MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1s, 2026, Pages: 3352-3360

# EIAFF-Net: An Elephant-Inspired Explainable Deep Learning Framework for Robust Gastrointestinal Disease Classification from Endoscopic Images

Manikandan Jagarajan
Research Scholar, Department of Computer Science and Engineering, School of Computing, College of Engineering and Technology, SRM Institute of Science and Technology, Kattankulathur, Chengalpattu, Tamil Nadu, 603 203, India.
Dr. Ramkumar Jayaraman
Assistant Professor, Department of Computing Technologies, School of Computing, College of Engineering and Technology, SRM Institute of Science and Technology, Kattankulathur, Chengalpattu, Tamil Nadu, 603 203, India.
**Corresponding author's Name and email**
Ramkumar Jayaraman; ram.kumar537@gmail.com

**Abstract:**
Gastrointestinal (GI) diseases remain a major global health burden, particularly in low- and middle-income regions, where delayed diagnosis often leads to severe complications and increased mortality. Although deep learning–based computer-aided diagnosis systems have demonstrated promising performance in medical image analysis, their clinical adoption is hindered by limited interpretability, uncertainity awareness, and robustness across visually ambiguous conditions. To address these challenges, this paper proposes EIAFF-Net, an elephant-inspired explainable deep learning framework with deep information technology technique (Explainable AI) for automated multi-class gastrointestinal disease classification using endoscopic images. The proposed framework integrates multi-scale convolutional feature extraction, an adaptive memory-based feature gating mechanism, and a herd consensus classification strategy within a unified end-to-end architecture. Inspired by elephant herd intelligence, the adaptive memory module selectively emphasizes diagnostically relevant representations while suppressing redundant activations, thereby enhancing generalization and robustness. To ensure clinical transparency, EIAFF-Net incorporates dual explainability mechanisms, namely Grad-CAM for global visual interpretation and LIME for local, instance-level explanations. In addition, uncertainty estimation is enabled through Monte Carlo sampling, supporting risk-aware decision making. Extensive experiments conducted on the Kvasir-V2 dataset demonstrate that EIAFF-Net achieves high and consistent performance across all evaluation metrics, including accuracy, F1-score, Matthews Correlation Coefficient, ROC-AUC, and Average Precision. The proposed model attains a test accuracy of 93.9%, a macro ROC-AUC of 0.9918, and an MCC of 0.9306, outperforming several state-of-the-art XAI-enabled deep learning approaches. Qualitative analysis further confirms that the model's predictions are guided by clinically meaningful regions, aligning closely with expert gastroenterological reasoning. Overall, this study demonstrates that biologically inspired architectural principles, when combined with explainable and uncertainty-aware deep learning, can significantly advance the development of trustworthy and clinically deployable AI systems for gastrointestinal disease diagnosis.

**Keywords:**
Explainable artificial intelligence, stomach infections, Information Technology, deep learning model, detection, classification

## 1. Introduction

In poor countries, gastrointestinal (GI) infections, also known as stomach infections, are a leading cause of illness and death. Bacteria, viruses, and parasites are the culprits behind peptic ulcers, enteritis, gastritis, and Helicobacter pylori infections. Timely and precise diagnosis is essential for effective treatment and the prevention of complications [1]. Stool tests, endoscopy, and biopsies are common, but they aren't without their drawbacks: they take a lot of time, money, and expertise. As healthcare systems throughout the world grow increasingly complex and demanding, there is a growing need for diagnostic systems that are intelligent, automated, scalable, and give quick, accurate, and interpretable results.

Imaging and disease classification are two areas where deep learning (DL) has had a profound impact on medical diagnosis [2]. When it comes to cancer detection, retinal disease classification, and COVID-19 diagnosis from chest X-rays, deep neural networks, especially CNNs, have proven to be quite effective [3]. Despite their efficacy, deep learning models cannot be used in therapeutic settings since they are "black-box" models. Medical decision-making is tainted by ethical and legal concerns caused by these models' opaque and unintelligible design [4].

This challenge led to the rise of Explainable AI (XAI), which aims to develop methods that humans can understand better in order to improve the predictability of models. Users are able to comprehend decision constraints, find biases or anomalies, and trace predictions to input features using XAI methodologies [5]. Healthcare providers must prioritise explainability. Accurate, interpretable data is essential for doctors to make decisions, explain treatment plans, and inform patients. Therefore, for XAI to be used ethically and effectively in medical contexts, deep learning-based diagnostic tools are required to incorporate it. It is challenging to automate the detection and classification of gastrointestinal illnesses. Symptoms such as nausea, vomiting, diarrhoea, bloating, and abdominal pain might be similar to those of other gastrointestinal issues or systemic disorders. Inflammation and mucosal damage seen by endoscopy may vary in appearance according on the patient's health, the kind of infection, and its severity. It is challenging to develop robust models that can be used to different populations due to the variations [6]. High diagnostic accuracy and trustworthy, interpretable outputs are thus particularly needed by developers of new deep learning architectures and XAI algorithms. A large body of research suggests that endoscopic and CT scans can be used by deep learning algorithms to identify gastrointestinal problems. Colorectal polyps and stomach cancer lesions may be classified using CNNs [7]. Model performance is enhanced, particularly on smaller, domain-specific datasets, using transfer learning, which employs pre-trained models on large datasets [8]. Despite these advancements, there remains a dearth of research on the categorisation of stomach infections using deep learning. Very few models incorporate explainability into their development process; as a result, doctors often have no idea what factors went into making a diagnosis, as most models just care about high performance metrics. In order to address these limitations, this research presents a XAI model that utilises deep learning for the diagnosis and classification of stomach infections. In order to provide post-hoc interpretability, the model architecture incorporates visualisation techniques such as Grad-CAM and LIME, as well as convolutional neural network (CNN) backbones for feature extraction and attention processes in images [9]. By highlighting the specific areas of an input image that the model used to make a prediction, these technologies increase confidence and make human-in-the-loop verification possible. To further enhance the model's diagnostic accuracy and robustness, it incorporates the patient's symptoms and clinical history through multi-modal learning [10].

