ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 12 Issue 03 March 2025

A WCE Image Classification on Kvasir Dataset using Hybrid CNN-LSTM

[1] C.V.Chakradhar, [2] Dr. T.Bhaskara Reddy

[1] Research Scholar, Department of CST, S.K.University, Anantapuramu, Andhra Pradesh, India
[2] Professor, Department of CST, S.K.University, Anantapuramu, Andhra Pradesh, India
Email: [2] chakradhar.viswa@gmail.com

***Abstract*— This study suggests a hybrid deep learning model that blends the benefits of LSTM and CNN networks to accurately classify Wireless Capsule Endoscopy (WCE) pictures from the Kvasir dataset to meet the growing need for automated medical image analysis. Enhancing the model's ability to handle intricate visual patterns, the CNN extracts spatial characteristics from the images, while the LSTM records temporal connections between these features. To reduce the impact of unbalanced data we use a weighted loss function and data augmentation. With an accuracy of 97.8%, experimental data show that our suggested model works superior to the state-of-the-art methods now in use. The accuracy and efficiency WCE and gastrointestinal image analysis could be greatly increased by this research.**

***Index Terms*— WCE Image, Kvasir Dataset, CNN-LSTM**

I. INTRODUCTION

Wireless Capsule Endoscopy (WCE) is a less nosy alternative than the traditional endoscopy and a diagnostic procedure that involves swallowing a small, A camera-equipped pill capsule records thousands of high-resolution pictures as it passes through the digestive tract and wirelessly sends them to a patient's wearable recording device.These images are then downloaded and analyzed by a physician to diagnose various gastrointestinal conditions, such as obscure bleeding, Crohn's disease, and small bowel tumours.WCE provides a comprehensive view of the small intestine, allowing the physicians to make more accurate diagnoses and improve more effective treatment plans. Here, a study proposes a novel deep learning-based approach to streamline WCE image analysis.

Kvasir's dataset is an extensive compilation of labelled GI images from WCE it provides a valuable resource for training and evaluating training and evaluating and has set a benchmark for deep learning models, particularly in tasks like object detection , image classification , Segmentation and contributing to timely and accurate diagnoses. It consists of high-quality images and videos captured during endoscopic examinations, covering a wide range of GI abnormalities such as polyps, esophagitis, ulcers, and normal mucosa. Its availability and comprehensiveness make it a critical tool in advancing computer-aided diagnostics in gastroenterology.

Both the Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) networks are essential for the diagnosis of gastrointestinal (GI) disorders and Wireless Capsule Endoscopy (WCE).

CNNs excel in feature extraction from WCE images by identifying patterns such as polyps, ulcers, or bleeding with high accuracy. Meanwhile, LSTM networks, designed to handle sequential data, are well-suited for analyzing video frames from WCE, capturing temporal dependencies critical for diagnosing motility disorders or tracking lesion progression.

Deep Convolutional Neural Networks (CNN) is very good at classifying medical pictures because they are very good at extracting spatial characteristics from images.

In this study, Combining CNNs for spatial feature learning with LSTMs for temporal sequence analysis provides a robust framework for improving GI diagnostics, enhancing both accuracy and efficiency in clinical settings, In order to improve feature extraction for wireless capsule endoscopy (WCE) image classification, a hybrid CNN-LSTM architecture was used.

This research emphasizes developing a robust and efficient system for WCE image analysis, which could enhance patient outcomes and reduce the diagnostic workload for healthcare professionals.

II. LITERATURE REVIEW

In order to gain a deeper understanding of gastrointestinal (GI) tract diseases and polyp detection, an extensive research effort was conducted; involving a thorough analysis of various research papers, previously developed systems, and currently implemented technologies. This literature review categorizes studies according to key technological advancements and methodological approaches in GI disease diagnosis.

There are many number of researches proposed sophisticated architecture that combine CNN with LSTM, which improves the precision of GI tract disease diagnosis.

For instance, a residual LSTM-layered CNN architecture was proposed to classify GI tract diseases, leveraging residual connections to enhance feature extraction and temporal sequence learning, which significantly improved diagnostic accuracy and computational efficiency [1]. Öztürk and

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 12 Issue 03 March 2025

Özkaya further refined this approach by developing an LSTM-based CNN model, optimizing the integration of spatial and temporal patterns, and achieving high classification performance across multiple datasets [2].

