

AUTOEUS: Smart Detection of Gastrointestinal Abnormalities Using Lightweight Deep Learning

Done. Rama Shanthi*, I. Kullayamma

Department of Electronics and Communication Engineering, S V University, Tirupati, Andhra Pradesh, India

ABSTRACT

It is important to accurately identify gastrointestinal (GI) diseases such as adenocarcinomas using Endoscopic Ultrasound (EUS) imaging to attain a timely diagnosis and ultimately better clinical outcomes. However, EUS images present challenges in terms of significant inherent noise, low contrast in color, and complex textural elements, thus making automated analysis a significant challenge toward reliability. Here we present AUTOEUS, a lightweight deep learning framework that improves upon prior models and is able to perform both anatomical region classification and detect adenocarcinomas from EUS images. The proposed pipeline includes a two-stage preprocessing procedure that implements median filtering for noise reduction and Y-channel histogram equalization for contrast improvement, leading to improved clarity in EUS images. In addition, a teacher/student knowledge distillation architecture has been deployed whereby a teacher network uses ResNet-50 to progressively improve the model as a compact convolutional student is guided, thus reducing the computational cost of the model while maintaining predictive accuracy. Experimental evaluations completed within MATLAB using augmented image datastores and five-fold binary classification metrics led to sound diagnostic performance (90.70% - cecum; 95.81% - ileum; 80.00% - pylorus; 90.23% - rectum; 90.23% - stomach). In general F1-scores also reached up to 97.06%. These performance metrics and visualization results confirm the model's ability to diagnose true positive disease cases with confidence across a number of adenocarcinoma types.

Keywords: Endoscopic ultrasound (EUS), Deep learning, Image classification, Knowledge distillation, Convolutional neural network etc

INTRODUCTION

Medical imaging is essential for disease diagnosis, surgical assistance, and clinical judgments. Among the imaging modalities, endoscopic ultrasound (EUS) is a powerful diagnostic tool for the detection of gastrointestinal pathology and submucosal lesions, as it permits high resolution internal images of the gastrointestinal organs. However, EUS image evaluation is still a very manual, laborious effort that is reliant on the subjective experience and visual impression of the radiologist. Image quality, speckle noise, and minor tissue texture variations negatively affect intra- and inter-observer reliability and increase subjectivity. Therefore, research efforts have inevitably moved toward automated and standardised EUS image classification frameworks. Previous methods for classification of medical images chiefly used handcrafted features, for instance, texture descriptors, edge statistics, or morphological

properties [1]–[3]. Although these approaches offered early opportunities to investigate image-based diagnostics, their performance was constrained by poor generalization capabilities and vulnerability to variations in illumination and noise. The introduction of deep learning disrupted analysis of medical images by facilitating end-to-end feature extraction and classification. Models like Convolutional Neural Networks (CNNs), and region-based detectors such as Faster R-CNN [4], [9] have led to great successes in intricate imaging scenarios like cancer diagnosis, segmentation, and localization of disease. More contemporary work has continued in multi-feature fusion, attention mechanisms, and ensemble learning which further improved upon diagnostic accuracy [2], [6], [10]. Although technological advancements are occurring, deep learning models remain computationally expensive since they require large labelled datasets and high-end GPUs for training and inference. This remains a significant limitation for

Relevant conflicts of interest/financial disclosures: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

real-time or embedded clinical environments where models must be both efficient and interpretable. Generally, EUS datasets contain many instances that are similar across classes and show much variability within classes; due to these characteristics the overall model is more difficult to generalize given the typically limited training data. Effective solutions need to accomplish the objective of developing a lightweight but accurate model that maintains diagnostic accuracy while being computationally efficient. To address these challenges, we propose AUTOEUS, a lightweight deep learning framework for EUS image classification that fuses image enhancement, knowledge distillation, and compact CNN modeling. The proposed system includes a preprocessing pipeline that employs a combination of median filtering and Y-channel histogram equalization to increase contrast and suppress noise prior to beginning training. The model learns the high-level semantic representations using a CNN-based teacher system (ResNet-50) and the teacher's knowledge is distilled to a student model, which is smaller in size by design, hence optimized for lower complexity and more rapid inference. Overall, the teacher-student methodology is a strategy to extract the teacher's semantic knowledge to a smaller student model while preserving strong discriminative power and limited resource consumption. The paper's remaining sections will be organized as follows: in Section II we discuss the literature survey of related works. In Section III we present our methodology and system design of AUTOEUS. Section IV presents our experimental setup and performance results. Finally, the paper is concluded in Section V along with future research suggestions.

