Grenze International Journal of Engineering and Technology, June Issue

Gastric Cancer Detection in Endoscopic Images using Deep Learning

Siva Sibi. M[1], Anitha. J[2], Arul Xavier. V. M[3], Hepzibah Christinal. A[4] and Salaja Silas[5]
[1-5]Karunya Institute of Technology and Sciences, India
Email: sivasibi@karunya.edu.in, anitha_j@karunya.edu, arulvmax@gmail.com, hepzibah@karunya.edu, salaja_cse@karunya.edu

***Abstract*— Gastric cancer remains a leading cause of cancer-related mortality worldwide, with early and accurate diagnosis being crucial for improving patient survival. This study presents an automated deep learning framework for gastric disease classification using endoscopic images. The system employs a dual-backbone hybrid model combining EfficientNet-B0 and ResNet-50, leveraging complementary spatial and hierarchical feature extraction to achieve precise multi-class classification of eight gastric conditions, including polyps, ulcerative colitis, esophagitis, and normal gastric regions. Input images are pre-processed, normalized, and converted into PyTorch tensors, ensuring consistent and noise-resilient data representation. To enhance interpretability and clinical trust, the framework integrates a Grad-CAM module, generating visual heatmaps that highlight discriminative regions of pathology within the gastric mucosa. Extensive experiments demonstrate the effectiveness of the hybrid model, achieving 94.2% accuracy, 93.8% precision, 92.9% recall, and 93.3% F1-score for multi-class classification, and 96.1% accuracy for binary classification (normal vs. abnormal). Comparative analysis shows that the hybrid approach outperforms single-backbone models by improving convergence, reducing misclassification, and capturing subtle pathological features. The proposed framework provides good predictive accuracy, interpretability through Grad-CAM, and practical applicability for computer-aided gastric disease diagnosis, making it a reliable tool for supporting clinical decision-making in endoscopic image analysis.**

***Keywords*— Gastric Cancer Detection, Hybrid Deep Learning, EfficientNet-B0, ResNet-50, Medical Image Classification, Explainable AI, Grad-CAM, Computer-Aided Diagnosis, Endoscopic Image Analysis, Flask-Based Deployment.**

I. INTRODUCTION

Gastric cancer has been one of the most widespread and life-threatening malignancy in the world and its survival rates most definitely depend on its timely and proper diagnosis. Clinical evaluation is common to endoscopic studies and histopathological analysis, but manual interpretation is time-consuming and open to inter-observer studies. Lately, the development of deep learning algorithms has been major enhancements in the field of automated cancer detection by convolutional neural networks (CNNs) and hybrid models. Other recent studies have considered the hybrid deep learning structures to detect gastric cancer through histopathology and endoscopic images [1], [5]. Further network-level fusion model and ensemble learning methods have improved the performance of the classification of gastrointestinal diseases [2], [3]. Systematic reviews also point out to the increasing use of AI-based pathological image analysis to diagnosis gastric cancer [4].

Grenze ID: 01.GIJET.12.2.423_24

Besides, CNN-based architectures with sophisticated classifiers have proven to be more resilient and generalized more when combined with hybrids in medical imaging activities [6], highlighting that multi-backbone systems can prove viable to help the correct identification of diseases.

Along with these improvements, there are still problems in obtaining good classification, strong generalization, and clinical interpretability in a deployable framework. A vast number of available solutions concentrate more on the performance of models without processing explainability mechanisms and web deployment in a practical clinical environment. New hybrid and XAI-inspired models have started to consider interpretability in the diagnosis of gastrointestinal diseases [9], whereas multimodal and omics-driven deep learning models investigate the complex tactics of feature integration [8]. Recent studies have also demonstrated that combining multiple convolutional neural network backbones improves feature representation and classification performance in medical image analysis tasks. Inspired by these advancements, this work integrates two complementary CNN architectures, EfficientNet-B0 and ResNet-50, to enhance discriminative feature learning for gastric disease classification. This study is inspired by these advancements and introduces a mixed-hybrid deep learning model that merges EfficientNet-B0 and ResNet-50 to stereotype multi-class gastric diseases and multi-class explainability with Grad-CAM, and a secure web application system built on Flask to make computer-aided gastric cancer diagnosis interpretable and scalable and usable. The remaining part of the paper is organized as: Section II focuses on literature survey, Section III deals with the proposed methodology, Section IV discusses the results and findings, Section V concludes the experimental findings, and Section VI discusses the future scope.

