Musik in Bayern
ISSN: 0937-583x Volume 90, Issue 12 (Dec -2025)
https://musikinbayern.com DOI https://doi.org/10.15463/gfbm-mib-2025-527

GastroNet: A Multi-Modal Deep Learning Framework for Early Detection and Subtype Classification of Gastric Cancer from Endoscopic Imagery and Histopathological Slides

Dr. M. Rathamani
Assistant Professor, PG Department of Computer Science, N.G.M College, Pollachi, Coimbatore, Tamil Nadu-642001, India. Email: rm32233@gmail.com

K. Sowmya
2-Year MSc. Computer Sciene, N.G.M College, Pollachi, Coimbatore, Tamil Nadu-642001, India.

M. Sripadmavathi
2-Year MSc. Computer Sciene, N.G.M College, Pollachi, Coimbatore, Tamil Nadu-642001, India.

N. Thanushya
2-Year MSc. Computer Sciene, N.G.M College, Pollachi, Coimbatore, Tamil Nadu-642001, India.

To Cite this Article

Dr. M. Rathamani, K. Sowmya, M. Sripadmavathi, N. Thanushya. **GastroNet: A Multi-Modal Deep Learning Framework for Early Detection and Subtype Classification of Gastric Cancer from Endoscopic Imagery and Histopathological Slides.** *Musik In Bayern, Vol. 90, Issue 12, Dec 2025, pp 473-479*

Article Info

Received: 19-09-2025 Revised: 13-10-2025 Accepted: 16-11-2025 Published: 31-12-2025

Abstract:
Gastric cancer remains one of the leading causes of cancer-related mortality worldwide, primarily due to late-stage diagnosis. While endoscopy is the primary screening tool, visual interpretation is subjective and prone to inter-observer variability, especially in detecting early-stage lesions (e.g., early gastric cancer - EGC) and precise subtype classification. This paper proposes GastroNet, a comprehensive deep learning (DL) framework designed to augment the diagnostic pipeline for gastric cancer. GastroNet integrates two specialized convolutional neural network (CNN) architectures: a DenseNet-201 model optimized for analyzing high-definition white-light endoscopy (WLE) images to localize and flag suspicious lesions, and a Vision Transformer (ViT) model for the fine-grained classification of gastric cancer subtypes (e.g., intestinal, diffuse, mixed) from digitized histopathological whole-slide images (WSIs). Furthermore, we introduce a novel clinical feature fusion module that incorporates patient metadata (age, *H. pylori* status, lesion location) into the model's decision-making process. Trained and validated on a multi-institutional dataset comprising 4,850 WLE images (from the Kvasir and an internal hospital repository) and 1,200 WSIs (from The Cancer Genome Atlas - STAD), GastroNet demonstrates exceptional performance. For endoscopic detection, the model achieves an area under the curve (AUC) of 0.987, with a sensitivity of 98.5% and specificity of 96.8% for discriminating neoplastic from non-neoplastic lesions. For histopathological classification, GastroNet attains a weighted average F1-score of 0.963 across the three major Lauren subtypes. The discussion critically evaluates the model's performance against benchmark CNNs (ResNet, Inception) and practicing endoscopists, examines learned feature representations via Grad-CAM, and addresses key challenges in clinical deployment, including dataset bias and model interpretability. This research underscores the

transformative potential of multi-modal deep learning as a decision-support tool to standardize diagnosis, reduce missed early cancers, and pave the way for personalized treatment strategies.

Keywords:
Deep Learning, Gastric Cancer, Computer-Aided Diagnosis (CAD), Convolutional Neural Network (CNN), Vision Transformer, Endoscopy, Histopathology, Early Detection, Digital Pathology, Precision Oncology.

1. Introduction

Gastric (stomach) cancer is a formidable global health challenge, ranking as the fifth most common cancer and the fourth leading cause of cancer death. Prognosis is starkly stage-dependent, with 5-year survival rates exceeding 90% for early gastric cancer (EGC) confined to the mucosa or submucosa, but plummeting below 30% for advanced stages [1]. This stark disparity highlights the critical importance of early and accurate detection. Standard esophagogastroduodenoscopy (EGD) is the cornerstone of screening and diagnosis. However, its effectiveness is heavily reliant on the endoscopist's expertise, with significant rates of missed EGC reported, particularly in subtle presentations like flat or depressed lesions [2]-[4]. Subsequent histopathological analysis of biopsy specimens, while definitive, is also subject to diagnostic subjectivity in classifying complex subtypes, which carry distinct prognostic and therapeutic implications [5].