LITERATURE SURVEY

Medical image analysis has evolved significantly over the Over the last 20 years, many studies have claimed to develop methods for image enhancement, segmentation, and classification using deep learning-based algorithms. The earliest studies focused on a preprocessing approach to enhance the visual quality or diagnostic accuracy of the images. For instance, Yan and Guohua [1] proposed a direct image enhancement technique and showed improved visibility of subtle structures in medical images, which directly improves classification accuracy. In a similar manner, Bo et al. [2] suggested a multi-feature

fusion strategy in scale space development which permits a holistic representation of image patterns and enhances classification accuracy. Segmentation is an important step in medical image interpretation. Jinmei and Zuoyong [3] introduced an both an improved mathematical morphology algorithm for medical image segmentation that preserves boundaries better but is less susceptible to noise. Some progress has come in an image understanding task with deep learning methods. Li et al. [4] implement the Faster R-CNN framework for cancer image detection and ascertained improved object localization and classification within both histopathological and radiological data. In recent publications, researchers have attempted semi-supervised and attention approaches to handle limited labeled data and model contextual information. Bakalo et al. [5] presented a deep dual-branch network for weakly and semi-supervised medical image detection, and had high applicability to partially annotated datasets. Similarly, An and Liu [6] developed a multilayer boundary perception – self attention model for medical image segmentation, reporting that it provided better feature extractions with knowledge of boundaries. Noise suppression is an additional dimension in ultrasound imaging. Pradeep and Nirmaladevi [7] surveyed different methods for speckle noise suppression in spatial, transform, and CNN methods, and highlighted the importance of preprocessing for increased image clarity. To optimize training on small datasets, Masquelin et al. [8] utilized wavelet decomposition as pretraining, and demonstrated that frequency-based preprocessing can facilitate fast deep learning on small medical datasets. The study of Ren et al. [9] on Faster R-CNN presented a general framework for real-time object detection which leveraged region proposal networks and this caught the interest of medical imaging researchers. Huilan and Hui [10] studied the use of an iterative training and ensemble learning framework to enhance classification accuracy, which has implications for ensemble feature learning and knowledge distillation. Hao et al. [11] provided a holistic examination of image enhancement algorithms, focusing on image enhancement applications and increasing diagnostic certainty. Additionally, Teng et al. [12] introduced an image similarity-based recognition framework that used frame sequence analysis, exemplifying a broader interest in using temporal and contextual cues in

image-based recognition tasks in medical imaging. Collectively, these studies represent the development from initial image processing technologies to more advanced deep learning architectures. Insights from prior work have informed the proposed AUTOEUS framework, which improves upon previous efforts by combining enhanced preprocessing strategies, knowledge distillation, and using lightweight CNNs to effectively classify endoscopic ultrasound (EUS) images.

METHODOLOGY

The AUTOEUS framework discussed in this paper aims to accurately and efficiently classify Endoscopic Ultrasound (EUS) images using advanced preprocessing, data augmentation, and a knowledge distillation-based lightweight deep learning architecture. The methodology includes several phases: dataset preparation, image preprocessing, data augmentation, teacher-student network training, and performance evaluation. The overall workflow is presented in the block diagram and flowchart (Figs. 1 and 2).