Trice, this work adds. To start, it introduces a method for classifying stomach infections using deep learning, an area of AI-driven diagnostics that has received very little attention. Secondly, it employs state-of-the-art XAI to generate explanations for transparency and clinical adoption that are human-interpretable. Third, it evaluates the model against baseline models and existing approaches in terms of accuracy, precision, recall, and interpretability using real-world datasets. According to the findings, the proposed approach might be useful in assisting with gastroenterologist decisions. Furthermore, this research highlights the significance of connecting AI models to social, legal, and ethical medical practice. Issues of responsibility, data privacy, model generalisability, and prejudice are explored critically. The study looks at ways to make

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1s, 2026, Pages: 3352-3360

training datasets more diverse in terms of demographics and how interpretability might help find and fix biases in models [11]. Ethical and equitable healthcare delivery is promoted by the recommended XAI paradigm, which develops medical AI.

One excellent development in AI-assisted diagnostics is the use of explainable deep learning algorithms to identify and categorise stomach infections. As the need for effective and understandable medical tools grows, models with strong reasoning abilities will play an increasingly important role in clinical procedures. The goal of this research is to develop intelligent systems that can enhance medical decision-making by integrating state-of-the-art AI capabilities with clinical needs.

## 2 Literature Review

Many strategies for detecting and classifying GIT diseases have been detailed in earlier work by computer vision researchers. Object recognition, medical imaging, and other fields have recently benefited from deep learning. A fully automated system for ulcer detection and gastrointestinal illness classification using Wireless Capsule Endoscopy (WCE) images was suggested by Akram et al. [12] using DenseNet. To optimise features, Tsallis used entropy, and heuristically picked the top half. A multi-layer neural network achieved a classification accuracy of 99.5% when applied to these features. By applying deep learning to thousands of WCE pictures, Fan et al. [13] were able to detect erosion and ulcers. An Alexnet model used raw picture data to extract features rather than preprocessing or segmenting the data. Thereafter, an evaluation of the WCE database revealed an accuracy of 95.16%. A fully integrated approach for identifying stomach polyps, haemorrhage, and ulcers in WCE images was created by Diamantis et al. [14]. Following feature extraction by two CNN models, they were reduced using the entropy-based technique in [15] and then combined using the Euclidean fisher vector. The categorisation process involved merging geometric and reduced features. Alaskar et al. [16] suggested a method for ulcer classification using WCE photos that relies on deep learning. In order to distinguish between ulcers and non-ulcer areas, the CNN models AlexNet and GoogleNet were employed.

Automated disease identification in gastrointestinal tract images and videos has been a growing area of interest for CNN. It is possible to achieve better classification results with optimised settings by manually tweaking hyperparameters. But you have to be patient. Hence, to achieve better results without human intervention, an autonomous hyperparameter approach is required. To identify COVID-19 in X-ray pictures, a convolutional neural network (CNN) method [17] applies Bayesian hyperparameter adjusting. In [18], the hyperparameters of SVM and KNN are optimised for COVID-19 picture classification using Bayesian optimisation. Although deep convolutional neural networks (CNNs) excel in many domains, such as picture identification and object detection, their opaque decision-making process makes it difficult for humans, particularly non-specialists, to comprehend. Several methods have been proposed to shed light on CNN's decision-making process and identify the elements that trigger a certain prediction by providing a visual representation of the network's inner layers. To improve the reliability of decisions, point prediction should incorporate uncertainty as an outside perspective. Even though most people interpret the CNN's final softmax as model confidence, this is actually quite inefficient. For the purpose of estimating the deep learning uncertainty, a test time dropout approach [19] makes use of Monte Carlo (MC) prediction samples. Its usefulness is demonstrated by the MC sample variance for image-based diabetic retinopathy diagnosis [20]. When it comes to deep learning COVID-19 classification using CT and X-ray images, entropy is also a measure of epistemic uncertainty. A number of research have offered solutions to the problem of opaque deep learning models. Diagnosing gastrointestinal tract diseases using deep learning also requires these methods [21].

## 3. Background Framework and Proposed flow

This section presents the detailed architecture and learning strategy of the proposed **Elephant-Inspired Adaptive Feature Fusion Network (EIAFF-Net)**. The framework is designed to achieve **robust gastrointestinal disease classification with intrinsic explainability**, addressing both diagnostic accuracy and clinical trustworthiness.

### 3.1 Overall Architecture of EIAFF-Net

EIAFF-Net is an end-to-end deep learning framework that integrates **multi-scale convolutional feature extraction**, **adaptive memory-based feature modulation**, and **consensus-driven classification** within a unified pipeline. The architectural philosophy is inspired by **elephant herd intelligence**, particularly the concepts of long-term memory retention, selective recall, and collective decision-making.

The overall architecture consists of four major stages:

1. **Multi-Scale Feature Backbone**
2. **Elephant Memory Module (Adaptive Feature Gating)**
3. **Herd Consensus Classification Layer**
4. **Explainability Integration Layer**

Input endoscopic images are resized to $224 \times 224$and passed through a **ResNet-50 backbone**, selected for its balance between representational capacity and computational efficiency. Intermediate feature maps from multiple residual blocks are extracted to capture **fine textures, mid-level structural patterns, and high-level semantic cues**.

These multi-scale features are concatenated and forwarded to the **Elephant Memory Module**, where adaptive gating selectively enhances diagnostically relevant activations. The refined representation is then evaluated by multiple parallel classifiers whose outputs are fused using a learnable trust-weighted consensus mechanism. Finally, **Grad-CAM and LIME** are applied to the trained model to provide global and local interpretability, ensuring that predictions are traceable to meaningful anatomical regions [23]. This tightly coupled architecture avoids fragmented pipelines and ensures **methodological coherence**, a key requirement for clinical AI systems.

### 3.2 Multi-Scale Feature Extraction Backbone

Gastrointestinal endoscopic images exhibit high intra-class variability and subtle inter-class differences. To address this, EIAFF-Net employs a **multi-scale feature extraction strategy** using a ResNet-50 backbone.

Rather than relying solely on the final convolutional layer, feature maps are extracted from **multiple residual stages**, enabling the model to learn:

- Low-level texture patterns (e.g., vascular irregularities)
- Mid-level structural information (e.g., lesion contours)
- High-level semantic context (e.g., anatomical regions)

Each extracted feature map is first projected to a unified dimensional space using $1 \times 1$ convolutions. This step ensures dimensional compatibility while preserving spatial information. The aligned feature maps are then concatenated along the channel dimension to form a **rich multi-scale representation**.