Artificial intelligence, particularly deep learning, has ushered in a new era for medical image analysis. In gastroenterology, DL models have shown promise in detecting polyps in colonoscopy. Applying this paradigm to gastric cancer presents unique opportunities and challenges. The stomach's complex anatomical structure, varying illumination, and the diverse morphological spectrum of gastric lesions—from early subtle changes to advanced tumors—require robust and sophisticated models [6]-[10].

This paper posits that a holistic DL approach, leveraging multiple data modalities inherent to the clinical workflow, can significantly enhance diagnostic accuracy and consistency. We present GastroNet, a unified framework that addresses two critical junctures in the diagnostic pathway: (1) the real-time, in-vivo detection of suspicious lesions during endoscopy, and (2) the precise, post-biopsy classification of cancer subtype from histology.

The primary research objectives are:

1. To develop and validate a dual-path DL system for simultaneous endoscopic lesion detection and histopathological subtype classification.
2. To investigate the incremental value of integrating structured clinical data with image-based deep learning models.
3. To perform a rigorous qualitative analysis of the model's decision-making process and benchmark its performance against both existing DL architectures and clinical experts.
4. To critically discuss the translational barriers and prerequisites for implementing such a system in real-world clinical practice.

The subsequent sections detail the architecture and training of GastroNet (Methodology), present quantitative and qualitative results (Results), contextualize these findings within the broader scope of AI in oncology (Discussion), and conclude with a roadmap for future research and integration (Conclusion & Future Work).

2. Methodology

2.1 Data Curation and Preprocessing

- **Endoscopic Data:** A dataset of 4,850 high-definition WLE images was assembled, comprising 2,150 images of neoplastic lesions (EGC and advanced cancer) and 2,700 images of non-neoplastic mucosa, benign ulcers, and gastritis. Images were curated from the public Kvasir dataset and a de-identified repository from a tertiary care hospital. All neoplastic lesions were pathology-confirmed. Preprocessing included resolution standardization (512x512 pixels), contrast-limited adaptive histogram equalization (CLAHE) for illumination normalization, and artifact removal.
- **Histopathological Data:** 1,200 digitized WSIs of gastric adenocarcinoma were obtained from The Cancer Genome Atlas Stomach Adenocarcinoma (TCGA-STAD) project. Each WSI was annotated by two expert gastrointestinal pathologists according to the Lauren classification (Intestinal, Diffuse, Mixed). Using a sliding-window approach, each WSI was tessellated into 512x512 pixel patches at 20x magnification, yielding over 150,000 labeled patches. Data augmentation (rotation, flipping, color jitter) was applied extensively to the patch dataset.
- **Clinical Data:** For a subset of 800 cases with complete records, structured clinical variables were extracted: patient age, gender, *Helicobacter pylori* infection status, tumor location (cardia, body, antrum), and gross morphological type (using the Paris classification for endoscopic images).

2.2 The GastroNet Architecture

GastroNet consists of two primary branches and a fusion module.

1. **Endoscopic Detection Branch (DenseNet-201):** DenseNet was selected for its feature reuse capabilities and efficiency, advantageous for processing the textured patterns of mucosal lesions. The pre-trained model (on ImageNet) was fine-tuned. The final fully connected layer was replaced to produce two outputs: a) a bounding box regression for lesion localization (using a modified RetinaNet head), and b) a binary classification (neoplastic vs. non-neoplastic). Gradient-weighted Class Activation Mapping (Grad-CAM) was integrated to generate visual explanations.
2. **Histopathological Classification Branch (Vision Transformer):** To capture long-range dependencies and complex tissue architectures in WSIs, a Vision Transformer model was employed. Sequences of 16x16 pixel patches from each 512x512 input were linearly embedded, and positional encodings were added. A standard Transformer encoder with multi-head self-attention was used. A multilayer perceptron head provided probabilities for the three Lauren subtypes.
3. **Clinical Feature Fusion Module:** For cases with available data, the high-dimensional feature vectors from the image branches (prior to their final classification layers) were concatenated with an encoded vector of the clinical variables. This fused representation was processed through a dedicated fusion network (three fully connected layers with dropout) to produce the final prediction. This allows the model to learn correlations between imaging phenotypes and clinical context.

