heterogeneous datasets.

3 Methods

4 Results

5 Discussion

6 Conclusion

Appendix

References

[1] R. Durgam, B. Panduri, V. Balaji, A. O. Khadidos, A. O. Khadidos, and S. Selvarajan. “Enhancing lung cancer detection through integrated deep learning and transformer models”. In: *Scientific Reports* 15.1 (2025), p. 15614 (cit. on p. v).

[2] A. B. Abdusalomov, M. Mukhiddinov, and T. K. Whangbo. “Brain tumor detection based on deep learning approaches and magnetic resonance imaging”. In: *Cancers* 15.16 (2023), p. 4172 (cit. on p. v).

[3] K. Sharma, Z. Uddin, A. Wadal, and D. Gupta. “Hybrid Deep Learning Framework for Classification of Kidney CT Images: Diagnosis of Stones, Cysts, and Tumors”. In: *arXiv preprint arXiv:2502.04367* (2025) (cit. on p. v).

[4] A. Mehmood, Y. Hu, and S. H. Khan. “A Novel Channel Boosted Residual CNN-Transformer with Regional-Boundary Learning for Breast Cancer Detection”. In: *arXiv preprint arXiv:2503.15008* (2025) (cit. on p. v).

[5] M. Ripart, H. Spitzer, L. Z. Williams, L. Walger, A. Chen, A. Napolitano, et al. “Detection of epileptogenic focal cortical dysplasia using graph neural networks: a MELD study”. In: *JAMA Neurology* 82.4 (2025), pp. 397–406 (cit. on p. v).

[6] M. Li, M. Meng, S. Ye, M. Fulham, L. Bi, and J. Kim. “Language-guided Medical Image Segmentation with Target-informed Multi-level Contrastive Alignments”. In: *arXiv preprint arXiv:2412.13533* (2024) (cit. on p. v).

[7] C. Li, H. Zhu, R. I. Sultan, H. B. Ebadian, P. Khanduri, C. Indrin, et al. “Mulmodseg: Enhancing unpaired multi-modal medical image segmentation with modality-conditioned text embedding and alternating training”. In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. IEEE. 2025, pp. 3581–3591 (cit. on p. v).

[8] W. Zhang, Z. Zhang, M. He, and J. Ye. “Organ-aware Multi-scale Medical Image Segmentation Using Text Prompt Engineering”. In: *arXiv preprint arXiv:2503.13806* (2025) (cit. on p. v).

[9] Y. Zhong et al. “Ariadne’s thread: Using text prompts to improve segmentation of infected areas from chest x-ray images”. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI 2023)*. Cham: Springer Nature Switzerland, 2023 (cit. on p. v).