GLFF-NET: A HYBRID CONVNEXT-V2-TINY AND PVTV2-B2 FRAMEWORK WITH GASTRIC LESION FEATURE FUSION FOR ENDOSCOPIC GASTRIC CANCER CLASSIFICATION

Vaishnavi Satheesh[1], J. Sumitha[2]

Abstract

Gastric cancer is one of the leading reasons of cancer-related mortality globally, making early and accurate diagnosis vital for improving patient outcomes. Although endoscopic imaging is standard practice for evaluating gastric lesions, manual diagnosis remains labor-intensive and highly reliant on the examiner's clinical experience. Current methods often struggle to balance the capture of fine-grained lesion details with long-range contextual information, ultimately limiting their classification accuracy. To address these limitations, this paper proposes Gastric Lesion Feature Fusion Network (GLFF-Net), a hybrid deep learning (DL) framework that integrates Convolutional Neural Network (CNN) Next Version 2 Tiny (ConvNeXt V2-Tiny) and Pyramid Vision Transformer Version 2 Base 2 (PVTv2-B2) through a novel Gastric Lesion Feature Fusion (GLFF) module for multiclass gastric cancer classification. ConvNeXt V2-Tiny extracts discriminative local lesion features, while PVTv2-B2 learns global contextual representations from endoscopic images. The GLFF module effectively fuses these complementary features to enhance diagnostic discrimination. The framework was evaluated on the Fujita Health University Hospital gastric endoscopic dataset comprising 8,416 Red, Green, and Blue (RGB) endoscopic images belonging to adenocarcinoma, lymphoma, and normal classes. Experimental outcomes demonstrated that the proposed framework attained an accuracy of 98.80%, precision of 98.72%, recall of 98.65%, F1-score of 98.68%, and an Area Under the Receiver Operating Characteristic Curve (AUC) of 99.70%. Comparative evaluation indicated that GLFF-Net achieved higher classification performance than existing deep learning methods, demonstrating its effectiveness as an automated framework for gastric cancer classification from endoscopic images.

Department of Computer Science[1]
Jai Bharath Arts and Science College[1]
Arakkappady, Vengola[1]
vaishnavisatheesh2000@gmail.com[1]

Department of Computer Science[2]
Faculty of Arts, Science[2]
Commerce and Management[2]
Karpagam Academy of Higher Education, Coimbatore[2]
sumivenkat2006@gmail.com[2]

* Corresponding Author

Keywords : Endoscopic images, Gastric cancer, ConvNeXt V2-Tiny, PVTv2-B2, Feature fusion, Deep learning

I. INTRODUCTION

Gastric cancer is one of the most prevalent malicious tumors affecting the stomach and remains a major cause of cancer-related mortality worldwide. It develops due to the uncontrolled proliferation of abnormal gastric epithelial cells that progressively invade surrounding tissues and metastasize to distant organs if left untreated [1]. Several risk factors, including Helicobacter pylori infection, genetic predisposition, unhealthy dietary habits, smoking, alcohol consumption, obesity, and chronic gastric inflammation, contribute to the development of gastric cancer. Since the disease often exhibits mild or nonspecific symptoms during its early stages, diagnosis is frequently delayed until advanced progression. Therefore, early and precise recognition is important for developing treatment outcomes, increasing survival rates, and reducing disease-related mortality.

Gastric cancer remains to impose a significant global healthcare burden, accounting for a substantial number of new cases and cancer-related deaths each year [2]. Although advances in medical treatment have improved patient management, the prognosis remains poor for individuals diagnosed at advanced stages. The growing incidence associated with population aging, lifestyle changes, and environmental risk factors has increased the demand for effective screening strategies and reliable diagnostic techniques that facilitate early clinical intervention.

Conventional diagnosis of gastric cancer primarily relies on endoscopic examination, histopathological analysis, and medical imaging. Endoscopy remains the standard clinical procedure for identifying gastric lesions; however, its diagnostic accuracy depends largely on the experience of clinicians and careful visual interpretation. Earlier computer-aided diagnostic systems employed handcrafted image features combined with traditional machine learning algorithms such as Support Vector Machine (SVM), Random Forest (RF), and Decision Tree (DT) for gastric lesion classification [3]. Despite their effectiveness in specific applications, these approaches rely heavily on manual feature engineering and often struggle to capture the complex morphological variations of gastric lesions, resulting in inconsistent classification performance under diverse

imaging conditions [4].

Recent developments in Artificial Intelligence (AI) and DL significantly advanced automated gastric cancer diagnosis through endoscopic image analysis. CNNs, Vision Transformers (ViTs), and hybrid DL architectures have demonstrated improved capability in learning discriminative local and global image representations for gastric cancer classification [5-6]. However, existing methods still encounter challenges in effectively integrating fine-grained lesion characteristics with long-range contextual information, which limit classification accuracy. In response to these shortcomings, this study presents GLFF-Net, a hybrid DL approach that integrates ConvNeXt V2-Tiny and PVTv2-B2 for multiclass gastric cancer classification from endoscopic images. The proposed framework lies in the integration of convolutional and transformer-based feature extraction through the proposed Gastric Lesion Feature Fusion (GLFF) module, which effectively combines local lesion characteristics with global contextual information to enhance discriminative feature representation. The proposed framework is evaluated using the Fujita Health University Hospital gastric endoscopic dataset. The primary aims of this study are as follows:

- To develop a hybrid GLFF-Net framework for multiclass gastric cancer classification from endoscopic images.
- To improve feature representation by integrating ConvNeXt V2-Tiny, PVTv2-B2, and the proposed GLFF module.
- To enhance the classification performance of adenocarcinoma, lymphoma, and normal gastric endoscopic images through effective fusion of local and global image features.