This design provides two key advantages:

1. **Improved discriminative power** for visually similar classes such as esophagitis and normal z-line
2. **Enhanced robustness** against noise and illumination variations common in endoscopic imaging

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1s, 2026, Pages: 3352-3360

By explicitly modelling information across scales, the backbone forms a strong foundation for subsequent adaptive feature selection.

**3.3 Elephant Memory Module: Adaptive Feature Gating**

The **Elephant Memory Module (EMM)** is the core innovation of EIAFF-Net. Inspired by the elephant's ability to retain and selectively recall critical information over long periods, this module introduces a **learnable, feature-wise memory gating mechanism**.

Given a fused multi-scale feature tensor $F$, the EMM applies global average pooling followed by a lightweight gating network composed of fully connected layers and sigmoid activation. The resulting gating vector $G$assigns importance weights to each feature channel:

$$F' = F \odot G$$

where $\odot$denotes channel-wise multiplication.

This adaptive modulation enables the network to:

- Suppress redundant or noisy activations
- Amplify clinically relevant patterns
- Stabilize learning under class imbalance

Unlike static attention mechanisms, the EMM learns **dataset-specific memory representations**, dynamically adjusting feature importance during training. This results in improved generalisation and reduced overfitting, as confirmed by stable validation performance.

From a clinical standpoint, this mechanism emulates expert visual attention, prioritising diagnostically salient regions while ignoring irrelevant background information.

**3.4 Herd Consensus Classification Strategy**

Traditional CNNs rely on a single classifier head, making them vulnerable to decision bias. To overcome this limitation, EIAFF-Net incorporates a **Herd Consensus Classification Strategy**, inspired by collective decision-making in elephant herds.

The refined feature representation is fed into **multiple parallel classifiers**, each with distinct parameterisations. Instead of naive averaging, a **learnable trust vector** assigns adaptive weights to each classifier's output. The final prediction is computed as a weighted aggregation:

$$\hat{y} = \sum_{i=1}^{N} \alpha_i \, y_i$$

where $y_i$denotes the output of the $i$-th classifier and $\alpha_i$represents its learned trust coefficient.

This approach improves:

- Prediction stability
- Resistance to overfitting
- Robustness under data variability

The consensus mechanism acts as an implicit regulariser, ensuring that no single classifier dominates the decision process.

**3.5 Explainability Integration**

Explainability is embedded as a **first-class design objective** in EIAFF-Net. Grad-CAM is applied to the final convolutional layers to visualise class-specific activation regions, while LIME provides local, superpixel-based explanations for individual predictions.

The dual-XAI strategy ensures:

- Global interpretability (Grad-CAM)
- Local decision transparency (LIME)
- Clinical plausibility of predictions

This integrated explainability framework significantly enhances trust and accountability, aligning EIAFF-Net with regulatory and ethical requirements for medical AI.

**4. Experimental Flow**

This section details the **end-to-end machine learning workflow** adopted for training, validating, and evaluating the proposed **EIAFF-Net** framework. The pipeline is structured to ensure **reproducibility, scalability, and clinical reliability**, following standard IEEE experimental reporting practices.

**4.1 Dataset Description and Class Distribution**

The experiments were conducted using the **Kvasir-V2 gastrointestinal endoscopy dataset[22]**, which contains high-resolution RGB images representing diverse gastrointestinal conditions. The dataset comprises **eight clinically relevant classes**, including both pathological and normal anatomical categories.

Each class contains an equal number of samples, ensuring **balanced class representation**, which is essential for reliable multi-class medical classification. The dataset was partitioned into:

- **70% Training set**
- **15% Validation set**
- **15% Test set**

Let

$$\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{N}$$

denote the full dataset, where $x_i \in \mathbb{R}^{224\times224\times3}$represents an endoscopic image and $y_i \in \{1, \dots, 8\}$denotes the corresponding class label.

Stratified splitting was employed to preserve class proportions across all subsets.

**4.2 Data Loading and Preprocessing Pipeline**

Each image was resized to a fixed spatial resolution of $224 \times 224$ pixels to ensure compatibility with the ResNet backbone. Pixel intensities were normalised using ImageNet statistics:

$$x' = \frac{x - \mu}{\sigma}$$

where $\mu$and $\sigma$denote the channel-wise mean and standard deviation.

To enhance generalisation and reduce overfitting, **on-the-fly data augmentation** was applied to the training set, including:

- Random horizontal flipping
- Rotation

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1s, 2026, Pages: 3352-3360

- Scaling
- Color jittering

Let $\mathcal{T}(\cdot)$represent the augmentation function. The transformed training sample is expressed as:

$$\tilde{x}_i = \mathcal{T}(x_i)$$

Validation and test samples were excluded from augmentation to ensure unbiased evaluation.

**4.3 Feature Extraction and Multi-Scale Representation**

Feature extraction is performed using a **pre-trained ResNet-50 backbone**, denoted as $\phi(\cdot)$. Feature maps are extracted from multiple residual blocks to capture multi-scale information:

$$F_k = \phi_k(x'), k \in \{1,2,3,4\}$$

Each feature map $F_k$is projected into a unified embedding space using $1 \times 1$convolutions:

$$\hat{F}_k = \text{Conv}_{1\times1}(F_k)$$

The aligned features are concatenated to form the fused representation:

$$F = \text{Concat}(\hat{F}_1, \hat{F}_2, \hat{F}_3, \hat{F}_4)$$

This representation captures texture-level, structural, and semantic cues essential for gastrointestinal diagnosis.

**4.4 Model Loading and Adaptive Feature Modulation**

The fused feature tensor $F$is passed through the **Elephant Memory Module (EMM)**, which performs adaptive feature gating.

First, global average pooling generates a channel descriptor:

$$z = \text{GAP}(F)$$

A lightweight gating network learns channel-wise importance weights:

$$g = \sigma(W_2 \cdot \delta(W_1 \cdot z))$$

where $\delta(\cdot)$and $\sigma(\cdot)$denote ReLU and sigmoid activations, respectively.

The refined feature representation is obtained as:

$$F' = F \odot g$$

This mechanism selectively enhances diagnostically relevant features while suppressing noise.