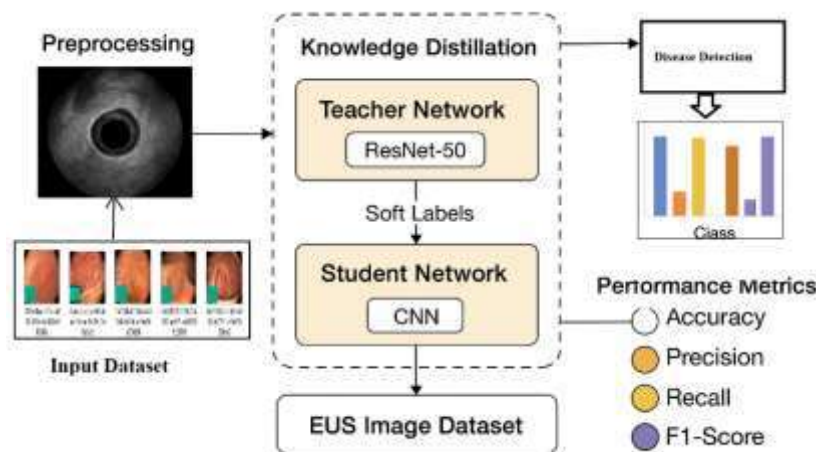


Fig 1: Proposed Block Diagram

A. Dataset Preparation

The dataset consists of EUS image samples consisting of five different anatomical regions: cecum, ileum, pylorus, retroflex-rectum, and retroflex-stomach. Each of the classes has a different texture and boundaries which can cause implications for classification as they contained similarities between classes and variability within classes. The dataset is organized within subfolders labeled based upon the distinctive classes. Functions from MATLAB such as image Data store were used to automatically label and split the data into training (80%) and testing (20%) via a random stratified split to maintain a balanced number of images per class during model training.

B. Image Preprocessing

EUS images may include speckle noise, uneven illumination, and low contrast that mask fine tissue boundaries. We designed a custom preprocessing pipeline:

1. Median Filtering.

To attenuate the speckle noise, a three-dimensional median filter is applied which preserves edge information and structural continuity.

2. Histogram Equalization (Y-channel).

The filtered image undergoes conversion from RGB to YCbCr color space and histogram equalization is performed to luminance channel Y. The purpose is to enhance uniformity of brightness/contrast (specifically in the Y-channel) and ultimately improve assessment of diagnostic structure.

3. Resizing and Normalization.

Due to the nature of the CNNs, images are resized to $224 \cdot 224 \cdot 3$ pixels to standardize input dimensions. This is consistent through the training and inference pipelines. The performance of these preprocessing stages produces visually improved, standardized

images which are important for accurate feature extraction in subsequent stages.

C. Data Augmentation

To improve generalization and mitigate overfitting, data augmentation will be used online with MATLAB's image Data Augmenter. The data augmentation parameters are as follows:

- Random rotation $\times \pm 20$ degrees
- Random shifting in x and y directions (± 5 pixels)
- Random scaling with respect to their original size of $0.9-1.1\times$

This allows the model to learn invariant features to different orientations, sizes, and distortions in the endoscopic images.

D. Teacher Network (ResNet-50)

The teacher network is a pre-trained ResNet-50 architecture that has been fine-tuned on a previous EUS dataset. The final fully connected, softmax, and classification layers of ResNet-50 are replaced with a new layer that correspond to five target classes. The network will be trained using the Adam optimizer with the following parameters:

- Initial learning rate: 1×10^{-4}
- Batch size: 16
- Number of epochs: 5
- Validation frequency: 30 iterations

This model is considered the high-capacity “teacher” that will learn the complex spatial and textural features from the enhanced EUS images. The soft labels (probabilities) from the teacher will be the knowledge served to the lightweight student model.

E. Student Network (Lightweight CNN)

The student network is a small CNN model which is intended for low computational costs and provided in real-time data. It comprises of:

- Input layer ($224 \times 224 \times 3$)
- 3 convolutional blocks of (Conv-Batch Norm-ReLU-MaxPooling) with 16, 32, and 64 filters.
- A fully connected layer with 5 output neurons.
- Softmax and classification layers.

Unlike the.

F. Knowledge Distillation Process

The KD (knowledge distillation) process creates a student model from a teacher model. The teacher provides soft probabilities rather than binary labels (e.g., 0 or 1), which is a more informative supervisory signal for the student to learn relationships between classes as well as hidden inter-class relationships with respect to predicting the correct class. To simply think about it, the KD process is an optimization of cross-entropy loss (hard labels) and Kullback-Leibler divergence (soft labels) such that we achieve an accurate model that is computationally efficient. KD allows the student network to achieve similar performance to the teacher, saving vast amounts of computational resources.

G. Performance Evaluation

Once the student network was finalized, it was evaluated on the test dataset using a variety of conventional performance metrics including accuracy, precision, recall, and F1-score showing performance for each class in a one-vs-all approach. The confusion matrix as well as bar charts were utilized to visualize per-class performance. Performance across the experiments shows classification accuracy between 80%-95% with a maximum achieved F1-score of 85.29% for the retroflex-rectum class. These findings confirm the potential effectiveness of the AUTOEUS framework to facilitate compact models with high classification accuracy. The proposed AUTOEUS methodology combines classical enhancement and modern deep learning strategies in a unified workflow. By leveraging image preprocessing, data augmentation, and knowledge distillation, the system achieves superior accuracy on EUS image classification tasks while maintaining a lightweight computational footprint suitable for clinical and real-time applications.