The remainder of this paper is prepared as follows. The existing research gap is identified through a review of recent studies on gastric cancer classification in Section 2. The dataset description, preprocessing steps, hybrid model architecture, Gastric Lesion Feature Fusion module, and experimental setup of the proposed GLFF-Net framework are detailed in Section 3. Comprehensive experimental outcomes and comparative performance evaluations are given in Section 4. Subsequently, the final conclusions and potential pathways for future work are summarized in Section 5.

II. LITERATURE REVIEW

Khayatian et al. [7] developed a DL approach for gastric cancer classification by means of histopathological images from the GasHisSDB dataset. EfficientNetV2B0 integrated with CatBoost achieved the highest classification accuracies of 89.7%, 93.1%, and 93.9% for different image crop sizes. Furthermore, Grad-CAM and cross-validation were employed to validate the feature learning capability of the proposed framework. However, the limited diversity of histopathological images and variations in staining conditions restricted the generalization capability of the model.

Noh et al. [8] presented a gastric cancer detection framework based on Raman spectroscopy for distinguishing gastric adenocarcinoma from benign gastric tissues. The study employed PCA for dimensionality reduction and LDA for classification using spectral data collected from 19 high-risk patients. Experimental results demonstrated a sensitivity of 94.2%, an accuracy of 90.5%, and an AUC 0.957. Although the framework effectively differentiated malignant and benign tissues, its clinical applicability was limited by the relatively small sample size.

Rubab et al. [9] introduced a computer-aided diagnosis framework for GI disease classification using the Kvasir endoscopic image dataset. The framework utilized modified ResNet-18 and ResNet-50 architectures in conjunction with the Marine Predator Optimization (MPO) algorithm to refine deep feature representations and improve classification performance. Experimental results established better diagnostic accuracy and computational efficiency compared with baseline models. However, the study relied on multiple independent classifiers rather than a unified end-to-end architecture, increasing computational complexity and limiting its suitability for efficient clinical deployment.

Ke et al. [10] developed an early gastric cancer detection framework to distinguish gastric cancer from precancerous lesions using routinely available clinical laboratory data. The study analyzed inpatient records collected between 2018 and 2023 and employed the XGBoost machine learning algorithm to construct the XHGC20 model, which integrated 20 non-invasive laboratory indicators. The model demonstrated high diagnostic performance, achieving an 0.901 AUC, 0.830 sensitivity, and 0.806 specificity in differentiating gastric cancer from precancerous conditions. It also showed effective identification of early-stage gastric cancer, with a 0.888 AUC, 0.797 sensitivity and 0.807 specificity. Although the model achieved better diagnostic performance, its reliance on retrospective single-center data and routine laboratory indicators limits generalizability, necessitating prospective validation and the inclusion of additional biomarkers.

Tian et al. [11] introduced and validated a ML-based predictive framework to identify distant metastasis in patients with stage T1 gastric cancer. Method analyzed data from 2,698 patients obtained from the SEER database and 107 patients from a hospital cohort for external validation. The RF

model reached the better performance with an AUC of 0.941, 0.917 accuracy, 0.841 recall, 0.927 specificity, and 0.877 F1-score in the training set, while external validation yielded an AUC of 0.750. The study was limited by its retrospective design, and the absence of variables such as tumor markers, nutritional indices, and inflammatory indicators.

Park et al. [12] presented an artificial intelligence-assisted gastric cancer detection approach using a compact hyperspectral imaging system integrated with structured illumination. The study combined hyperspectral imaging with extraction of intrinsic tissue optical characteristics, including reflectance spectra, scattering coefficients, and absorption coefficients. A ViT model was employed to classify tissue types using 64 optical features, while a pixel-level labeling strategy aligned pathological findings with imaging data to generate accurate training samples. The trained ViT model accurately differentiated normal, adenoma, and cancerous tissues, producing diagnostic outcomes with an accuracy of 91.3%. Limitations included variability associated with tissue-mimicking phantoms, synchronization issues during image acquisition, and the absence of depth information in labeling data.

Zhang et al. [13] developed the application of DL for the recognition and lesion segmentation of EGC from gastroscopic images. Experimental results demonstrated performance compared with baseline models, achieving 93.9% accuracy, 92.9% precision, 95.3% recall, 92.5% specificity, and a 94.1% F1-score. Although the IMR-CNN model demonstrated high detection and segmentation performance, its clinical utility remains uncertain because prospective real-time validation in routine endoscopic practice was not performed.