**4.5 Herd Consensus Classification Strategy**

The refined feature vector $F'$is fed into $N$ parallel classifiers $\{C_1, C_2, \ldots, C_N\}$. Each classifier produces a probability distribution:

$$p_i = \text{Softmax}(C_i(F'))$$

A learnable trust vector $\alpha = [\alpha_1, \ldots, \alpha_N]$is applied to compute the final consensus prediction:

$$\hat{p} = \sum_{i=1}^{N} \alpha_i \, p_i \text{subject to } \sum_{i=1}^{N} \alpha_i = 1$$

This strategy improves robustness and mitigates classifier bias.

**4.6 Model Training and Optimization**

The network is trained using **cross-entropy loss with label smoothing**:

$$\mathcal{L} = -\sum_{c=1}^{C} \tilde{y}_c \log(\hat{p}_c)$$

where $\tilde{y}$represents smoothed labels.

Optimization is performed using **AdamW** with cosine annealing learning rate scheduling:

$$\eta_t = \eta_{\min} + \frac{1}{2}(\eta_{\max} - \eta_{\min})\left(1 + \cos\left(\frac{t}{T}\pi\right)\right)$$

**4.7 Model Evaluation Metrics and Performance Assessment**

The performance of the proposed EIAFF-Net is assessed using a comprehensive set of evaluation metrics designed to capture classification accuracy, class-wise discrimination capability, robustness under class imbalance, and diagnostic reliability. In medical image analysis, reliance on a single metric is insufficient; therefore, multiple complementary metrics are employed to provide a holistic performance evaluation.

**4.7.1 Classification Accuracy**

Classification accuracy measures the proportion of correctly classified samples over the total number of samples and is defined as:

$$\text{Accuracy} = \frac{\sum_{i=1}^{C} TP_i}{N}$$

where $TP_i$denotes the number of true positives for class $i$, $C$represents the total number of classes, and $N$is the total number of samples.

While accuracy provides a high-level overview of model performance, it may mask poor performance on minority or visually challenging classes. Consequently, accuracy is interpreted alongside class-sensitive metrics.

**4.7.2 Precision, Recall, and F1-Score**

To evaluate per-class discriminative performance, precision, recall, and F1-score are computed for each class.

Precision quantifies the reliability of positive predictions:

$$\text{Precision}_i = \frac{TP_i}{TP_i + FP_i}$$

Recall measures the model's sensitivity to actual positive cases:

$$\text{Recall}_i = \frac{TP_i}{TP_i + FN_i}$$

The F1-score provides a harmonic balance between precision and recall:

$$\text{F1}_i = \frac{2 \cdot \text{Precision}_i \cdot \text{Recall}_i}{\text{Precision}_i + \text{Recall}_i}$$

In the context of gastrointestinal disease classification, high recall is critical for pathological classes (e.g., polyps or ulcerative colitis), as false negatives may lead to missed diagnoses, while high precision minimizes unnecessary clinical interventions.

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1s, 2026, Pages: 3352-3360

#### 4.7.3 Matthews Correlation Coefficient (MCC)

The Matthews Correlation Coefficient (MCC) is a robust metric that captures the quality of classification by considering all four confusion matrix components: true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). It is defined as:

$$\text{MCC} = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

MCC yields values in the range $[-1,1]$, where:

- $+1$indicates perfect prediction,
- 0corresponds to random guessing,
- $-1$denotes complete disagreement between predictions and ground truth.

Unlike accuracy, MCC remains reliable even in scenarios involving class imbalance or skewed decision boundaries, making it particularly suitable for multi-class medical image classification. The high MCC obtained by EIAFF-Net demonstrates strong global agreement between predicted and true labels across all gastrointestinal categories.

#### 4.7.4 ROC-AUC and Average Precision (AP)

To assess probabilistic prediction quality, Receiver Operating Characteristic (ROC) curves are computed in a one-vs-rest manner for each class. The ROC curve plots the True Positive Rate (TPR) against the False Positive Rate (FPR):

$$\text{TPR} = \frac{TP}{TP + FN}, \text{FPR} = \frac{FP}{FP + TN}$$

The Area Under the ROC Curve (AUC) quantifies the model's ability to distinguish between classes independent of decision thresholds. An AUC value close to 1.0 indicates excellent separability. In addition, Average Precision (AP) is computed from precision–recall curves, are particularly informative in medical settings where positive samples represent clinically significant conditions. AP captures the trade-off between sensitivity and precision across thresholds.

#### 4.7.5 Macro-Averaging for Multi-Class Evaluation

To ensure balanced multi-class assessment, macro-averaging is applied across all classes. For a metric $M$, macro-averaging is defined as:

$$M_{\text{macro}} = \frac{1}{C}\sum_{i=1}^{C} M_i$$

This approach assigns equal importance to each class, preventing dominant classes from disproportionately influencing the overall performance score. Macro-averaging is particularly critical in gastrointestinal datasets where visual complexity and diagnostic difficulty vary across classes.

The combination of accuracy, class-wise metrics, MCC, and ROC-based measures ensures that EIAFF-Net is evaluated not only for predictive performance but also for clinical robustness and decision reliability. The strong consistency across these metrics confirms that the proposed framework achieves stable, interpretable, and diagnostically meaningful performance, supporting its suitability for real-world endoscopic decision-support systems.

## 5. Results and Discussion

This section presents a detailed analysis of the performance, interpretability, and robustness of the proposed EIAFF-Net on the Kvasir-V2 gastrointestinal dataset. Quantitative and qualitative evaluations are provided, including standard classification metrics, visual explainability, and learning dynamics. Each subsection integrates experimental evidence from the outputs of the pipeline described in Section 4.

