Deep feature analysis, classification with AI-driven gastrointestinal diagnostics

Janagama Srividya[1*] and *Harikrishna* Bommala[1]

[1]Department of CSE, Bharatiya Engineering Science & Technology Innovation University, Anantapur, Andhra Pradesh.
[1]Department of Computer Science and Engineering, CMR Technical Campus, kandlakoya, Medchal, TS.
[1]Department of CSE, KG Reddy College of Engineering & Technology, Hyderabad, Telangana, India

Abstract. Several AI-based methods have substantially progressed the area of medical image and video-based diagnostics, which encompasses radiography, pathology, endoscopy, and the categorization of gastrointestinal (GI) diseases. When it comes to classifying numerous GI disorders, the majority of prior research that relies solely on spatial cues performs poorly. While some prior research has made use of temporal features trained on a 3D convolution neural network, these studies have focused on a very small subset of the gastrointestinal system and have used very few classes. To address these concerns, we introduce an all-inclusive AI-based system for classifying different GI illnesses using endoscopic recordings. This system can extract spatial and temporal data concurrently, leading to improved classification performance. For temporal variables, we employ a long short-term memory model; for spatial features, we employ two independent residual networks in cascade mode.

1 Introduction

Wireless capsule endoscopy (WCE) has emerged as a crucial tool in the research of gastrointestinal disorders due to its non-invasive, painless, and safe manner of providing comprehensive imaging of the whole GI tract. This technology has changed the way gastrointestinal illnesses are diagnosed and treated since its inception in 2001 [1]. It has been extensively used in clinical practice since then. Nevertheless, there is a major downside: patients need 6-8 hours to do a GI examination, which produces an astonishing 50,000 to 80,000 photos. Specialists in gastrointestinal disorders who devote a great deal of effort todiagnosing and analyzing images face a formidable obstacle in this regard. Novel approaches to gastrointestinal diagnostics are urgently required to address this critical issue. A total of $4.59 billion was the value of the worldwide market for gastrointestinal diagnostics in 2021.

* Corresponding author: haribommala@gmail.com

From 2022 to 2030, it is anticipated to increase at a CAGR of 4.58%. The increasing demand for technologically improved molecular tests, the rising prevalence of gastrointestinal infections and malignancies, and the increasing demand for point-of-care (POC) testing for gastrointestinal illnesses are all contributing factors to this market expansion [2].

When new diagnostic tools are introduced, they speed up the market's velocity even further. The gastrointestinal diagnostics industry is poised for significant growth, thanks to recent regulatory approvals and the launch of cutting-edge diagnostic technologies. This study intends to go further into technical breakthroughs based on artificial intelligence (AI) in light of these developments and difficulties within the field of gastrointestinal diagnostics.

1.1 Research Objectives

The following are the stated goals of the research:

•The goal of deep feature analysis and classification in AI-driven gastrointestinal diagnostics is to improve disease detection and classification by revealing previously unseen linkages and patterns in large amounts of medical data.

•To investigate the potential of algorithms that draw inspiration from nature to optimize feature selection in gastrointestinal (GI) diagnostics;specifically, to find the most useful features and improve classification accuracy by emulating biological and ecological processes.

•The goal of this study is to find out how to optimize AI-based GI diagnoses using algorithms and approaches inspired by biology, with the goal of identifying crucial diagnostic features and improving diagnostic model performance through the application of optimization strategies inspired by nature.