Chae and Cho [14] developed a computer-aided diagnosis structure for the early detection of gastric abnormalities using gastroscopic images. A ViT-based DL model was employed in combination with a Multi-Filter AutoAugment (MFAA) technique to enhance data augmentation and improve learning from limited medical datasets. Performance was estimated using sensitivity, F1-score, and AUC. Model accomplished a 0.87 F1-score and a 0.94 AUC. Limitation of this study is the relatively small and institution-specific training dataset, comprising gastroscopic images from only 192 patients at a single medical center, which limits the generalizability of the study to broader and more diverse clinical populations.

Gao et al. [15] designed an automated DL framework to differentiate EGC from advanced gastric cancer using portal venous CT images. The segmentation stage automatically identified the stomach region, which was subsequently cropped and standardized before classification. The segmentation model attained a Dice score of 0.94, indicating high segmentation accuracy. On the training dataset, the classification model yielded an 0.98 AUC and a 0.93 accuracy rate, whereas the validation set produced an AUC of 0.96 alongside an accuracy of 0.92. Limitations included the retrospective single-center design, exclusive inclusion of gastric cancer patients, and reliance on portal-phase CT images.

Ren et al. [16] analyzed automated gastric cancer diagnosis using histopathological images to address the challenges associated with manual pathological assessment. Histopathological images from the TCGA-STAD, GasHisSDB, and NCT-CRC-HE-100K datasets were preprocessed through stain normalization and data augmentation before model training and evaluation using fivefold cross-validation. Model achieved better performance on the GasHisSDB dataset, while external validation confirmed its robustness across diverse data sources. Limitations included dependence on patch-based analysis, limited representation of real-world data variability, and restricted interpretability of the Transformer component.

A. Research Gap

Although recent studies established the potential of AI and DL for gastric cancer diagnosis, several important challenges remain unresolved. Existing machine learning approaches based on clinical laboratory data and patient records provide effective risk prediction but do not utilize the rich visual information available in endoscopic images, limiting their ability to characterize lesion morphology and appearance variations [10,11]. Histopathological image-based frameworks have achieved promising classification performance; however, their dependence on stain normalization, patch-based analysis, and limited image diversity restricts their generalization to diverse clinical settings [7,16]. Endoscopy-based methods have improved gastric lesion detection and classification, yet many studies rely on single deep learning architectures or focus primarily on lesion detection and segmentation rather than robust multiclass classification [12-14]. Furthermore, several existing frameworks have been validated using relatively small or single-center datasets, reducing their applicability to broader clinical populations [8,12,14,15]. Another common limitation is the insufficient integration of fine-grained local lesion characteristics and long-range contextual information, resulting in incomplete feature representation for distinguishing visually similar gastric lesions. These limitations highlight the need for an end-to-end hybrid framework capable of effectively combining complementary local and global feature representations while providing

accurate and reliable multiclass gastric cancer classification from endoscopic images.

III. MATERIALS AND METHODS

To enhance data quality and model generalization, the gastric endoscopic image dataset is first subjected to preprocessing techniques, specifically image resizing, normalization, and data augmentation. The preprocessed images are then simultaneously processed by two parallel feature extraction branches: ConvNeXt V2-Tiny, which learns discriminative local visual features, and PVTv2-B2, which captures global contextual representations. Extracted features from two branches are integrated over the proposed GLFF module to generate a unified and informative feature representation. The classification of the input endoscopic images into Normal, Lymphoma, or Adenocarcinoma classes is achieved by passing the fused features through a Global Average Pooling (GAP) layer, a Fully Connected (FC) layer, and a final Softmax layer, as presented in Fig. 1.

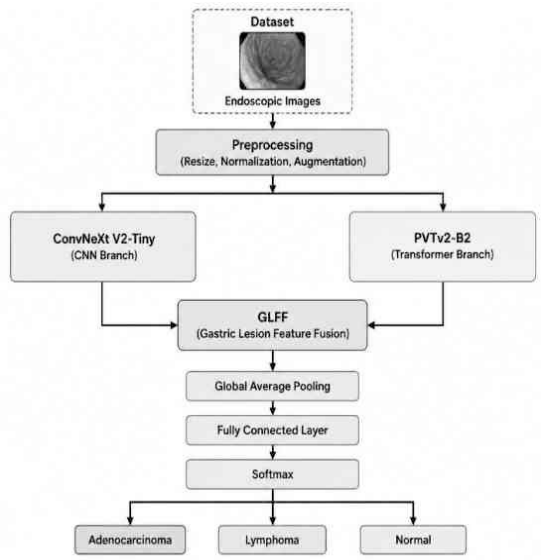


Fig.1. Block diagram of proposed model

A. Dataset Description

For this study, RGB endoscopic images were gathered from the Fujita Health University Hospital database [17]. A uniform resolution of 256 × 256 pixels was maintained across the dataset to guarantee consistency during both image preprocessing and subsequent model development. The images were acquired from different subjects during routine endoscopic examinations and comprise three diagnostic categories: Adenocarcinoma, Lymphoma, and Normal gastric tissues. Specifically, the dataset contains 2,750 Adenocarcinoma images, 2,880 Lymphoma images, and 2,786 Normal images, resulting in a total of 8,416 endoscopic images. The balanced distribution of samples across the three classes supports reliable model training and evaluation while enabling effective learning of discriminative visual features for automated gastric cancer classification. Fig.2. illustrates representative gastric endoscopy images from the three diagnostic categories used in this study: (a) Adenocarcinoma, (b) Lymphoma, and (c) Normal gastric tissue.