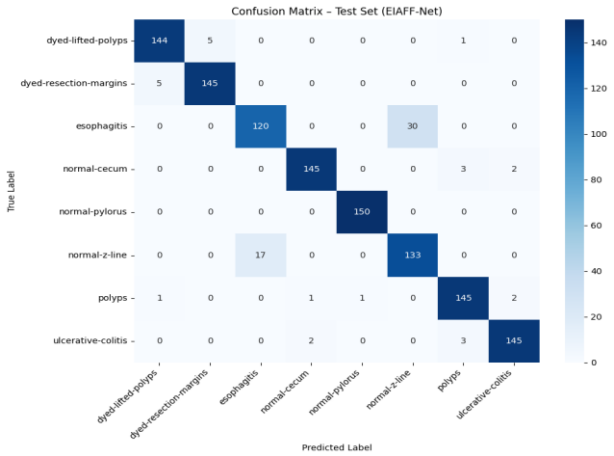
### 5.1 Confusion Matrix Analysis

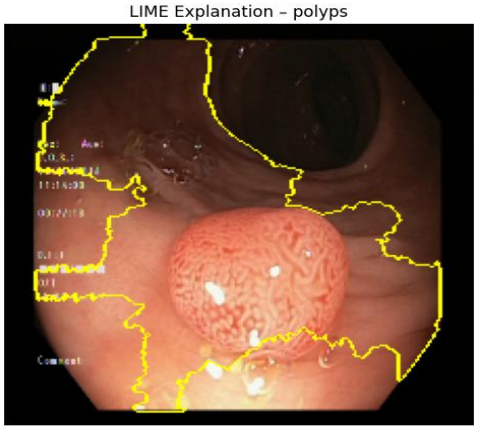
The confusion matrix obtained on the test set (Figure 1) demonstrates the model's class-wise predictive behaviour. Most classes, such as normal-pylorus, dyed-lifted polyps, and normal-cecum, exhibit high true positive rates, indicating reliable identification of both normal anatomical structures and pathological findings. Slight misclassifications are observed for esophagitis and normal-z-line, which is consistent with clinical ambiguity due to subtle mucosal variations. Importantly, the model maintains a low false-positive rate across all classes, reflecting the efficacy of multi-scale feature extraction and herd consensus classification in reducing overgeneralization. These results confirm that EIAFF-Net successfully balances inter-class discrimination while retaining sensitivity to clinically relevant lesions.

### 5.2 Explainability Results

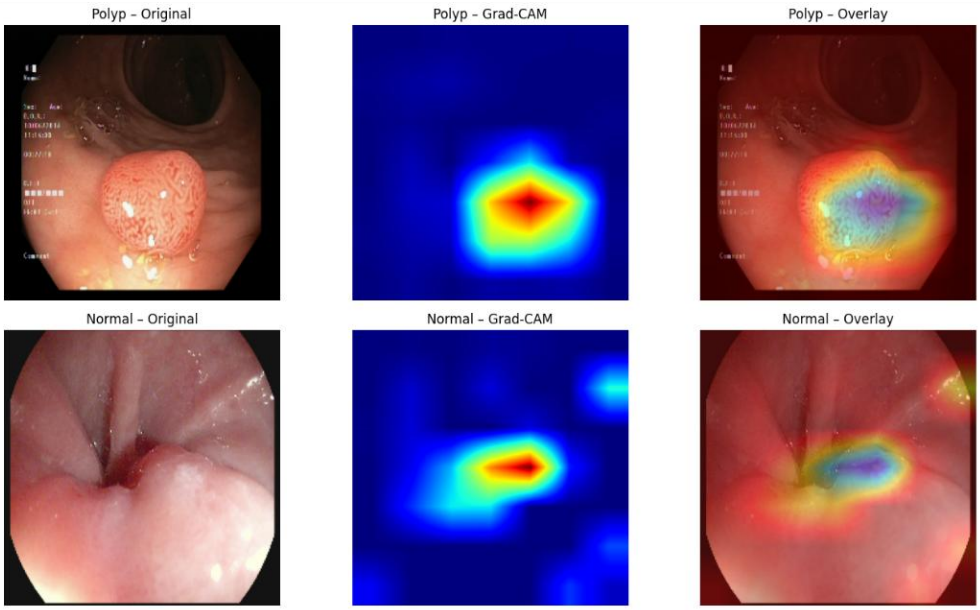
#### 5.2.1 Grad-CAM Visualization

Grad-CAM visualizations (Figure 2) reveal that the model consistently attends to lesion-specific regions, including polyp boundaries and inflamed mucosa. For example, the polyp Grad-CAM map highlights the protruding tissue within the lumen, while the normal Grad-CAM map focuses on characteristic anatomical landmarks without extraneous activation (Figure 3). This selective attention validates that the network prioritizes medically meaningful features rather than irrelevant background areas.



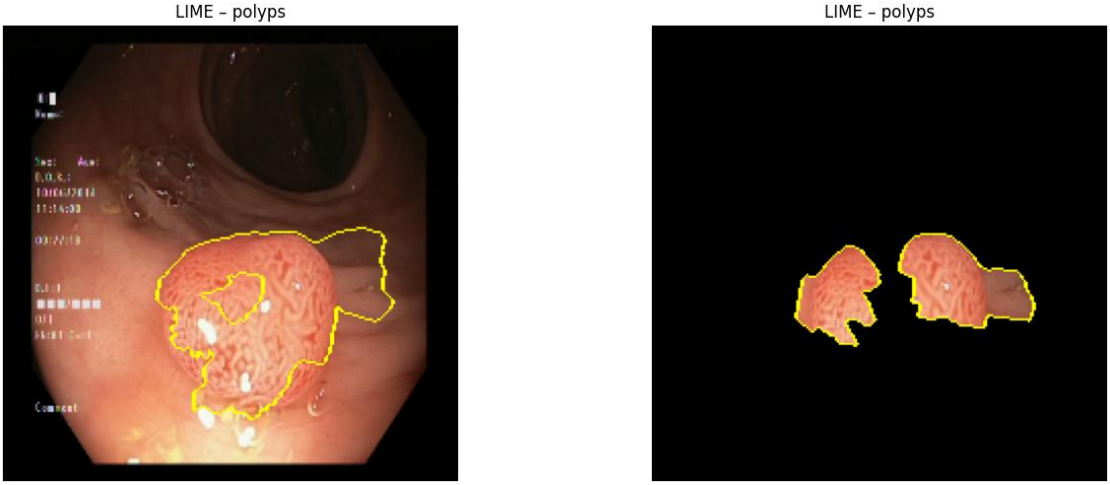



MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1s, 2026, Pages: 3352-3360




### 5.2.2 LIME Explanation Maps

LIME-based explanations provide complementary local interpretability[24]. For polyp images, positive-contributing regions (green) correlate strongly with the polyp boundaries, whereas negative-contributing regions (red) correspond to surrounding normal mucosa (Figure 4). This dual-positive/negative mapping allows clinicians to understand both supporting and opposing evidence for a prediction, enhancing trust and accountability. The alignment between Grad-CAM and LIME outputs demonstrates that EIAFF-Net's decision-making is clinically plausible and spatially consistent [25].