2 Literature Survey

Two deep learning models utilizing the Moth-Crow approach were suggested [3] for the diagnosis of gastrointestinal illnesses. The process began with boosting the image's contrast before moving on to data augmentation and pretrained DL models. In order to classify the images, a machine learning algorithm was fed the extracted features together with the distance canonical correlation method's features. In the meanwhile, deep learning models for GI disease classification and machine learning techniques for classifying universal features were introduced [4]. One data set was used for training the systems, while another data set was used for evaluation. For the purpose of classifying the gastroenterology dataset, Yogapriya et al. [5] employed three convolutional neural networks (CNN). When there weren't enough photographs in the dataset, data augmentation was employed to fill in the gaps. With an accuracy of 96.33% and a recall of 96.37%, the VGG16 model outperformed the others. Also, a CNN with coding and decoding layers to extract spatial characteristics was suggested [6]. To achieve this level of parity, data augmentation was employed. ResNet50 achieved a precision of 90.28% and GoogLeNet 91.38%. In order to categorize the Kvasir dataset for the purpose of detecting GI disorders, [7] suggested five CNN models. This section describes the usage of an image processing method to eliminate artifacts and a data augmentation technique to get sufficient images. Accuracy was 92.3 percent for the ResNet model and 90.3 percent for the Inception-v3 model. To identify polyps using the self-attention technique, Wan et al. [8] introduced the YOLOv5 model. When features are extracted to improve information-rich channels and deactivate unimportant ones, self-attention is engaged. A recall accuracy rate of 98.9% and a precision rate of 91.5% were achievedby the YOLOv5 model. [9] suggested a technique that could be used to train a dataset and then be applied to additional datasets; the approach was successful in achieving an F1 score of 91% on an internal dataset. Meanwhile, a two-network strategy was [10] for the classification of gastrointestinal disorders, using a pretrained deep learning network that achieved a sensitivity level of 98.13%. A Jaccard index of 75.18% was attained by the second network that relied on lesion extraction and isolation from the healthy zone. The features of gastrointestinal disorders can be extracted and classified using DenseNet-201 and Inception v3, [11]. Before merging the features of the two models, we trained the system with more data, extracted deep features, and then utilized the dragonfly optimization approach and fused feature classification to improve the system.Using a number of criteria, [13] laid up a method for diagnosing gastrointestinal disorders. The matrix-based technique was used to train VGG16 to extract and integrate features. The PSO approach was used to choose the characteristics based on their fitness. The features that were chosen were categorized using the cubic SVM method. One convolution neural network (CNN) model for gastrointestinal illness detection was [13]. In order to acquire the ulcer area, a customized mask was used for segmentation, and RCNN was trained on the dataset. After ResNet101 extracted the features, Grasshopper optimization using a fitness function was used to optimize them. In the meanwhile, the gastroenterology data set was classified using a CNN [14]. After being fed into the LSTM, all of the layers were utilized to categorize the features of the pooled layer map. For the goal of testing the system's performance using independent endoscopic pictures [15] proposed utilizing the ResNet50 model to train images from the GI dataset. The system achieved a sensitivity level of 90.8% and an accuracy rate of 94.7% in cases of active EoE.

3 Methodology

Video data can be classified using spatiotemporal features by utilizing our suggested classification architecture, which incorporates cascaded CNN and LSTM deep networks. Our network's primary strength is the significant improvement in performance it achieves when classifying a sequence of n consecutive images of varied length (I_1,I_2,I_3,…,In). For instance, improved classification performance is a direct outcome of using more sequential photos. Our cascaded deep learning model outperformed models that relied solely on convolutional neural networks (CNNs). This is due to the fact that convolutional neural network (CNN) models analyze each input picture independently in an image dataset, whereas in a video dataset, both spatial and temporal features are included. Overall classification performance suffers when CNN models ignore time. In order to overcome the limitations of earlier methods in the medical field that utilized spatial information, our research involved enhancing classification performance by incorporating a spatial variant of a recurrent neural network (LSTM) into the traditional CNN model. As shown in Figure 1, our proposed classification system has a general outline. There are three distinct phases to the complete framework: initial feature extraction (spatial), feature extraction (temporal), and classification. The provided input sequence of endoscopic frames was subjected to a distinct set of deep learning techniques in each level. Consequently, 37 distinct types of gastrointestinal disorders were used to forecast the input sequence's final class designation. The sections that follow provide a comprehensive breakdown of each step.

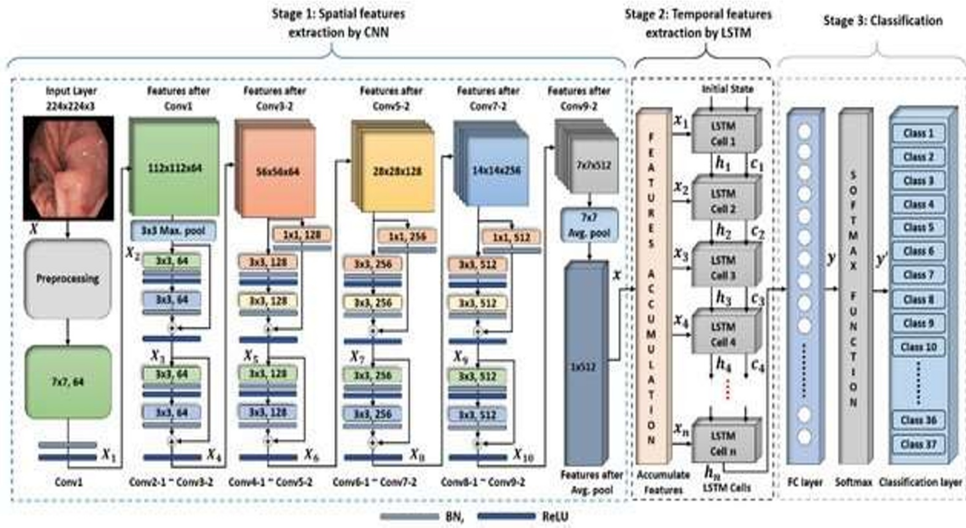

Fig.1. An outline of the suggested deep architecture for the categorization of various GI disorders using cascaded convolutional neural networks and long short-term memories (LSTMs).