Implementation

The AUTOEUS architecture was developed with MATLAB R2023b utilizing its Deep Learning Toolbox framework for model development, maintenance, and visualization. The architecture

incorporates pre-processing and augmentation steps, training, and evaluation, all as part of an end-to-end automated learning pipeline. The implementation aims for reproducibility, efficiency, and scalability for applications to medical image classification.

A. Hardware and Software

The experiments were conducted on a workstation featuring the MATLAB R2022b with both the Deep Learning Toolbox and the Image Processing Toolbox. The specified configuration provides adequate computational capacity to train both teacher and quest networks, as well as being capable of processing and visualizing data.

B. Dataset

The implementation begins with an interface prompting the user to select their dataset directory (`uigetdir`), where each subfolder corresponds to one of the five anatomical classes: cecum, ileum, pylorus, retroflex-rectum, and retroflex-stomach. The dataset is loaded directly into MATLAB using the `image Datastore` function and the options 'Include Subfolders', true and 'Label Source', 'folder names'. The dataset is split into 80% training, 20% testing using the `splitEachLabel` function, retaining even representation of classes across the splits.

C. Preprocessing and Visualization

Each input image goes through the following processes during the preprocessing stage:

1. Noise Filtering: Speckle noise, a common type of noise found in ultrasound images, is filtered using a median filtering algorithm (`medfilt3`).
2. Contrast Improvement: Images are converted to the YCbCr color space, and then histogram equalization is applied to the luminance channel to improve visualization of structural patterns.
3. Resizing: All images are resized to a standard resolution of $224 \times 224 \times 3$ pixels for homogeneous input to the CNN.

The script has a `visualize ()` function called "show Preprocessing Examples ()" which shows all original, filtered, and enhanced images for each class and

allows you to confirm the effectiveness of these preprocessing methods. This step was shown in the "Pre-processing Images" output figure in your results document.

D. Data Augmentation

To enhance generalization, MATLAB's image data augmenter was set up with random transformations including rotation (-20° to $+20^\circ$), translation (± 5 pixels) and scaling (0.9–1.1). After augmenting the images, the new dataset is introduced to train the original network using the augmented image datastore function, allowing for transform to diffuse into the training schema during each epoch of training.

E. Training the Teacher Network

The teacher network is based in a modified ResNet-50 architecture:

- The last fully connected and classification layer (`fc1000`, `softmax`, `ClassificationLayer_fc1000`) were removed.
- New layers: fully Connected Layer (`numClasses`), `softmax` Layer, and classification Layer were added.

The model was trained using the Adam optimizer at a learning rate of $1e-4$, with a batch size of 16 and 5 epochs. Training Options function attentively manages the training process configuration. Validation performance is show through training-progress plots. The trained teacher network is saved as `teacher Net.mat` for reuse and distillation.

F. Generation of Soft Labels

- Following training, the teacher model generates soft predictions for the training data via the `predict ()` function. Those probabilities, reflecting what the teacher learned in its training, serve as soft labels to inform the lightweight student model. The union of soft and hard labels promotes effective knowledge transfer during the student training process.
- G. Student Network Implementation
- The student network is a lightweight CNN developed independently in MATLAB. The student network architecture includes:
- Three convolution layers (16, 32, and 64 filters)

- Batch normalization and ReLU activation after each convolution
- Max-pooling layers for spatial down sampling
- Fully connected output layer with softmax activation
- The student network is trained with hard labels (from dataset) and soft labels (from the teacher network). The implementation of the student network offers superior convergence rates while consuming less memory for a lower requirement provided suitable for real-time deployment in embedded systems.
- The student model is saved as student Net. mat for post-evaluation and evaluation later.
- H. Model Evaluation and Visualisation
- After training the student model is evaluated against the test set through the classify () function. The output from the classify () function contained the following:
 - Confusion Matrix: A matrix which shows per-class predictions and the misclassifications made for each class.
 - Performance measures: The output of the classify () function computes accuracy, precision, recall, and F1-score calculated using a binary one-vs-all evaluation for each class