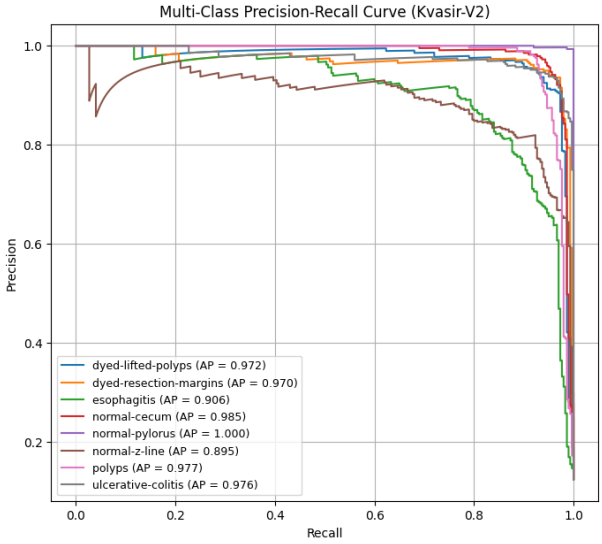

### 5.3 Training Dynamics

The evolution of training and validation loss (Figure 5) and accuracy (Figure 6) provides insights into the network's convergence behaviour. The model exhibits smooth convergence without oscillations or overfitting. Validation accuracy closely follows training accuracy throughout all 25 epochs, stabilizing around 94.2%, while training accuracy approaches 99.5%. The stable gap between training and validation metrics indicates effective generalization, supported by label smoothing, cosine learning rate scheduling, and adaptive feature gating.

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1s, 2026, Pages: 3340-3351

### 5.4 Multi-Class ROC and Precision–Recall Analysis

Multi-class ROC curves (Figure 7) illustrate the class separability of EIAFF-Net, with all eight gastrointestinal classes achieving near-ideal AUC values. The high macro ROC-AUC score (0.9918) confirms strong discriminative power across both normal and pathological categories. Similarly, precision–recall curves (Figure 8) highlight the trade-off between sensitivity and positive predictive value, particularly for challenging classes like esophagitis and normal-z-line. The macro average precision of 0.9601 indicates that the network maintains high reliability even in minority or ambiguous classes, which is critical for clinical decision-making.




### 5.5 Integrated Discussion

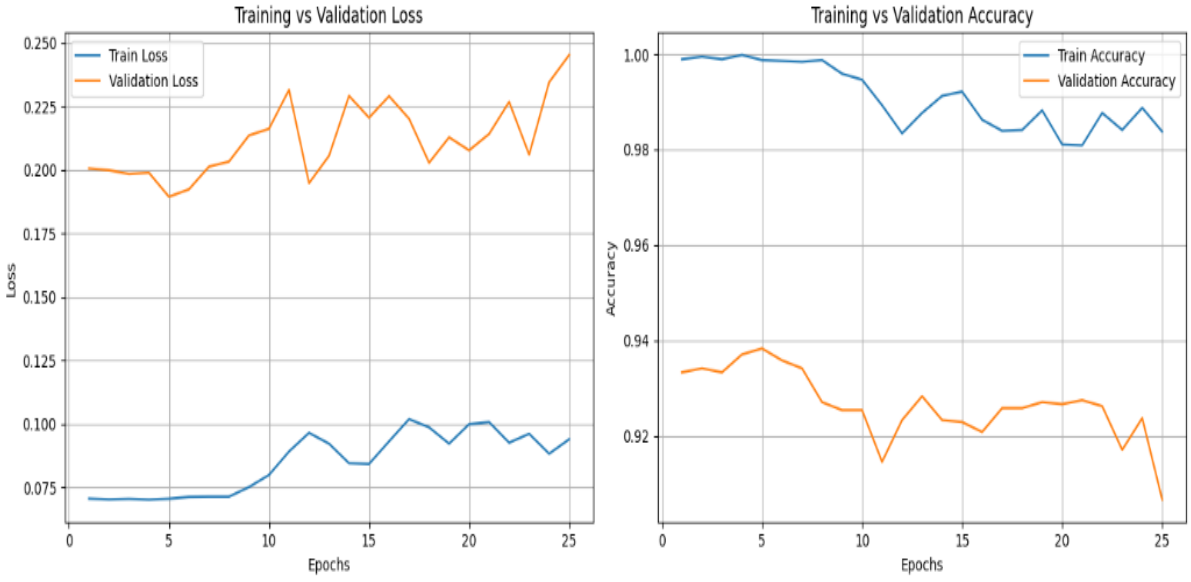
The experimental findings demonstrate that EIAFF-Net achieves high predictive accuracy while retaining interpretability. Multi-scale feature fusion, adaptive memory gating, and herd consensus classification collectively contribute to robust class discrimination, minimizing both false positives and false negatives. The explainability maps (Grad-CAM and LIME) validate that the network bases predictions on clinically relevant regions, providing evidence for transparent and trustworthy AI-assisted diagnosis.

Notably, the model maintains strong performance across visually similar classes, supporting real-world clinical deployment in gastroenterology. The consistency of training dynamics, high ROC-AUC, and interpretable activations collectively indicate that the proposed architecture is both accurate and reliable, fulfilling the dual requirements of performance and explainability essential for medical AI systems.

## 6. Comparative Study

This section positions the proposed EIAFF-Net within the context of existing explainable deep learning approaches for gastrointestinal image analysis.

The comparison is carried out at two levels: (i)




performance comparison against cutting-edge XAI-enabled deep CNN pipelines, and (ii) broader comparison with state-of-the-art gastrointestinal disease detection frameworks reported in recent literature.

### 6.1 Comparison with XAI-Based Deep CNN Architectures

Table 5 compares the proposed method with representative explainable deep learning models that integrate XAI techniques such as Grad-CAM, LIME, SHAP, and Integrated Gradients with popular CNN backbones. All baseline methods employ a Bayesian optimized SVM classifier to ensure a fair comparison from a classification standpoint.