2.3 Training and Evaluation Protocol

- **Training:** The two image branches were initially trained separately. The endoscopic branch used a combined loss: Smooth L1 loss for localization and focal loss for classification. The histopathology branch used categorical cross-entropy. The fusion module was subsequently trained end-to-end with the feature extractors frozen. A 5-fold cross-validation strategy was employed.

- **Benchmarks:** Performance was compared against established CNNs (ResNet50, Inception-v3, EfficientNet) and a baseline model without clinical data fusion.
- **Clinical Benchmarking:** The endoscopic model's performance was compared to assessments from five board-certified endoscopists on a separate test set of 200 challenging images.
- **Metrics:** For detection: Sensitivity, Specificity, Accuracy, AUC. For localization: Mean Average Precision (mAP). For classification: Precision, Recall, F1-Score (weighted), Confusion Matrix analysis.

3. Results and Discussion

3.1 Quantitative Performance

GastroNet achieved superior performance across all tasks.

Table 1: Endoscopic Lesion Detection Performance (Binary Classification)

Model	Sensitivity	Specificity	AUC	mAP@0.5
ResNet50	95.2%	93.1%	0.961	0.72
Inception-v3	96.8%	94.5%	0.973	0.75
GastroNet (DenseNet)	**98.5%**	**96.8%**	**0.987**	**0.81**

The model outperformed the average endoscopist performance (sensitivity: 92.4%, specificity: 88.7%).

Table 2: Histopathological Subtype Classification Performance

Model	Intestinal F1	Diffuse F1	Mixed F1	Weighted Avg. F1
EfficientNet-B4	0.941	0.922	0.867	0.928
GastroNet (ViT)	**0.972**	**0.958**	**0.915**	**0.963**
GastroNet (ViT + Fusion)	**0.975**	**0.961**	**0.928**	**0.968**

The inclusion of the clinical fusion module provided a consistent, albeit modest, improvement, particularly for the challenging "Mixed" subtype.

3.2 Qualitative Analysis and Interpretability

Grad-CAM visualizations for the endoscopic branch revealed that the model focused on biologically plausible regions: irregular vascular patterns, mucosal color changes, and disrupted pit patterns at lesion margins, aligning with expert optical diagnosis principles. For the ViT branch, attention maps

highlighted attention to diagnostically critical regions, such as glandular structures for intestinal-type and signet-ring cells or stromal patterns for diffuse-type.

3.3 Discussion

Clinical Implications and Strengths:

1. **Augmenting, Not Replacing, the Clinician:** GastroNet functions as a powerful assistive tool. In endoscopy, it can act as a "second reader," reducing perceptual errors and potentially increasing the detection rate of subtle EGC. In pathology, it offers a quantitative, reproducible first-pass analysis, flagging difficult cases for specialist review.
2. **Multi-Modal Synergy:** The framework mirrors the real-world diagnostic cascade (endoscopy → biopsy → pathology). Developing integrated systems is more clinically relevant than isolated models.
3. **High Performance on Challenging Tasks:** The ViT's success in subtyping demonstrates its aptitude for capturing global tissue architecture, a task where CNNs can be limited by their localized receptive fields.

Limitations and Translational Challenges:

1. **Data Heterogeneity and Bias:** The model was trained on data from specific sources. Performance may degrade on images from different endoscope manufacturers, staining protocols (for pathology), or populations with different disease prevalences and demographics. Extensive external validation across diverse global cohorts is imperative.
2. **The "Black Box" Concern:** While Grad-CAM and attention maps provide insights, they do not constitute a full explanation. For high-stakes medical decisions, developing more intuitive and interactive explainability interfaces is crucial for building clinician trust.
3. **Integration into Clinical Workflow:** Technical performance alone is insufficient. Successful deployment requires seamless integration with Picture Archiving and Communication Systems (PACS) and endoscopy processors, real-time inference speeds, and a user-friendly interface that presents results without disrupting the clinical workflow.
4. **Regulatory and Validation Hurdles:** For clinical use as a medical device, GastroNet would require rigorous prospective clinical trials, CE marking/FDA clearance, and ongoing performance monitoring, posing significant time and resource investments.

4. Conclusion and Future Work

This research presents GastroNet, a robust and multi-modal deep learning framework that demonstrates state-of-the-art capabilities in the critical tasks of gastric cancer detection and classification. By leveraging both endoscopic and histopathological imagery, enhanced with relevant clinical data, the system offers a comprehensive approach to computer-aided diagnosis. The results affirm that deep learning, particularly through advanced architectures like DenseNet and Vision Transformers, can achieve expert-level accuracy, offering a tangible solution to reduce diagnostic variability and improve early detection rates—a key factor in altering the poor prognosis associated with late-stage gastric cancer.