4 Classification

In the last stage of classification, instead of using all the outputs (i.e., $h1,h2,h3,\ldots,hn$), the output feature vector is chosen as the output hn of the LSTM cell at the last time step n. Figure 1 shows the final classification process, which involves a stack of layers including FC, softmax, and classification. The FC layer that follows the last LSTM cell's output has the same number of nodes as the classes. The main goal of the FC layer is to find bigger patterns in the images by combining all the spatiotemporal information gathered by the previous layers.

5 Results and Discussion

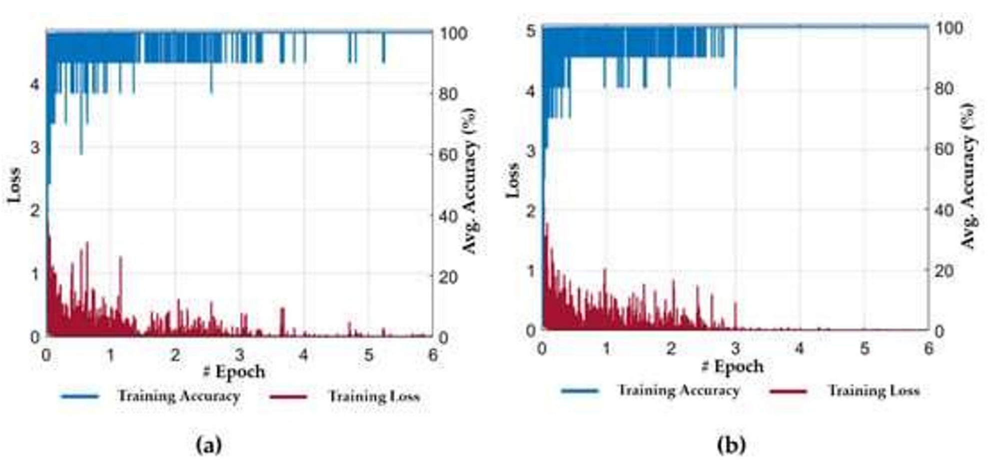

Fig. 2. Plots of training loss and accuracy for the first stage, which involves using ResNet18 to extract spatial features: (a) First step cross-validation; and (b) Second step cross-validation.

Both the training loss and the accuracy for the two folds of cross-validations are displayed in Figure 2, along with various epoch counts. After a certain number of epochs, we can claim that the model we've selected has been trained sufficiently when the training loss approaches zero and the training accuracy approaches 100%. Training loss and accuracy progress are displayed in Figure 3 for both folds of cross-validations. Once the training loss is close to zero and the training accuracy is close to 100% after a specific number of iterations in the first epoch, our model's second stage, the Long Short-Term Memory (LSTM) layer, has achieved optimal convergence.

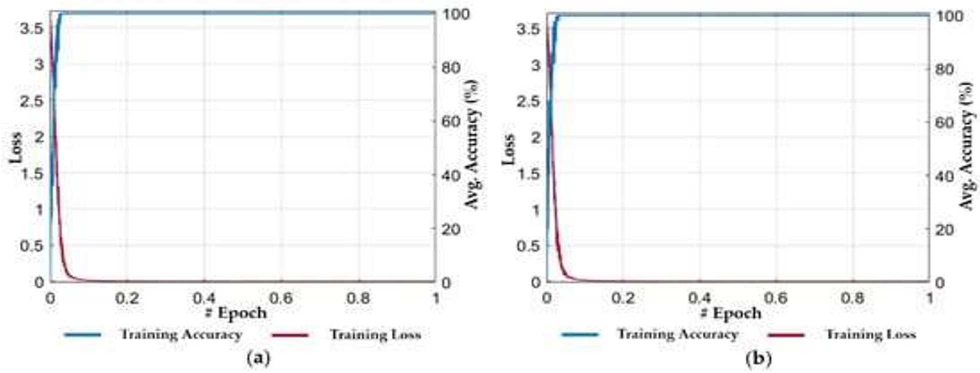

Fig.3. In the second phase, we use LSTM to extract temporal properties; here we can see training loss and accuracy plots: (a) An approach to cross-validation known as first-degree, and (b) an approach known as second-degree,

Figure 3 further shows that LSTM's convergence is more rapid and less choppy than ResNet18's (in the first stage). Instead of utilizing the successive frames for temporal feature extraction, an intermediate dataset including discriminating spatial feature vectors was used, which is the main cause of this outcome.