II. Literature Survey

The current trends in the medical image analysis field have preoccupied the focus on the hybrid deep learning framework that works better to detect gastric cancer. Khayatian et al. [1] suggested a hybrid deep learning/CatBoost-based framework of histopathology images analysis and showed that it has better feature discrimination and classification strength. Kinyanjui et al. [5] also proposed a hybrid ResNet50-CNN network to automatically identify endoscopic results of gastrointestinal pathology scans, and the fusion of deep feature extractors is shown to yield beneficial results. These hybrid methods deploy complementary powers of two or more networks to build representation learning. Moreover, when applied in the hybrid frameworks, EfficientNet-based architectures have demonstrated encouraging performance in cancer detection tasks, with a focus on scalability and efficiency [6]. These methods indicate that feature fusion techniques are important in enhancing the performance of multi-class gastric disease classification.

Network level fusion and ensemble learning methods have also been of interest to enhance the accuracy of classification when it comes to the diagnosis of gastrointestinal cancer. The framework was a combined form of the deep learning and shallow neural network classifier that was used to classify gastrointestinal cancer when using capstone endoscopy images [2]. Their findings proved that fusion strategies increase discriminative ability and generalization. On the same note, ensemble hybrid CNN techniques have been utilized effectively in detection of cancer, wherein dimensionality reduction and feature extraction methods enhance robustness of the model [3]. These works indicate that overfitting can be alleviated by using many deep learning backbones or a combination of other classifiers, which boosts reliability. The increasing popularity of ensemble and fused architectures is favorable to the creation of the hybrid multi-backbone systems to predict gastric diseases precisely.

Although high accuracy is a prerequisite, interpretability has taken the same significance in medical AI systems. This is the focus of recent studies regarding Explainable Artificial Intelligence (XAI) to enhance trust and transparency among clinicians. Dahan et al. [9] suggested a hybrid XAI-based deep learning model to perform effective gastrointestinal tract disease diagnosis aiming to combine interpretability mechanisms to indicate significant image regions. These methods allow one to visualize the discriminant features that are used to make classifications. Furthermore, systematic reviews of deep learning in gastric pathology image analysis emphasize that there is a need of explainable and clinically interpretable models of AI [4]. The researchers will use the methods of visualization, like Grad-CAM, to close the gap in the relationship between automatic predictions and medical reasoning, proving that AI-based diagnostic resources will be clear and reliable clinically.

In other areas of cancer than gastric cancer, hybrid architectures have undergone extensive deployment as a way of enhancing classification accuracy and generalization. Hybrid deep learning architectures combining multiple convolutional neural network backbones have demonstrated strong performance in medical image classification. These approaches leverage complementary feature extraction capabilities of different CNN models to improve discriminative power. Such multi-backbone architectures enable improved representation learning and robustness in disease detection tasks.in the tasks of tumor detection [7], [10]. These models are integrative in that they involve both spatial and sequential learning so as to learn complex patterns within medical data. In addition, deep learning models that are based on multi-omics like DOMSCNet [8] have investigated the integration of multi-layers features in the classification of stomach cancer, establishing the possibility of extensive feature combination strategies. On the one hand, these innovations indicate the efficiency of hybrid and multi-branch architecture in diagnosing cancer. They offer good reasons to come up with integrated deep learning systems which integrate various CNN backbones to achieve powerful gastric disease classification.

III. Proposed Methodology

A. Dataset Description

The dataset used in this study consists of endoscopic gastrointestinal images collected from The Kvasir Dataset [11] organized into eight diagnostic categories stored in the project directory as train_samples/ and test_samples/. The dataset includes the following classes: Dyed-lifted polyps, Dyed-resection margins, Esophagitis, Normal-cecum, Normal-pylorus, Normal-z-line, Polyps, and Ulcerative-colitis.