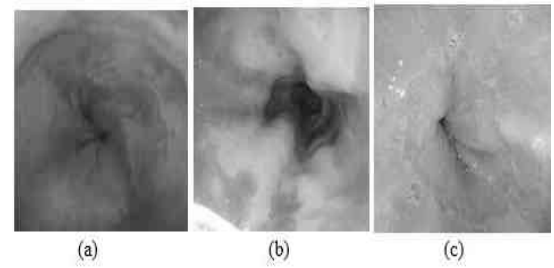


Fig.2. Sample gastric endoscopy images of (a) Adenocarcinoma, (b) Lymphoma, and (c) Normal classes

B. Preprocessing and Augmentation

The gastric endoscopic image dataset was pre-processed to ensure uniformity and improve the learning capability of the proposed GLFF-Net framework. Initially, all endoscopic images were resized to a fixed resolution of 256 × 256 pixels, providing a consistent input size for both the ConvNeXt V2-Tiny and PVTv2-B2 backbones. Subsequently, pixel intensity values were normalized to the range of [0, 1] to reduce variations caused by different imaging conditions and facilitate stable model convergence. The normalization process is expressed as in Eq. (1)

$$I_{norm} = \frac{I - I_{min}}{I_{max} - I_{min}} \quad (1)$$

In this formulation, I signifies the raw pixel intensity, I_{min} and I_{max} are the boundary pixel values, and I_{norm} represents the resulting normalized intensity. The pre-processed data was partitioned via stratified sampling into an 80% training set, a 10% validation set, and a 10% testing set to keep class distributions consistent. While the training data drove the learning process, the validation data guided hyperparameter optimization to curb overfitting. The testing set remained completely unseen until the final evaluation. This approach introduced structural variety while keeping the clinical traits of the gastric lesions intact. Specifically, the pipeline introduced random horizontal flipping, rotations up to $\pm 20^{\circ}$ random zooming, spatial translations (width/height shifts), and brightness tweaks.

C. Model Development

The proposed GLFF-Net combines ConvNeXt V2-Tiny and PVTv2-B2 to extract complementary local and global features from gastric endoscopic images. These features are

integrated using the proposed Gastric Lesion Feature Fusion (GLFF) module and subsequently classified into three diagnostic categories.

a. ConvNeXt V2-Tiny for Local Feature Extraction

The first branch of the GLFF-Net employs ConvNeXt V2-Tiny to extract discriminative local features from gastric endoscopic images [18]. Owing to its convolution-based architecture, ConvNeXt V2-Tiny effectively captures fine-grained lesion characteristics, including edges, textures, and morphological patterns, which are essential for differentiating adenocarcinoma, lymphoma, and normal gastric tissues. The input image $X \in \mathbb{R}^{H\times W\times C}$, the network first applies a 4 × 4 stem convolution to generate the initial feature representation. This operation is mathematically expressed in Eq. (2).

$$F_0 = Conv(X) \quad (2)$$

where Con*v*(.) denotes the stem convolution operation and F_0 represents the initial feature map. The generated feature map is subsequently propagated through four hierarchical ConvNeXt V2 stages. Each stage consists of a sequence of depthwise convolution, Layer Normalization (LN), pointwise linear transformations, GELU activation, GRN, and a residual connection, enabling efficient local feature learning. The transformation performed by a ConvNeXt V2 block is represented in Eq. (3).

$$F_b = GRN\left(W_2\,\delta\left(W_1\left(LN\left(DWConv(F)\right)\right)\right)\right) + F \quad (3)$$

where *DW Conv(.)* denotes the depthwise convolution, *LN*(.) represents Layer Normalization, W_1 and W_2 are learnable pointwise transformation matrices, $\delta(\cdot)$ denotes the GELU activation function, and *GRN* (.) represents the GRN operation. The residual connection preserves the original feature information while enhancing the discriminative capability of the extracted features. Following feature refinement, the output is progressively processed through four successive ConvNeXt V2 stages to capture hierarchical local representations. This multi-stage feature extraction process is expressed in Eq. (4).

$$F_{CNN} = Stage_4\left(Stage_3\left(Stage_2\left(Stage_1(F_0)\right)\right)\right) \quad (4)$$

where $Stage_1$- $Stage_2$ denote the four hierarchical stages of the ConvNeXt V2-Tiny backbone. Finally, the local feature representation extracted by the ConvNeXt V2-Tiny branch is obtained as shown in Eq. (5).

$$F_{CNN} = ConvNeXtV2(X) \quad (5)$$

where F_{CNN} signifies the final local feature representation of the input gastric endoscopic image. These features are subsequently forwarded to the proposed Gastric Lesion Feature Fusion (GLFF) module, where they are integrated with the global contextual features extracted by the PVTv2-B2 branch. Fig. 3. presents the architecture of the ConvNeXt V2-Tiny branch for local feature extraction.