Table 5. Evaluation of the suggested approach in comparison with cutting-edge XAI-enabled CNNs

| XAI-Based Methodology | Classifier | Accuracy (%) |
|---|---|---|

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1s, 2026, Pages: 3340-3351

| | | |
|---|---|---|
| Grad-CAM + ResNet50 | Bayesian Optimized SVM | 85.3 |
| LIME + VGG19 | Bayesian Optimized SVM | 84.7 |
| SHAP + DenseNet121 | Bayesian Optimized SVM | 87.1 |
| Integrated Gradients + Inception-ResNet-V2 | Bayesian Optimized SVM | 86.4 |
| Proposed EIAFF-Net | End-to-End Herd Consensus Classifier | 93.9 |

The proposed EIAFF-Net clearly outperforms all compared methods by a significant margin of 6–9% absolute accuracy gain. Unlike the baseline approaches, which rely on single-backbone feature extraction followed by an external classifier, EIAFF-Net employs multi-scale feature fusion, adaptive memory gating, and consensus-driven classification within a unified architecture.

Moreover, the explainability mechanisms in baseline models are largely post-hoc and loosely coupled to the feature learning process. In contrast, EIAFF-Net integrates explainability-aware design choices, ensuring that the learned representations themselves remain clinically meaningful.

### 6.2 Comparison with State-of-the-Art Gastrointestinal Diagnostic Models

A broader comparison with previously published gastrointestinal disease detection frameworks is presented in Table 6. This comparison highlights not only predictive accuracy but also the presence of explainability, uncertainty handling, and optimization strategies, which are increasingly demanded in medical AI systems.

**Table 6. Comparison of the proposed approach with state-of-the-art methods**

| Attribute | [26] | [12] | [19] | Proposed (EIAFF-Net) |
|---|---|---|---|---|
| Year | 2019 | 2020 | 2022 | 2025 |
| Disease Coverage | Gastritis, Ulcer | Barrett's Esophagus, Polyp | Esophagitis | Gastric Erosion, Polyps, Antral Gastritis, Fundic Inflammation, Normal |
| Accuracy (%) | 91.7 | 93.2 | 87.5 | 93.9 |
| Explainable AI (XAI) | NA | NA | NA | Grad-CAM, SHAP/LIME |
| Uncertainty Handling | NA | NA | NA | Dropout Inference, Monte Carlo Sampling |

This comparison demonstrates that earlier works primarily focus on accuracy-driven diagnosis with limited disease coverage and no explicit mechanisms for interpretability or uncertainty estimation. The proposed framework not only achieves higher accuracy but also introduces a holistic diagnostic pipeline, combining explainability, robustness, and optimization within a single system.

From a clinical deployment perspective, this multidimensional superiority is critical, as regulatory and ethical requirements increasingly demand transparency, reliability, and risk awareness alongside predictive performance.

### 6.3 Discussion of Comparative Findings

The comparative results confirm that EIAFF-Net advances the state of the art by addressing key limitations of prior approaches. While earlier methods rely heavily on post-hoc explanations and static feature representations, the proposed framework integrates adaptive memory mechanisms and collective decision-making, leading to superior generalization and interpretability. The ability of EIAFF-Net to maintain high performance across a broader spectrum of gastrointestinal conditions further reinforces its suitability for real-world endoscopic diagnostic support.

## 7. Future Work and Limitations:

Despite the strong empirical performance demonstrated by EIAFF-Net, certain limitations remain that open important avenues for future investigation. A primary limitation of the present study is its reliance on a single benchmark dataset, namely Kvasir-V2. Although this dataset is widely accepted and clinically validated, evaluating the proposed framework across multiple datasets and institutions is essential to fully establish its robustness under diverse imaging protocols, endoscopic devices, and patient demographics. Furthermore, the current implementation is restricted to static endoscopic images, whereas clinical endoscopy is inherently temporal. Extending EIAFF-Net to incorporate temporal modelling through recurrent architectures or transformer-based video analysis could significantly enhance diagnostic consistency and capture disease progression more effectively.

From an interpretability and safety perspective, while Grad-CAM and LIME provide valuable visual explanations, future research may benefit from incorporating counterfactual and concept-based explainable AI approaches that allow clinicians to explore "what-if" scenarios and better understand borderline or ambiguous cases. In addition, although uncertainty handling mechanisms such as Monte Carlo sampling have been introduced, deeper integration of confidence calibration and risk-aware decision thresholds would further improve patient safety in real-world deployment. Future extensions of this work will therefore focus on multi-institutional validation, real-time inference optimization for clinical environments, temporal endoscopic video analysis, and advanced uncertainty-aware explainability, with the overarching goal of transitioning EIAFF-Net from a high-performing research prototype into a clinically deployable and trustworthy AI system for gastrointestinal disease diagnosis.

## 8. Conclusion:

This work presented EIAFF-Net, a novel elephant-inspired explainable deep learning framework for automated gastrointestinal disease detection and multi-class classification from endoscopic images. The proposed approach was designed to address two critical challenges in medical image analysis: achieving high diagnostic accuracy while simultaneously ensuring model transparency and clinical interpretability. EIAFF-Net integrates multi-scale feature extraction, adaptive memory-based feature modulation, and a herd consensus classification strategy within a unified end-to-end architecture. This design enables the model to effectively capture fine-grained texture patterns, structural variations, and high-level semantic cues that are essential for differentiating visually similar gastrointestinal conditions. The adaptive memory module further enhances robustness by selectively emphasizing diagnostically relevant features while suppressing redundant activations, leading to stable convergence and improved generalization. Extensive experiments conducted on the Kvasir-V2 dataset demonstrate that the proposed framework achieves strong and consistent performance across all evaluation metrics, including accuracy, F1-score, Matthews Correlation Coefficient, ROC-AUC, and Average Precision. The close alignment between validation and test performance confirms the model's generalization capability, while the high MCC and ROC-AUC values indicate reliable and balanced multi-class discrimination.

Comparative analysis against state-of-the-art XAI-enabled deep learning models further establishes the superiority of EIAFF-Net in terms of both predictive accuracy and methodological completeness. Beyond quantitative performance, the integration of Grad-CAM and LIME provides clinically meaningful visual explanations, highlighting lesion-specific regions that align with gastroenterological expertise.