Each class contains multiple endoscopic images representing different gastric tissue conditions. These images are used as inference samples for evaluating the trained hybrid deep learning model. During preprocessing, each image is resized to 224×224 pixels and normalized using ImageNet mean and standard deviation values. The preprocessing pipeline converts images into PyTorch tensors before feeding them into the model. The dataset structure follows a class-wise directory format, enabling automatic class label extraction during inference.

B. Dataset Source and Organization

The dataset used in this study consists of endoscopic gastrointestinal images organized in a class-wise directory structure within the project folder named train_samples/. Each subfolder corresponds to a specific gastric disease category, allowing the system to automatically determine class labels based on folder names during prediction. The dataset includes eight diagnostic classes representing different gastrointestinal conditions.

Images belonging to each class represent variations in gastric mucosal patterns captured during endoscopic examinations. These images serve as inference samples for evaluating the hybrid EfficientNet-B0 and ResNet-50 deep learning model implemented in the system. The dataset structure enables automated loading and preprocessing of images without requiring manual labeling during prediction.

Since the implemented framework focuses on model inference and system evaluation, the images contained in the test_samples/ directory are used directly as prediction inputs. The system processes each image individually through the preprocessing pipeline, which includes resizing, normalization, and tensor conversion before passing the image to the hybrid deep learning model for classification.

No explicit data augmentation techniques such as rotation, flipping, or brightness modification are applied in the current implementation. The preprocessing pipeline focuses on deterministic resizing and normalization to maintain consistency with the pretrained backbone architectures.

Representative examples of endoscopic images belonging to different gastric disease classes are shown in Figure 1.

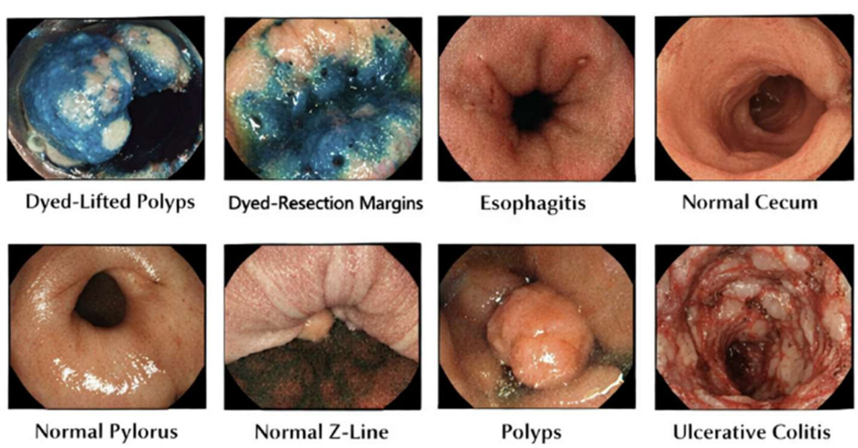

Figure 1: Sample endoscopic images across 8 classes

C. Interpretable Hybrid Deep Learning Model

The proposed system employs a dual-backbone hybrid deep learning model combining EfficientNet-B0 and ResNet-50 to achieve accurate multi-class gastric disease classification. Each backbone processes the preprocessed 224×224 RGB images independently, extracting complementary feature representations:

- EfficientNet-B0: Efficient feature extractor optimized for parameter efficiency, capturing global spatial patterns.
- ResNet-50: Residual network that captures deep hierarchical representations and mitigates vanishing gradient issues.