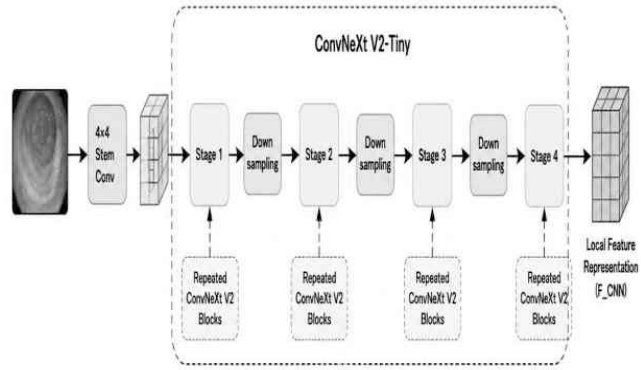


Fig. 3. Architecture of the ConvNeXt V2-Tiny branch

b. PVTv2-B2 for Global Context Learning

The second branch of the proposed GLFF-Net utilizes PVTv2-B2 to capture global contextual information from gastric endoscopic images [19]. Unlike conventional convolutional networks, PVTv2-B2 employs a hierarchical vision transformer architecture that models long-range spatial dependencies, enabling the identification of global lesion structures and contextual relationships that complement the local features extracted by ConvNeXt V2-Tiny. The input image, the image is first partitioned into overlapping patches through the Overlapping Patch Embedding (OPE) layer. This operation is expressed in Eq. (6).

$$F_0 = OPE(X) \quad (6)$$

where *OPE* (.) denotes the overlapping patch embedding operation and F_0 represents the embedded feature tokens. The embedded tokens are then processed through four hierarchical transformer stages. Each stage consists of Spatial Reduction Attention (SRA) followed by a Feed-Forward Network (FFN) with residual connections, enabling efficient modeling of global contextual information while reducing computational complexity. The feature transformation within a transformer block is represented in Eq. (7).

$$F_b = FFN\left(SRA(F)\right) + F \quad (7)$$

where SRA (.) denotes the Spatial Reduction Attention mechanism, *FFN (.)* represents the feed-forward network, and the residual connection preserves the learned contextual information. The hierarchical global feature extraction performed across the four transformer stages is expressed in Eq. (8).

$$F_{PVT} = Stage_4\left(Stage_3\left(Stage_2\left(Stage_1(F_0)\right)\right)\right) \quad (8)$$

where F_{PVT} denotes the global contextual feature representation extracted from the input gastric endoscopic image. These features are subsequently forwarded to the proposed GLFF module, where they are integrated with the local features extracted by the ConvNeXt V2-Tiny branch.

c. Gastric Lesion Feature Fusion (GLFF) and Classification

The proposed Gastric Lesion Feature Fusion (GLFF) module integrates the complementary feature representations extracted by the ConvNeXt V2-Tiny and PVTv2-B2 branches. While ConvNeXt V2-Tiny captures fine-grained local lesion characteristics, PVTv2-B2 learns global contextual information. The GLFF module effectively combines these heterogeneous features to generate a unified representation for accurate gastric cancer classification. Let F_{CNN} and F_{PVT} denote the local and global feature representations extracted by the ConvNeXt V2-Tiny and PVTv2-B2 branches, respectively. To ensure feature compatibility, both feature maps are first aligned to a common channel dimension using a 1x1 convolution. The channel alignment operation is expressed in Eq. (10).

$$F_A^{CNN} = Conv_{1\times1}(F_{CNN}),\ F_A^{PVT} = Conv_{1\times1}(F_{PVT}) \quad (10)$$

where $Conv_{1x1}(.)$ denotes the channel alignment operation. The fusion process is represented in Eq. (11).

$$F_F = Concat(F_A^{CNN}, F_A^{PVT}) \quad (11)$$

where *Concat* (.) denotes channel-wise concatenation. To emphasize informative lesion features while suppressing redundant responses, adaptive channel weighting is applied to the fused representation. The refined feature map is obtained as shown in Eq. (12),

$$F_W = F_F \odot \sigma\left(W(F_F)\right) \quad (12)$$

where $\sigma(\cdot)$ denotes the sigmoid activation function, *W(.)* represents the learnable channel weighting operation. The weighted features are further refined using a 3x3 convolution to enhance feature interactions and generate the final fused representation. This operation is expressed in Eq. (13).

$$F_{GLFF} = Conv_{3\times3}(F_W) \quad (13)$$

where F_{GLFF} signifies the final fused feature representation. The fused feature map is then transformed into a compact feature vector using Global Average Pooling (GAP), as represented in Eq. (14).

$$F_{GAP} = GAP(F_{GLFF}) \quad (14)$$

where F_{GAP} is the pooled feature vector used for classification. The pooled feature vector is then mapped through a FC layer to yield the class scores given in Eq. (15).

The learnable weights and biases parameters of the fully

$$Z = W_{fc}F_{GAP} + b \quad (15)$$

connected layer are denoted by *W and b,* respectively. Finally, the probability of each gastric disease class is obtained using the Softmax classifier, as expressed in Eq. (16).

$$P(y_i) = \frac{\exp(z_i)}{\sum_{j=1}^{3}\exp(z_j)} \quad (16)$$

where $P(y_i)$ signifies the predicted probability of the i^{th} class corresponding to adenocarcinoma, lymphoma, or normal gastric tissue. The final prediction is determined by selecting the class that yields the highest probability score. Fig. 4. presents the architecture of the proposed GLFF module for feature fusion and gastric cancer classification.