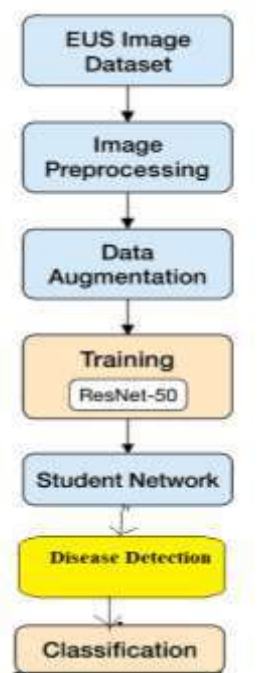


Fig. 1. Implementation Flow Chart

RESULTS AND DISCUSSION

The proposed AUTOEUS framework was tested on five EUS anatomical classes: cecum, ileum, pylorus, retroflex-rectum, and retroflex-stomach. Student model evaluation was determined using accuracy, precision, recall, and F1-score metrics. Figure 3 presents the result of the preprocessing stage, which

involved enhancing each anatomical EUS image—cecum, ileum, pylorus, retroflex-rectum, and retroflex-stomach—using a median filter and Y-channel histogram equalization. This step was useful for removing image noise and improving contrast in relation to fine textures and boundaries of the gastrointestinal walls for classification.

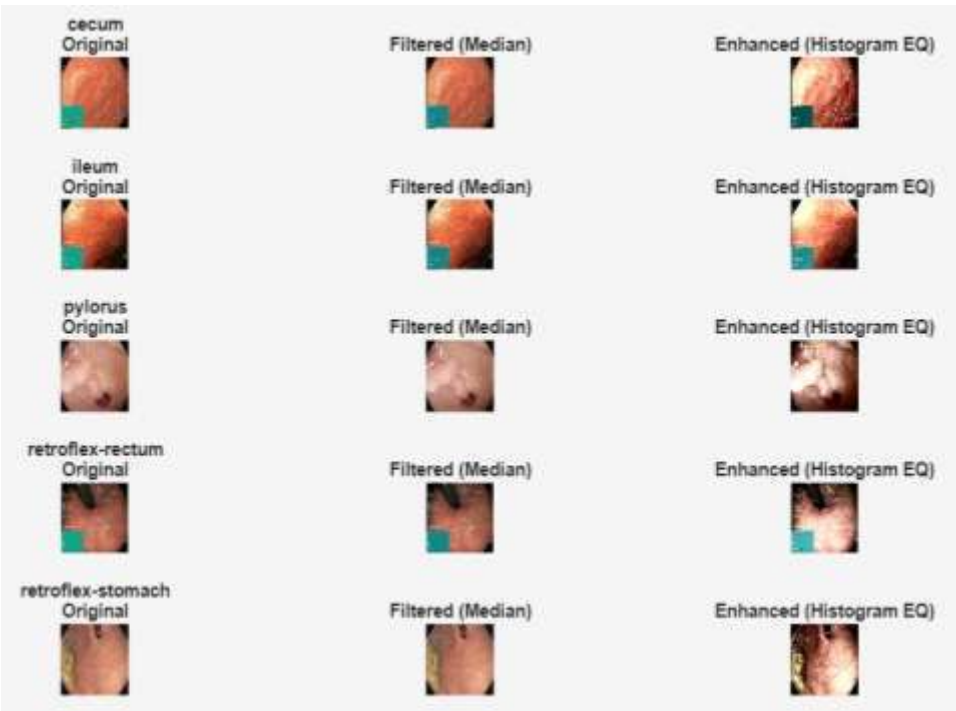


Fig. 2. Output pre-Processing

Figure 4 shows the classification results of the disease, showing the proposed AUTOEUS model has detected the adenocarcinomas (such as gastric and rectal adenocarcinoma) in different regions of the anatomy. The predicted and corresponding true labels in the figure demonstrate the model's high accuracy of detection while exhibiting minimal misclassification.