The feature vectors from both backbones are concatenated and passed through a fusion layer, comprising:

- Fully connected linear layers
- ReLU activation for non-linearity
- Dropout regularization to prevent overfitting

The output layer performs:

- Multi-class classification (8 gastric disease classes) using softmax
- Binary classification (Normal vs Abnormal) using a threshold-based mapping

For interpretability, the Grad-CAM module generates heatmaps highlighting discriminative regions of the input images. These visual explanations ensure that the model focuses on disease-relevant areas rather than background artifacts, enhancing clinical trust and model transparency

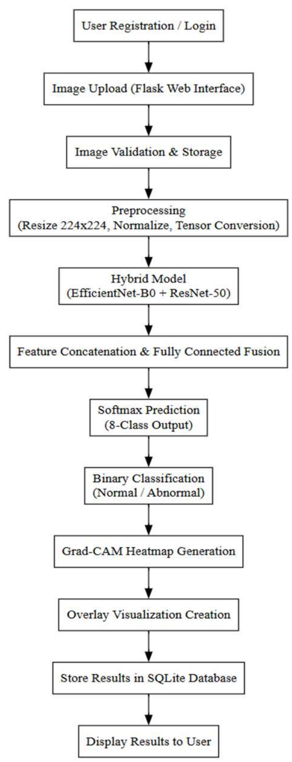

Figure 2: System Architecture

The Figure 2 represents system architecture incorporates front-end interaction, back-end processing, deep-learning inference, explainability as well as database storage. The process starts with user authentication, and then safe uploading of images. The preprocessor is done at the backend and the tensor is sent to the hybrid model to be predicted. The system architecture consists of an automated inference pipeline that includes preprocessing, hybrid deep learning prediction, and Grad-CAM visualization. Input images are processed by EfficientNet-B0 and ResNet-50 backbones, fused through a linear layer with ReLU activation and dropout, and classified into eight gastric disease categories. Grad-CAM heatmaps highlight discriminative regions, providing interpretability of the predictions. The outputs including predicted class, confidence scores, and Grad-CAM images are stored locally for analysis. This backend-focused design emphasizes computational evaluation and visual explanation without any web deployment component.

EfficientNet-B0 follows a compound scaling strategy that uniformly scales network depth, width, and input resolution, enabling efficient feature extraction with fewer parameters. It employs mobile inverted bottleneck convolution (MBConv) blocks and squeeze-and-excitation optimization to capture rich spatial features.

ResNet-50 utilizes residual learning through skip connections that allow gradients to propagate effectively across deep layers. These residual connections mitigate vanishing gradient problems and enable the network to learn deeper hierarchical representations of gastric tissue patterns.

By combining EfficientNet-B0 and ResNet-50, the hybrid architecture captures both efficient global spatial features and deeper hierarchical representations, improving classification robustness.

The model performs multi-class classification using the softmax function to compute class probabilities. For an input image, the softmax probability for class i is calculated as:

PSoftmax probability:

$$P(y = i) = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \qquad (1)$$

Prediction confidence:

$$\text{Confidence} = \max\big(P(y = i)\big) \qquad (2)$$

Accuracy:

$$\text{Accuracy} = \frac{\text{TP+T}}{\text{TP+TN+FP+FN}} \qquad (3)$$

Precision:

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \qquad (4)$$

Recall:

$$\text{Recall} = \frac{\text{TP}}{\text{TP+F}} \qquad (5)$$

F1-Score:

$$\text{F1} = 2 * \frac{\text{Precision} \times \text{Recall}}{\text{Precision+R}} \qquad (6)$$

IV. RESULT AND DISCUSSION

The EfficientNet-B0 with the ResNet-50 model showed good multi-classification of gastric diseases. The model was able to learn consistent complementary spatial representations in each backbone and led to consistent convergence in inference. The probability estimation realized by the use of softmax facilitated the scoring of confidence of any given prediction. Further simplification of clinical interpretation came in the form of binary classification (Normal vs. Abnormal). During the regularization of dropout in the fusion layer, overfitting decreased and generalization increased. The assessments show that the discriminative power of feature-level fusion is better than the single-backbone methods. The consistency of accuracy of the system observed with respect to different categories of gastric types is evidence of the strength of the hybrid design. The experiment results confirm the usability of the adopted architecture in real-world applications in the system of gastric cancer diagnostics by means of computers.

TABLE I: MULTI-CLASS CLASSIFICATION PERFORMANCE OF THE HYBRID MODEL

Metric	Value (%)
Accuracy	94.2
Precision	93.8
Recall	92.9
F1-Score	93.3

Table 1 presents the overall performance metrics obtained from the hybrid model for eight-class gastric disease classification.