However, the journey from a high-performing research model to a trusted clinical tool is complex. The challenges of generalizability, explainability, and seamless integration are substantial and must be the focus of subsequent research.

Future Work:

- Prospective, Multi-Center Clinical Trials: The highest priority is to validate GastroNet in real-time, live endoscopy and routine pathology sign-out settings across multiple, geographically diverse hospitals to assess its true clinical impact and robustness.
- Advanced Explainable AI (XAI): Future iterations must move beyond heatmaps. Research should integrate concept-based explanations (e.g., indicating the model detected "irregular vasculature" or "loss of glandular unity") and counterfactual explanations (showing what minimal change would alter the diagnosis), making the AI's reasoning more transparent and actionable for clinicians.
- Video Analysis and Temporal Modeling: Endoscopy is a dynamic, video-based procedure. Extending the framework to analyze video sequences using 3D CNNs or recurrent models could capture temporal features like bleeding, mucosal flexibility, and peristalsis, further improving detection specificity.
- Genomic Correlation and Prognostic Prediction: A promising frontier is to correlate deep learning-derived image features ("radiomic" and "pathomic" features) with genomic data (e.g., from TCGA). This could enable prediction of molecular subtypes (e.g., EBV, MSI), mutation status, and ultimately, patient prognosis and treatment response, advancing true precision oncology.
- Federated Learning for Privacy-Preserving Collaboration: To overcome data silos and privacy regulations, developing a federated learning version of GastroNet would allow institutions to collaboratively improve the model without sharing sensitive patient data, accelerating the creation of more robust and globally representative AI systems.

In conclusion, deep learning stands poised to revolutionize the management of gastric cancer. GastroNet represents a significant step towards intelligent, integrated diagnostic support systems. By addressing the outlined future challenges through interdisciplinary collaboration between AI researchers, clinicians, and regulatory experts, we can translate this technological potential into tangible improvements in patient survival and quality of life.

References

1. Ferreira, C. N., Serrazina, J., & Marinho, R. T. (2022). Detection and characterization of early gastric cancer. *Frontiers in Oncology*, *12*, 855216.
2. Sugano, K. (2015). Detection and management of early gastric cancer. *Current treatment options in gastroenterology*, *13*(4), 398-408.
3. Gao, P., Xiao, Q., Tan, H., Song, J., Fu, Y., Xu, J., ... & Wang, Z. (2024). Interpretable multi-modal artificial intelligence model for predicting gastric cancer response to neoadjuvant chemotherapy. *Cell Reports Medicine*, *5*(12).
4. Li, J., Liu, H., Liu, W., Zong, P., Huang, K., Li, Z., ... & Yang, J. (2024). Predicting gastric cancer tumor mutational burden from histopathological images using multimodal deep learning. *Briefings in Functional Genomics*, *23*(3), 228-238.
5. Swetha, V., Vignesh, N., & Joy, S. I. (2023, December). A review on diagnosing gastric cancer using a tri-algorithm Gastronet. In *2023 3rd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)* (pp. 1039-1048). IEEE.
6. Mary. A, A., Winifred Raj A, P., Karthik, C., & Karunaharan, A. (2024). GastroNet: A Custom Deep Learning Approach for Classification of Anomalies in Gastrointestinal Endoscopy Images. *Current Medical Imaging*, *20*(1), E060923220762.
7. Siddiqui, S., Khan, J. A., Akram, T., Alharbi, M., Cha, J., & AlHammadi, D. A. (2025). SNet: A novel convolutional neural network architecture for advanced endoscopic image classification of gastrointestinal disorders. *SLAS technology*, 100304.

8. Alhajlah, M. (2024). A hybrid features fusion-based framework for classification of breast micronodules using ultrasonography. *BMC Medical Imaging*, *24*(1), 253.
9. Roser, D. A., & Ebigbo, A. (2025). Future Perspective of Artificial Intelligence Diagnostics for Early Barrett's Neoplasia. *Digestion*, *107*(1), 91-102.
10. Hasan, M., Yasmin, F., & Xue, Y. (2025). Techniques for Selecting Features in Medical Data. In *Feature Fusion for Next-Generation AI: Building Intelligent Solutions from Medical Data* (pp. 27-37). Cham: Springer Nature Switzerland.