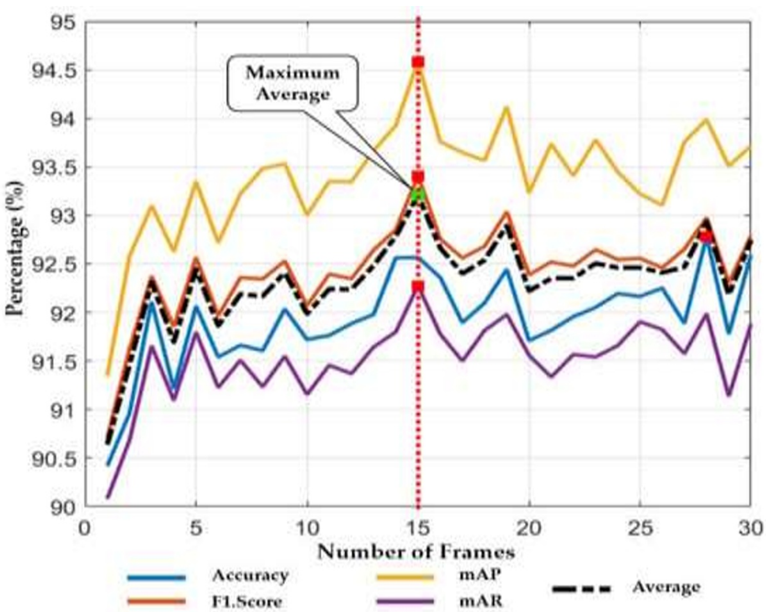

Fig. 4. framework's classification performance as a function of LSTM frames count (n).

In Figure 4, the green square box shows the maximum average performance, whereas the red square box illustrates the greatest performance with respect to the several performance metrics (i.e., accuracy, F1 score, mAP, and mAR). The best accuracy, according to the total maximum average performance, was achieved with 15 frames (n = 15).
The greatest area under the curve (AUC) value of 97.057% shows that our suggested strategy beats all of the existing baseline methods (Figure 5). Here are some more baseline methods: SqueezeNet (82.131%), AlexNet (87.332%), GoogLeNet (91.097%), VGG19 (92.039%), VGG16 (94.060%), InceptionV3 (95.000%), ResNet50 (95.724%), and ResNet18 (95.705%). The result of two-fold cross-validations is used to display all of these ROC curves. The performance gap can be better understood by referring to the expanded view provided by the left-hand figure 5.

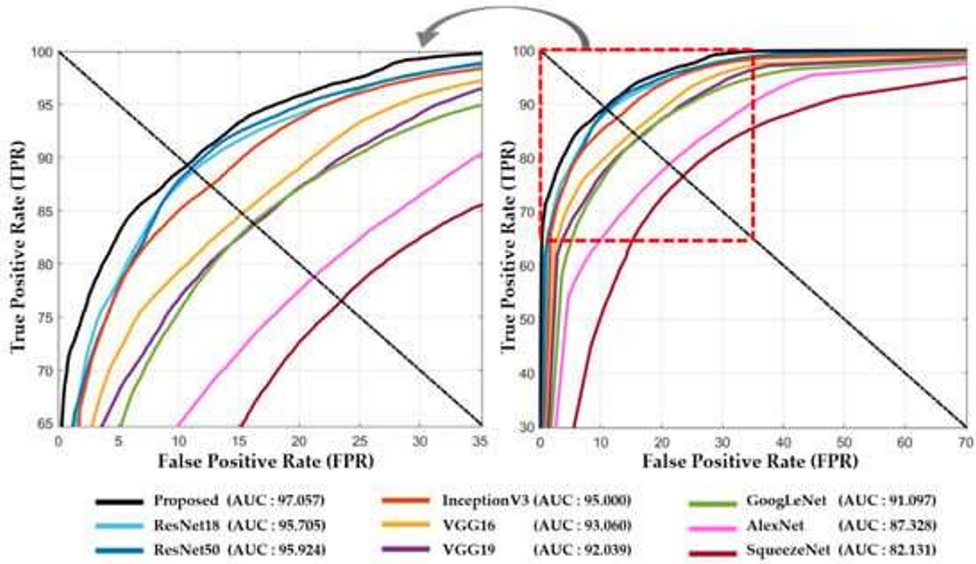

Fig. 5. Area under the curve (AUC) receiver operating characteristic curves for our suggested technique and various baseline models