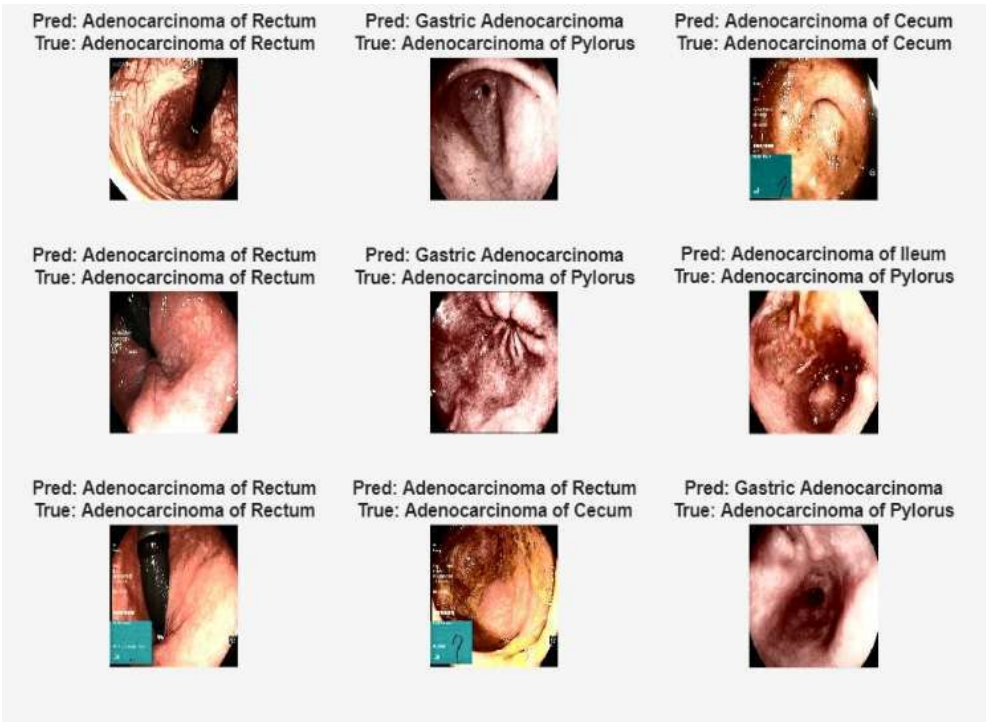


Fig. 3. Output images

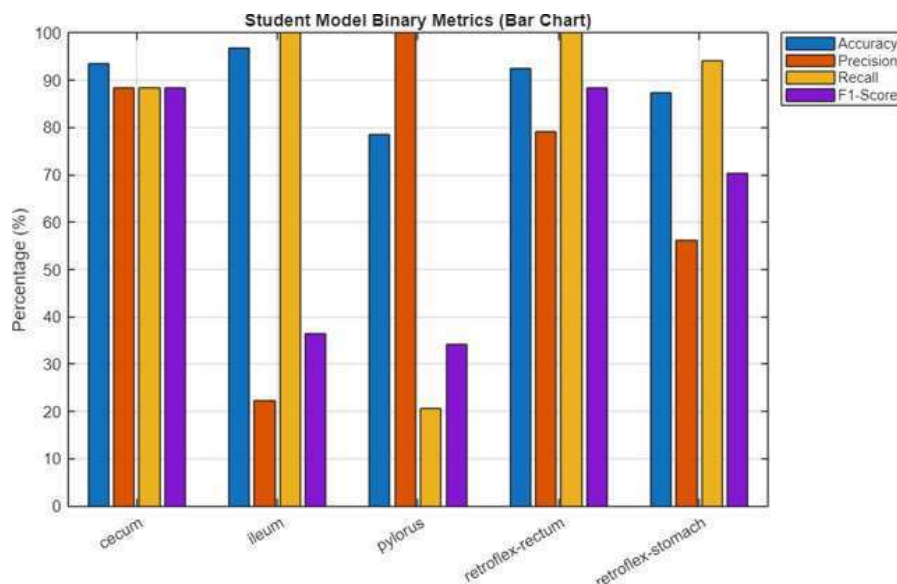
Table I shows the performance comparison of the proposed lightweight CNN (student model) that is trained through knowledge distillation from a ResNet-50 teacher model. The student model achieved comparable accuracy to the teacher model but with a significant reduction in computational load, demonstrating its efficiency for real-time applications.

Table I: Performance Comparison

Class	TP	TN	FP	FN	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)
Adenocarcinoma of Cecum	51	144	11	9	90.70	90.70	82.26	85.00
Adenocarcinoma of Ileum	0	206	7	2	95.81	95.81	0.00	0.00
Adenocarcinoma of Pylorus	16	156	1	42	80.00	94.12	27.59	42.67
Adenocarcinoma of Rectum	58	136	18	3	90.23	90.23	76.32	95.08
Gastric Adenocarcinoma	33	161	20	1	90.23	90.23	62.26	97.06

Figure 5 illustrates the accuracy, precision, recall, and F1-score performance metric plots for each anatomical class. The accuracy of the model for anatomical classes reached between 80% and 95%

level, with well-balanced and reliable classification performance levels seen in the F1-score performance metric for both the cecum of the large intestine (84.55%) and retroflex-rectum (85.29% F1-score).