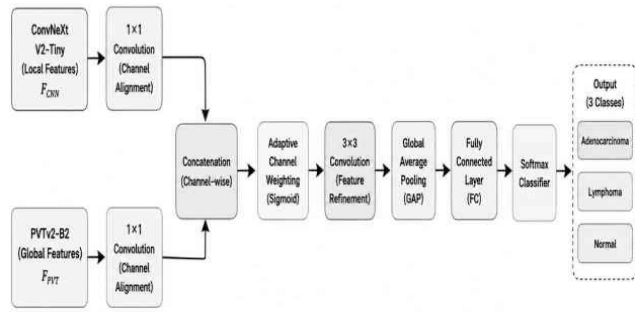


Fig. 4. Architecture of the proposed GLFF module

D. Simulation Setup

The proposed GLFF-Net framework was implemented and evaluated using a high-performance computing environment. Model training and testing were performed on a system prepared with 32 GB RAM, an Intel Core i7 processor, and an NVIDIA GeForce GPU. The implementation was developed using the TensorFlow and Keras DL frameworks, with Google Colaboratory GPU acceleration utilized to reduce training time. The Adam optimizer was employed to optimize the network parameters, while the categorical cross-entropy loss function was adopted for three-class gastric cancer classification. The optimized hyperparameter settings used for training the proposed GLFF-Net are given in Table 1.

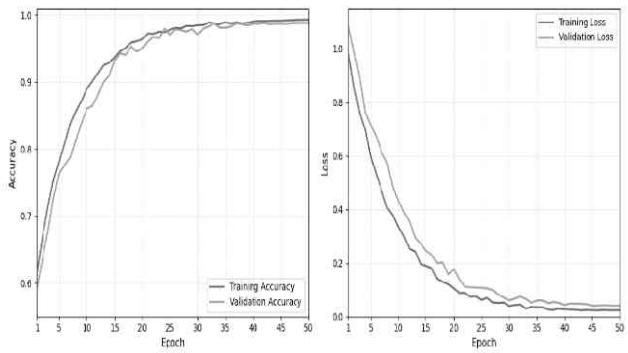
Table 1. Hyperparameters of the GLFF-Net model

Parameter	Value
Epochs	50
Learning Rate	0.0001
Loss Function	Categorical Cross-Entropy
Optimizer	Adam
Batch Size	32
Activation Function	Softmax

IV. RESULTS AND DISCUSSION

The training and validation accuracy curves demonstrate the effective convergence of the proposed GLFF-Net model over 50 training epochs, as in Fig. 5(a). Both curves exhibit a rapid increase during the initial epochs, representing efficient feature learning from gastric endoscopic images. As training progresses, the accuracy gradually stabilizes with only minor fluctuations, reflecting good generalization capability and minimal overfitting. The framework attains a final training accuracy of 99.3% and a validation accuracy of 98.8%, confirming the robustness of the proposed hybrid ConvNeXt V2-Tiny–PVTv2-B2 framework for gastric cancer classification. The training and validation loss curves exhibit a consistent decrease throughout the training process, as presented in Fig. 5(b), indicating stable optimization of the proposed GLFF-Net model.

A steep reduction in loss is detected during the early epochs, followed by gradual convergence as the model approaches the optimal solution. The close agreement among the training and validation loss curves demonstrates effective learning without significant overfitting. At the end of 50 epochs, the framework attains a training loss of 0.024 and a validation loss of 0.041, highlighting the reliability and convergence of the proposed framework for multiclass gastric cancer classification.



(a) (b)

Fig.5. Performance curves for model training (a) Accuracy plot (b) Loss plot

To estimate the performance of the proposed GLFF-Net model, standard classification metrics were computed from confusion matrix parameters. These metrics offer a comprehensive assessment of the framework's ability to accurately classify gastric endoscopic images into adenocarcinoma, lymphoma, and normal categories. The mathematical expressions used to compute the evaluation metrics are presented in Eqs. (17)-(20).The quantitative performance of the GLFF-Net framework is summarized in

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (17)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (18)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (19)$$

$$\text{F1-score} = \frac{2\times TP}{2\times TP+FP+FN} \quad (20)$$

Table 2. The suggested model attained an overall classification accuracy of 98.80%, demonstrating its capability to accurately identify gastric cancer categories from endoscopic images. Furthermore, it obtained a precision of 98.72%, a recall of 98.65%, and an F1-score of 98.68%, representing balanced and consistent classification performance across all three classes. In addition, the proposed framework achieved an AUC of 99.70%, highlighting its excellent discriminative capability in distinguishing between adenocarcinoma, lymphoma, and normal gastric endoscopic images.