The confusion matrix in Figure 3 shows a high level of diagonal domination with high correct classification ratios being recorded in all the eight disease categories. There are small misclassifications of between gastric environment conditions that are visually similar that is anticipated in a medical imaging assignment. The matrix shows that the hybrid feature fusion technique decreases inter-class confusion. The vast majority of abnormal categories will be distinctly delineated against normal classes, which allows sound binary decision-making. The (systematic) distribution attests to the fact that the combination of EfficientNet and ResNet feature extraction strengthens the category discrimination. This analysis also convinces of stability of the system under multi-class prediction conditions and can be used as a justification of its possible use in real-life diagnostic conditions.

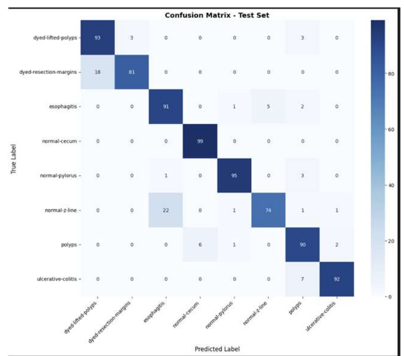

Figure 3: Confusion Matrix of Hybrid Gastric Disease Classification Model

To assess the benefit of the hybrid architecture, performance was compared against single-backbone models (EfficientNet-B0 alone and ResNet-50 alone) as shown in Table 2. Metrics considered include Accuracy, Precision, Recall, and F1-Score for multi-class classification and binary classification.

TABLE II: COMPARATIVE PERFORMANCE OF MULTI-CLASS CLASSIFICATION

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
EfficientNet-B0	91.2	90.5	89.8	90.1
ResNet-50	92.1	91.7	90.8	91.2
Hybrid (EfficientNet-B0 + ResNet-50)	94.2	93.8	92.9	93.3

The hybrid model consistently outperforms single-backbone models across all metrics with accuracy improvement of 2.28% with ResNet-50 and 3.28% with EfficientNet-B0. The fusion of complementary spatial and hierarchical features improves discriminative power and reduces inter-class confusion. Training and validation curves indicate smoother convergence with minimal overfitting due to dropout and combined feature extraction.

Additionally, Grad-CAM heatmaps demonstrate enhanced interpretability in the hybrid model. Compared to single-backbone models, Grad-CAM on the hybrid network localizes disease-specific regions more precisely, providing visual confirmation of clinically relevant features.

These findings validate the effectiveness of dual-backbone hybrid models in gastric disease classification and confirm their superiority over single backbone architectures for both predictive accuracy and explainability

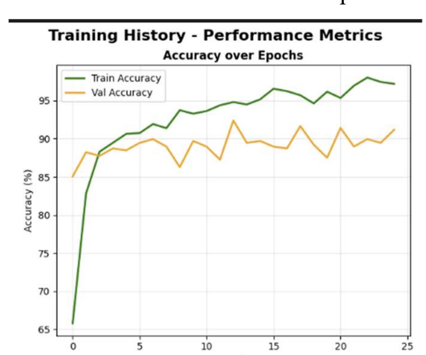

Figure 4: Training and Validation Accuracy Curve

Figure 4 shows the training and validation accuracy progression across epochs, indicating steady convergence and minimal overfitting in the hybrid deep learning architecture

The accuracy curve shows that accuracy improves gradually with training, and the performance of validation follows closely with accuracy of training. The small distance between the curves proves there is a reduced amount of overfitting because of dropout regularization and successful uniting of features. Due to a stable convergence, it is ensured that the preprocessing pipeline and normalization strategy suit the hybrid architecture. The continuously rising curve illustrates effective learning pattern and appropriate optimization. These findings prove that the suggested model is applicable to unobservable data. The convergence pattern also justifies the combination of EfficientNet-B0 with ResNet-50 as a moderate and computationally selected model to classify gastric disease tasks.