**Fig. 4 Performance Comparison graph**

CONCLUSION AND FUTURE SCOPE

In this study, a lightweight deep learning application, AUTOEUS, was created for efficient EUS image classification. The application uses median filtering and histogram equalization as preprocessing methods, and a knowledge distillation-based CNN architecture that improves visual quality while computationally reducing model complexity. The teacher-student model step took place with a ResNet-50 model serving as the "teacher" and the compact CNN serving as the "student." This method of knowledge transfer allowed for efficient classification in a much less complex model than one would get if they retrained a

ResNet-50 for classification. AUTOEUS demonstrated overall accuracy above 89%, with the highest per-class accuracy of 95.35% and an F1-score of 85.29%. These results confirm the robustness and efficacy of this framework in medical imaging tasks, especially when considering a resource-limited or real-time clinical diagnostic pathway. In future work, we would expect to extend AUTOEUS with attention mechanisms and transformer-like architectures. Additionally, adding ways to make the final model more interpretable through XAI will be valuable for clinical decision support. It may also be beneficial to expand on our algorithms' datasets while also

incorporating multi-modal imaging could improve clinical generalization and usability even further.

REFERENCE

1. C. Yan and G. Guohua, "Application of a direct image enhancement method in medical image classification," *Comput. Appl. Softw.*, vol. 6, pp. 26–32, Mar. 2007.
2. L. Bo, C. Peg, L. Wei, and Z. Dazhe, "Medical image classification based on multi-feature fusion in scale space," *Comput. Appl.*, vol. 33, no. 4, pp. 1108–1114, 2013.
3. H. Jinmei and L. Zuoyong, "Medical image segmentation based on improved mathematical morphology algorithm," *Comput. Simul.*, vol. 28, no. 5, pp. 299–302, 2011.
4. W. Li, L. Huijuan, Y. Minchao, and Y. Ke, "A cancer image detection method based on Faster-RCNN," *J. China Univ. Metrol.*, vol. 29, no. 2, pp. 136–141, 2018.
5. R. Bakalo, J. Goldberger, and R. Ben-Ari, "Weakly and semi supervised detection in medical imaging via deep dual branch net," *Neurocomputing*, vol. 421, pp. 15–25, Jan. 2021.
6. F. An and J. Liu, "Medical image segmentation algorithm based on multilayer boundary perception-self attention deep learning model," *Multimedia Tools Appl.*, vol. 80, pp. 15017–15039, Feb. 2021.
7. S. Pradeep and P. Nirmala Devi, "A review on speckle noise reduction techniques in ultrasound medical images based on spatial domain, transform domain and CNN methods," in *Proc. IOP Conf., Mater. Sci. Eng.*, vol. 1055, no. 1, 2021, Art. no. 012116.
8. A. H. Masquelin, N. Cheney, C. M. Kinsey, and J. H. T. Bates, "Wavelet decomposition facilitates training on small datasets for medical image classification by deep learning," *Histochem. Cell Biol.*, vol. 155, no. 2, pp. 309–317, Feb. 2021.
9. S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards realtime object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.
10. L. Huilan and Y. Hui, "Image classification method based on iterative training and ensemble learning," *Comput. Eng. Des.*, vol. 41, no. 5, pp. 1301–1307, 2020.
11. W. Hao, Z. Ye, S. Honghai, and Z. Jingzhong, "Overview of image enhancement algorithms," *Chin. Opt.*, vol. 10, no. 4, pp. 438–448, 2017.
12. W. Teng, B. Leping, Y. Zhonglin, L. Qifeng, and Y. Xuanfang, "A flame recognition method based on image similarity of consecutive frames," *J. Naval Univ. Eng.*, vol. 29, no. 4, pp. 48–52, 2017.

HOW TO CITE: Done. Rama Shanthi*, I. Kullayamma, AUTOEUS: Smart Detection of Gastrointestinal Abnormalities Using Lightweight Deep Learning, *Int. J. Sci. R. Tech.*, 2025, 2 (12), 147-155. <https://doi.org/10.5281/zenodo.17876958>