6 Conclusion

Based on endoscopic recordings using convolutional neural networks (CNNs) and long short-term memory (LSTMs), we present a novel method for gastrointestinal disease classification in this research. In addition, we provide a system for retrieving endoscopic videos based on class predictions using the classification framework we propose. When compared to features learned from spatial information alone, the suggested spatiotemporal features-based approach allows for the encoding of better discriminative representations of multiple endoscopic images. It follows that merging geographical and temporal data improves classification and retrieval efficiency. To evaluate the effectiveness of the proposed method, we utilized the KVASIR database in conjunction with a publically available dataset from GastroLab. Also, in order to compare the different state-of-the-art approaches fairly, we used the same dataset and experimental protocol for each.

References

1. H. Sung, J. Ferlay, R. L. Siegel et al., "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," CA: A Cancer Journal for Clinicians, vol. 71, no. 3, pp. 209–249, 2021.
2. K. Pogorelov, M. Riegler, S. L. Eskeland et al., "Efficient disease detection in gastrointestinal videos–global features versus neural networks," Multimedia Tools and Applications, vol. 76, no. 21, pp. 22493–22525, 2017.
3. M. Khan, K. Muhammad, S. Wang et al., "Gastrointestinal diseases recognition: a framework of deep neural network and improved moth-crow optimization with dcca fusion," Human-centric Computing and Information Sciences, vol. 12, 2022.
4. V. Thambawita, D. Jha, H. L. Hammer et al., "An extensive study on cross-dataset bias and evaluation metrics interpretation for machine learning applied to gastrointestinal tract abnormality classification," ACM Transactions on Computing for Healthcare, vol. 1, no. 3, pp. 1–29, 2020.
5. J.Yogapriya, V. Chandran, M. Sumithra, P. Anitha, P. Jenopaul, and C. Suresh Gnana Dhas, "Gastrointestinal tract disease classification from wireless endoscopy images using pretrained deep learning model," Computational and Mathematical Methods in Medicine, vol. 2021, Article ID 5940433, 12 pages, 2021.
6. Z. M. Lonseko, P. E. Adjei, W. Du et al., "Gastrointestinal disease classification in endoscopic images using attention-guided convolutional neural networks," Applied Sciences, vol. 11, no. 23, Article ID 11136, 2021.
7. Dheir and S. A. Abu-Naser, "Classification of anomalies in gastrointestinal tract using deep learning," International Journal of Academic Engineering Research, vol. 6, pp. 15–28, 2022.
8. J. Wan, B. Chen, and Y. Yu, "Polyp detection from colorectum images by using attentive YOLOv5," Diagnostics, vol. 11, no. 12, p. 2264, 2021.
9. Nogueira-Rodríguez, M. Reboiro-Jato, D. Glez-Peña, and H. López-Fernández, "Performance of convolutional neural networks for polyp localization on public colonoscopy image datasets," Diagnostics, vol. 12, no. 4, p. 898, 2022.
10. E. H. Dulf, M. Bledea, T. Mocan, and L. Mocan, "Automatic detection of colorectal polyps using transfer learning," Sensors, vol. 21, no. 17, p. 5704, 2021.
11. F. Mohammad and M. Al-Razgan, "Deep feature fusion and optimization-based approach for stomach disease classification," Sensors, vol. 22, no. 7, p. 2801, 2022.
12. M. A. Khan, S. Kadry, M. Alhaisoni et al., "Computer-aided gastrointestinal diseases analysis from wireless capsule endoscopy: a framework of best features selection," IEEE Access, vol. 8, pp. 132850–132859, 2020.
13. M. A. Khan, M. A. Khan, F. Ahmed et al., "Gastrointestinal diseases segmentation and classification based on duo-deep architectures," Pattern Recognition Letters, vol. 131, pp. 193–204, 2020.
14. Ş. Öztürk and U. Özkaya, "Residual LSTM layered CNN for classification of gastrointestinal tract diseases," Journal of Biomedical Informatics, vol. 113, Article ID 103638, 2021.
15. E. Okimoto, N. Ishimura, K. Adachi, Y. Kinoshita, S. Ishihara, and T. Tada, "Application of convolutional neural networks for diagnosis of eosinophilic esophagitis based on endoscopic imaging," Journal of Clinical Medicine, vol. 11, no. 9, p. 2529, 2022.