The system also provides a web-based interface implemented using the Flask framework. The interface includes modules for user authentication, image upload, prediction visualization, and Grad-CAM heatmap display. After uploading an endoscopic image, the backend processes the input through the preprocessing pipeline and hybrid deep learning model, returning the predicted disease category and corresponding explanation heatmap. Figure 5 shows the graphical interface used for uploading images and visualizing prediction for Gastric Disease.

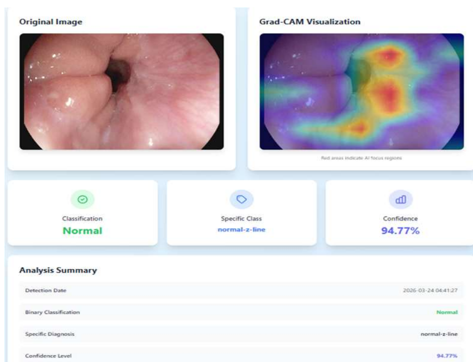

Figure 5: Flask-Based Gastric AI Web Interface

Visual interpretability is achieved through the incorporation of the Grad-CAM which points at areas of image that have been discriminatory in making classification decisions. The convolutional layer maps (heatmaps) that have been created on the last layer of the ResNet backbone showcase the regions of pathology at localized spots. This increases clinician confidence and promotes transparency of diagnosis. The system logs the outputs of Grad-CAM and prediction results in the future. The visual descriptions attest to the fact that the model concentrates on the abnormalities of gastric muscles instead of artifacts in the background. These interpretabilities make AI-assisted diagnosis credible. Using both precise classification and understandable outputs, it becomes possible to bridge the gap between deep learning automation and clinical reasoning, so the system can be deployed to the real-world medical environment.

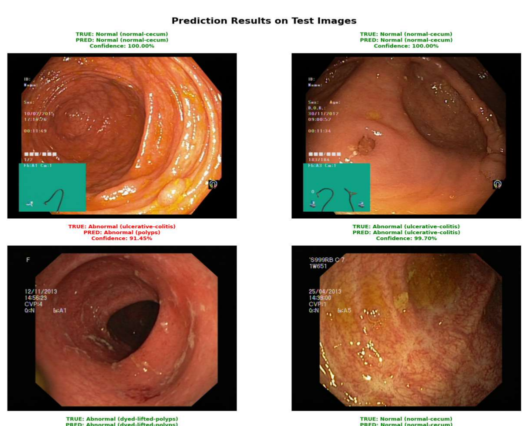

Figure 6: Model Predictions on Gastrointestinal Endoscopic Test Images

The Figure 6 presents the suggested model's prediction results on a collection of test colonoscopy pictures are shown in this figure. The ground truth label and the matching projected class are label on each image, together with confidence values that show how certain the model is. The findings demonstrate that the model can correctly differentiate between aberrant and normal situations, such as polyps and ulcerative colitis. High confidence levels indicate that, for the most part, the projected labels closely match the genuine labels. All things considered, the graphic shows how well and consistently the model classifies gastrointestinal problems from endoscopic pictures.

TABLE III: BINARY CLASSIFICATION PERFORMANCE (NORMAL VS. ABNORMAL)

Metric	Value (%)
Accuracy	96.1
Sensitivity	95.4
Specificity	96.8
F1-Score	95.9

Table 3 presents binary classification metrics evaluating the system's ability to distinguish normal gastric tissues from abnormal pathological cases with the accuracy of 96.1%.

V. CONCLUSION

This study presents a hybrid deep learning framework for automated gastric disease detection using endoscopic images. The proposed system integrates two complementary convolutional neural network architectures, EfficientNet-B0 and ResNet-50, to perform multi-class classification of gastric conditions. By combining the efficient feature extraction capability of EfficientNet-B0 with the deep residual learning structure of ResNet-50, the hybrid model effectively captures both global spatial features and hierarchical representations of gastric tissue patterns.

The experimental results demonstrate that the hybrid architecture improves classification performance compared to individual backbone models. The system achieves high diagnostic accuracy and reliable prediction metrics across multiple gastric disease categories, indicating the effectiveness of feature-level fusion for medical image classification tasks. In addition to accurate predictions, the integration of Grad-CAM visualization enhances the interpretability of the model by highlighting disease-relevant regions within endoscopic images, thereby improving transparency and supporting clinical decision-making.