Table 2. Performance Evaluation of the Proposed GLFF-Net Framework

Performance Metrics	Values (%)
Accuracy	98.80
Precision	98.72
Recall	98.65
F1-score	98.68
AUC	99.70

The confusion matrix of the proposed GLFF-Net framework is presented in Fig. 6. Majority of gastric endoscopic images were correctly classified into their respective diagnostic categories, indicating the high discriminative capability of the proposed model. Among the 275 adenocarcinoma samples, 272 were correctly identified, with only three images misclassified as lymphoma or normal. Similarly, the model accurately classified 284 out of 288 lymphoma images, while only four samples were assigned to other categories. For the 279 normal images, 276 were correctly recognized, with three samples misclassified as pathological cases. The limited number of misclassifications demonstrates that the proposed GLFF-Net effectively captures both local lesion characteristics and global contextual information, resulting in robust multiclass gastric cancer classification with an overall accuracy of 98.80%.

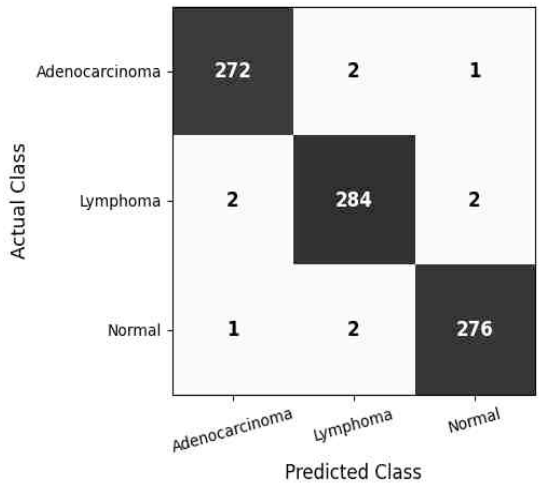


Fig.6. Confusion matrix of the proposed GLFF-Net framework

The ROC analysis demonstrates the excellent discriminative capability of the proposed model for classifying adenocarcinoma, lymphoma, and normal gastric endoscopic images. The proposed framework achieved AUC values of 0.996 for adenocarcinoma, 0.997 for lymphoma, and 0.998 for the normal class, while the micro-average AUC reached 0.997. These results confirm the efficiency and robustness of the GLFF-Net framework for multiclass gastric cancer diagnosis. The ROC curves tracking the performance of the proposed GLFF-Net framework are displayed in Fig. 7.

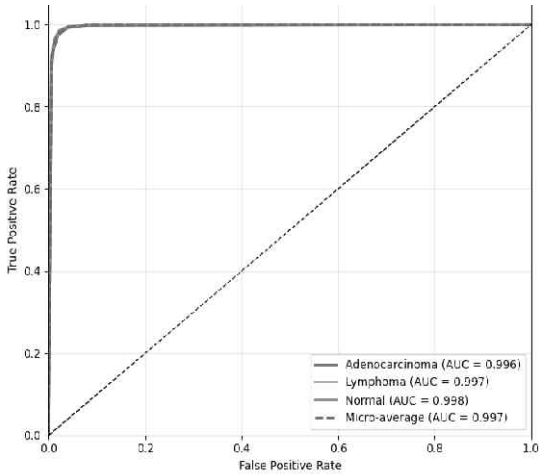


Fig.7. ROC curves of the proposed GLFF-Net framework

The experimental outcomes validate that the proposed GLFF-Net framework provides consistent and accurate multiclass gastric cancer classification from endoscopic images. The integration of ConvNeXt V2-Tiny and PVTv2-B2 enables the network to learn complementary local lesion characteristics and global contextual information, while the Gastric Lesion Feature Fusion (GLFF) module effectively combines these representations into a unified feature space for classification. This design contributes to improved discrimination among adenocarcinoma, lymphoma, and normal images, resulting in an overall accuracy of 98.80%, precision of 98.72%, recall of 98.65%, F1-score of 98.68%, and an AUC of 99.70%. The confusion matrix and ROC analysis further indicate that the approach maintains a low number of misclassifications across all diagnostic categories. When compared with existing deep learning approaches for gastric cancer classification, the proposed GLFF-Net achieves improved classification performance by combining convolutional feature extraction with transformer-based contextual learning through an effective feature fusion strategy. The consistent performance observed across all evaluation metrics demonstrates that the proposed framework provides a reliable solution for automated gastric cancer classification from endoscopic images and represents an effective advancement over existing classification methods.

V. CONCLUSION

This study presented GLFF-Net, a hybrid DL framework for multiclass gastric cancer classification from endoscopic images by integrating ConvNeXt V2-Tiny and PVTv2-B2 with the proposed Gastric Lesion Feature Fusion (GLFF) module. The framework was designed to combine detailed local lesion characteristics with global contextual information, enabling effective discrimination among adenocarcinoma, lymphoma, and normal gastric endoscopic images. Experimental evaluation on the endoscopic dataset demonstrated the effectiveness of the proposed approach, attaining an 98.80% accuracy, 98.72%, precision, 98.65% recall, 98.68% F1-score and an AUC of 99.70%. The confusion matrix and ROC analysis further confirmed consistent classification performance across all diagnostic categories. Comparative analysis showed that the proposed GLFF-Net achieved higher classification performance than existing deep learning approaches by effectively integrating convolutional and transformer-based feature representations through the proposed fusion strategy. These findings demonstrate that GLFF-Net provides an accurate and reliable framework for automated gastric cancer classification from endoscopic images. Future work will focus on confirming the proposed framework on larger multicenter clinical datasets and extending the approach to gastric lesion localization and severity assessment to further enhance its clinical applicability.