MSW MANAGEMENT -Multidisciplinary, Scientific Work and Management Journal
ISSN: 1053-7899
Vol. 36 Issue 1s, 2026, Pages: 3340-3351

The consistency between global and local explainability outputs reinforces trust in the model's decision-making process and supports its suitability for real-world clinical adoption. Overall, this study demonstrates that biologically inspired architectural principles, when combined with modern explainable AI techniques, can significantly advance the development of trustworthy, interpretable, and high-performance medical AI systems. EIAFF-Net represents a promising step toward practical AI-assisted gastrointestinal diagnostics, offering clinicians a reliable decision-support tool that balances accuracy, transparency, and robustness.

References

[1] Zhang, M., Pan, J., Lin, J., Xu, M., Zhang, L., Shang, R., ... & Yu, H. (2023). An explainable artificial intelligence system for diagnosing Helicobacter Pylori infection under endoscopy: a case–control study. *Therapeutic advances in gastroenterology*, *16*, 17562848231155023.

[2] Sharma, D. K., Chatterjee, M., Kaur, G., & Vavilala, S. (2022). Deep learning applications for disease diagnosis. In *Deep learning for medical applications with unique data* (pp. 31-51). Academic Press.

[3] Malik, H., Anees, T., Din, M., & Naeem, A. (2023). CDC Net: Multi-classification convolutional neural network model for detection of COVID-19, pneumothorax, pneumonia, lung Cancer, and tuberculosis using chest X-rays. *Multimedia Tools and Applications*, *82*(9), 13855-13880.

[4] Roy, D. S. (2025). Machine Unlearning Models for Medical Care. *Exploration of Transformative Technologies in Healthcare 6.0*, 273.

[5] Tritscher, J., Krause, A., & Hotho, A. (2023). Feature relevance XAI in anomaly detection: Reviewing approaches and challenges. *Frontiers in Artificial Intelligence*, *6*, 1099521.

[6] Fefferman, D. S., & Farrell, R. J. (2005). Endoscopy in inflammatory bowel disease: indications, surveillance, and use in clinical practice. *Clinical Gastroenterology and Hepatology*, *3*(1), 11-24.

[7] Ozawa, T., Ishihara, S., Fujishiro, M., Kumagai, Y., Shichijo, S., & Tada, T. (2020). Automated endoscopic detection and classification of colorectal polyps using convolutional neural networks. *Therapeutic advances in gastroenterology*, *13*, 1756284820910659.

[8] Cui, Y., Song, Y., Sun, C., Howard, A., & Belongie, S. (2018). Large scale fine-grained categorization and domain-specific transfer learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4109-4118).

[9] Ennab, M., & Mcheick, H. (2025). Advancing AI Interpretability in Medical Imaging: A Comparative Analysis of Pixel-Level Interpretability and Grad-CAM Models. *Machine Learning and Knowledge Extraction*, *7*(1), 12.

[10] Ouyang, Y., Wu, Y., Wang, H., Zhang, C., Cheng, F., Jiang, C., ... & Li, Q. (2023). Leveraging historical medical records as a proxy via multimodal modeling and visualization to enrich medical diagnostic learning. *IEEE Transactions on Visualization and Computer Graphics*, *30*(1), 1238-1248.

[11] Meng, C., Trinh, L., Xu, N., Enouen, J., & Liu, Y. (2022). Interpretability and fairness evaluation of deep learning models on MIMIC-IV dataset. *Scientific Reports*, *12*(1), 7166.

[12] Khan, M. A., Sharif, M., Akram, T., Yasmin, M., & Nayak, R. S. (2019). Stomach deformities recognition using rank-based deep features selection. *Journal of medical systems*, *43*, 1-15.

[13] Fan, S., Xu, L., Fan, Y., Wei, K., & Li, L. (2018). Computer-aided detection of small intestinal ulcer and erosion in wireless capsule endoscopy images. *Physics in Medicine & Biology*, *63*(16), 165001.

[14] Diamantis, D. E., Iakovidis, D. K., & Koulaouzidis, A. (2019). Look-behind fully convolutional neural network for computer-aided endoscopy. *Biomedical signal processing and control*, *49*, 192-201.

[15] Sharif, M., Attique Khan, M., Rashid, M., Yasmin, M., Afza, F., & Tanik, U. J. (2021). Deep CNN and geometric features-based gastrointestinal tract diseases detection and classification from wireless capsule endoscopy images. *Journal of Experimental & Theoretical Artificial Intelligence*, *33*(4), 577-599.

[16] Alaskar, H., Hussain, A., Al-Aseem, N., Liatsis, P., & Al-Jumeily, D. (2019). Application of convolutional neural networks for automated ulcer detection in wireless capsule endoscopy images. *Sensors*, *19*(6), 1265.

[17] Ucar, F., & Korkmaz, D. (2020). COVIDiagnosis-Net: Deep Bayes-SqueezeNet based diagnosis of the coronavirus disease 2019 (COVID-19) from X-ray images. *Medical hypotheses*, *140*, 109761.

[18] Nour, M., Cömert, Z., & Polat, K. (2020). A novel medical diagnosis model for COVID-19 infection detection based on deep features and Bayesian optimization. *Applied Soft Computing*, *97*, 106580.

[19] Gal, Y., & Ghahramani, Z. (2016, June). Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning* (pp. 1050-1059). PMLR.

[20] Leibig, C., Allken, V., Ayhan, M. S., Berens, P., & Wahl, S. (2017). Leveraging uncertainty information from deep neural networks for disease detection. *Scientific reports*, *7*(1), 1-14.

[21] Bach, S., Binder, A., Montavon, G., Klauschen, F., Müller, K. R., & Samek, W. (2015). On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. *PloS one*, *10*(7), e0130140.

[22] Pogorelov, K., Randel, K. R., Griwodz, C., Eskeland, S. L., de Lange, T., Johansen, D., ... & Halvorsen, P. (2017, June). Kvasir: A multi-class image dataset for computer aided gastrointestinal disease detection. In *Proceedings of the 8th ACM on Multimedia Systems Conference* (pp. 164-169).

[23] Wu, J., Chen, X. Y., Zhang, H., Xiong, L. D., Lei, H., & Deng, S. H. (2019). Hyperparameter optimization for machine learning models based on Bayesian optimization. *Journal of Electronic Science and Technology*, *17*(1), 26-40.

[24] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016, August). " Why should i trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 1135-1144).

[25] Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. (2017). Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision* (pp. 618-626).

[26] Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., ... & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical image analysis*, *42*, 60-88.