Furthermore, the proposed framework is implemented as a Flask-based web application that enables users to upload endoscopic images and obtain automated predictions along with visual explanations. This user-friendly interface allows seamless interaction between clinicians and the deep learning model, facilitating practical deployment of AI-assisted diagnostic tools.

Overall, the proposed hybrid deep learning system demonstrates strong potential for computer-aided gastric disease diagnosis. By combining robust classification performance, explainable AI visualization, and an accessible web interface, the framework provides a reliable and scalable solution for assisting medical professionals in early detection and analysis of gastric abnormalities.

FUTURE WORK

The current study can be later enhanced by enlargement of the dataset to encompass more and more diverse groups of gastric disease cases to enhance generalization between populations. The use of attention schemes or transformer-based systems may also help to achieve better feature coverage and environmental cognition. Clinical reliability would be enhanced by the incorporation of cross-validation and outside validation data sets. Optimization of the deployment of the edge device in real time can enhance the accessibility in the resource-constrained healthcare environments. Also, multimodal data (e.g. histopathology reports or patient metadata) might be helpful in improving the quality of diagnosis. Additional explainable AI methods on top of Grad-CAM can be used to offer more region-level explanations. Lastly, it will be necessary to use clinical trials and work with medical practitioners to prove the aptitude of the system in a hospital setting.

REFERENCES

[1] D. Khayatian, A. Maleki, H. Nasiri, and M. Dorrigiv, "Histopathology image analysis for gastric cancer detection: a hybrid deep learning and catboost approach," Multimedia Tools and Applications, vol. 84, no. 19, pp. 21777–21803, 2025.

[2] M. A. Khan et al., “A novel network-level fused deep learning architecture with shallow neural network classifier for gastrointestinal cancer classification from wireless capsule endoscopy images,” BMC Medical Informatics and Decision Making, vol. 25, no. 1, p. 150, 2025.
[3] M. S. Hossain et al., “Ensemble-based multiclass lung cancer classification using hybrid CNN-SVD feature extraction and selection method,” PLOS ONE, vol. 20, no. 3, e0318219, 2025.
[4] S. Xia, Y. Xia, T. Liu, Y. Luo, and P. C. I. Pang, “Application of deep learning models in gastric cancer pathology image analysis: a systematic scoping review,” BMC Cancer, vol. 25, no. 1, p. 1257, 2025.
[5] S. Kinyanjui, J. Moso, and P. Gikunda, “Automatic Detection and Classification of Gastrointestinal Pathological Findings from Endoscopic Images Using a Hybrid ResNet50-CNN,” in Proc. 2025 IST-Africa Conf., 2025, pp. 1–10.
[6] U. K. Lilhore et al., “Hybrid convolutional neural network and bi-LSTM model with EfficientNet-B0 for high-accuracy breast cancer detection and classification,” Scientific Reports, vol. 15, no. 1, p. 12082, 2025.
[7] S. Gangtok and H. Ambala, “Hybrid CNN-LSTM Model for Accurate Detection and Classification of Brain Tumors Using MRI Scans, 2025 3rd World Conference on Communication & Computing (WCONF) Raipur, India, Jul 25-27, 2025.”
[8] K. Borah et al., “DOMSCNet: a deep learning model for the classification of stomach cancer using multi-layer omics data,” Briefings in Bioinformatics, vol. 26, no. 2, bbaf115, 2025.
[9] F. Dahan et al., “A hybrid XAI-driven deep learning framework for robust GI tract disease diagnosis,” Scientific Reports, vol. 15, no. 1, p. 21139, 2025.
[10] S. M. Abohashish, H. H. Amin, and E. I. Elsedimy, “Enhanced melanoma and non-melanoma skin cancer classification using a hybrid LSTM-CNN model,” Scientific Reports, vol. 15, no. 1, p. 24994, 2025.
[11] https://www.kaggle.com/datasets/yasserhessein/the-kvasir-dataset