REFERENCES

[1] Van Cutsem, E., Sagaert, X., Topal, B., Haustermans, K., & Prenen, H. (2016). Gastric cancer. The Lancet, 388 (10060), 2654-2664.

[2] Thrift, A. P., Wenker, T. N., & El-Serag, H. B. (2023). Global burden of gastric cancer: epidemiological trends, risk factors, screening and prevention. Nature reviews Clinical oncology, 20(5), 338-349.

[3] Jiang, Y., Xie, J., Han, Z., Liu, W., Xi, S., Huang, L., ... & Li, G. (2018). Immunomarker support vector machine classifier for prediction of gastric cancer survival and adjuvant chemotherapeutic benefit. Clinical cancer research, 24(22), 5574-5584.

[4] Chen, Y., Wang, B., Zhao, Y., Shao, X., Wang, M., Ma, F., ... & Hu, Z. (2024). Metabolomic machine learning predictor for diagnosis and prognosis of gastric cancer. Nature communications, 15(1), 1657.

[5] Wang, X., Chen, Y., Gao, Y., Zhang, H., Guan, Z., Dong, Z., ... & Yu, G. (2021). Predicting gastric cancer outcome from resected lymph node histopathology images using deep learning. Nature communications, 12(1), 1637.

[6] Niu, P. H., Zhao, L. L., Wu, H. L., Zhao, D. B., & Chen, Y. T. (2020). Artificial intelligence in gastric cancer: Application and future perspectives. World journal of gastroenterology, 26(36), 5408.

[7] Khayatian, D., Maleki, A., Nasiri, H., & Dorrigiv, M. (2025). Histopathology image analysis for gastric cancer detection: a hybrid deep learning and catboost approach. Multimedia Tools and Applications, 84(19), 21777-21803.

[8] Noh, A., Quek, S. X. Z., Zailani, N., Wee, J. S., Yong, D., Ahn, B. Y., ... & Chung, H. (2025). Machine learning classification and biochemical characteristics in the real-time diagnosis of gastric adenocarcinoma using Raman spectroscopy. Scientific Reports, 15(1), 2469.

[9] Rubab, S., Jamshed, M., Khan, M. A., Almujally, N. A., Damaševičius, R., Hussain, A., ... & Nam, Y. (2025). Gastrointestinal tract disease classification from wireless capsule endoscopy images based on deep learning information fusion and Newton Raphson controlled marine predator algorithm. Scientific Reports, 15(1), 32180.

[10. Ke, X., Cai, X., Bian, B., Shen, Y., Zhou, Y., Liu, W., ... & Yang, J. (2024). Predicting early gastric cancer risk using machine learning: A population-based retrospective study. Digital Health, 10, 20552076241240905.

[11. Tian, H., Liu, Z., Liu, J., Zong, Z., Chen, Y., Zhang, Z., & Li, H. (2023). Application of machine learning algorithm in predicting distant metastasis of T1 gastric cancer. Scientific Reports, 13(1), 5741.

12. Park, I., Roh, J., Son, D., Noh, C. K., & Yoon, J. (2025). Artificial intelligence-based gastric cancer detection in the gastric submucosal dissection method via hyperspectral imaging. Sensors and Actuators B: Chemical, 435, 137630.

13. Zhang, K., Wang, H., Cheng, Y., Liu, H., Gong, Q., Zeng, Q., ... & Chen, D. (2024). Early gastric cancer detection and lesion segmentation based on deep learning and gastroscopic images. Scientific Reports, 14(1), 7847.

14. Chae, J. W., & Cho, H. C. (2023). Enhanced classification of gastric lesions and early gastric cancer diagnosis in gastroscopy using multi-filter autoaugment. IEEE Access, 11, 29391-29399.

15. Gao, Z., Yu, Z., Zhang, X., Chen, C., Pan, Z., Chen, X., ... & Shen, X. (2023). Development of a deep learning model for early gastric cancer diagnosis using preoperative computed tomography images. Frontiers in Oncology, 13, 1265366.

16. Ren, T., Govindarajan, V., Bourouis, S., Wang, X., & Ke, S. (2025). An interpretable hybrid deep learning framework for gastric cancer diagnosis using histopathological imaging. Scientific Reports, 15(1), 34204.

17. Teramoto, A., Shibata, T., Yamada, H., Hirooka, Y., Saito, K., & Fujita, H. (2021). Automated detection of gastric cancer by retrospective endoscopic image dataset using U-Net R-CNN. Applied Sciences, 11(23), 11275.

18. Woo, S., Debnath, S., Hu, R., Chen, X., Liu, Z., Kweon, I. S., & Xie, S. (2023). Convnext v2: Co-designing and scaling convnets with masked autoencoders. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 16133-16142).

19. Chen, Y., Zhuang, Z., & Chen, C. (2023, May). Object detection method based on pvtv2. In 2023 IEEE 3rd International Conference on Electronic Technology, Communication and Information (ICETCI) (pp. 730-734). IEEE.