Another notable study developed a hybrid CNN-LSTM model to classify wireless capsule endoscopy (WCE) images as either bleeding or normal. The CNN component extracted spatial features, while the LSTM captured temporal dependencies, outperforming traditional WCE classification techniques in diagnostic accuracy [3]. Likewise, a graph convolutional neural network (GCNN) for weakly supervised anomaly localisation in capsule endoscopy movies was proposed by Adewole et al.

This method utilized spatial relationships between image regions to localize abnormalities, thereby reducing the need for extensive labelling and expediting the diagnostic process [5].

Furthermore, a number of research highlighted the significance of poorly supervised and self-supervised learning strategies. Pascual et al. introduced a time-based self-supervised learning method for WCE image analysis, leveraging temporal patterns in unlabeled data to train models effectively [10]. They further explored dilated CNNs for WCE abnormality detection, highlighting their ability to capture broader contextual information without increased computational costs [11]. This methodology facilitated efficient diagnostic processes while reducing dependency on labelled data [12].

Finally, other works explored complementary aspects of GI imaging. For example, Goel et al. demonstrated the significance of colour space selection in improving WCE abnormality detection [9]. Additionally, studies investigating LSTM-based localization techniques for WCE highlighted the potential of temporal sequence learning in enhancing spatial positioning within the GI tract, aiding clinical interpretation [8]. These diverse approaches collectively advance the field of GI disease detection and contribute to more accurate, efficient, and automated diagnostic solutions.[13][14]

III. METHODOLOGY

3.1 foundation and context of the Proposed Method

The CNN-LSTM architecture leverages CNNs to extract spatial features from images, while LSTMs process these features to capture sequence dependencies. This dual approach suits the Kvasir dataset, where CNNs capture distinct visual features of various gastrointestinal diseases, and LSTM layers aggregate these features over multiple image frames to reinforce classification accuracy across categories.

3.2 Proposed Method Overview

The Kvasir dataset, characterized by diverse gastrointestinal diseases, presents a unique challenge for image classification. The CNN-LSTM architecture is used to address this.

While LSTM's efficiently capture temporal dependencies within a sequence, CNN are superior extracting spatial characteristics from images. This synergy allows the model to learn discriminative visual features from individual images and then utilize LSTM layers to integrate these features across multiple frames, enhancing classification accuracy for different disease categories.

This study investigates the effectiveness of various CNN architectures, such as AlexNet and ResNet, as feature extractors in conjunction with LSTM layers. By incorporating LSTM layers that process the outputs of the CNN's pooling layers, the model can better capture the sequential relationships present in the image data. This approach aims to improve classification accuracy across the imbalanced and diverse classes within the Kvasir dataset. In order to capture temporal dependencies within WCE image sequences, the suggested model integrates long short-term memory networks (LSTMs) with CNNs for spatial feature extraction.

This hybrid approach is designed to improve the classification accuracy of WCE data by effectively learning both spatial and temporal patterns.

3.3 Parameter Setting

Key parameters include CNN kernel size, number of LSTM units, dropout rate, and learning rate adjustments. Using the Kvasir dataset, various CNN architectures are tested, and specific settings are optimized to handle data imbalance and maintain model generalizability across multiple classes.

Table 1: Parameters Used in Proposed Model.

Hyper parameter	Value
Activation Function	Softmax
Cost Function	Categorical Entropy
Learning Rate	0.001
Optimizer	Adam
Epochs	50
Dropout Ratio	0.2
Batch Size	32
Training Callbacks	Early Stopping

3.3.1. Dataset Preparation

Preprocessing:

- Resize the images to a fixed size (e.g., 224x224).
- Normalize the pixel values to the range [0, 1].
- Augment the dataset with transformations (rotation, flipping, brightness/contrast adjustments, and zooming) to improve generalization and increase data diversity.
- Separate the dataset into test, validation, and training sets (for example, 70%, 15%, and 15%).
- Consider the temporal relationship between frames, especially for dynamic processes like polyp removal or

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 12 Issue 03 March 2025

bleeding.

3.3.2. Model Architecture

Feature Extraction (CNN):

- We use a pre-trained CNN backbone (e.g., ResNet50, VGG16, or EfficientNet) for transfer learning.
- To extract high-level spatial characteristics, eliminate the last classification layers.
- Adjust the Kvasir dataset's pre-trained layers to better fit its domain.

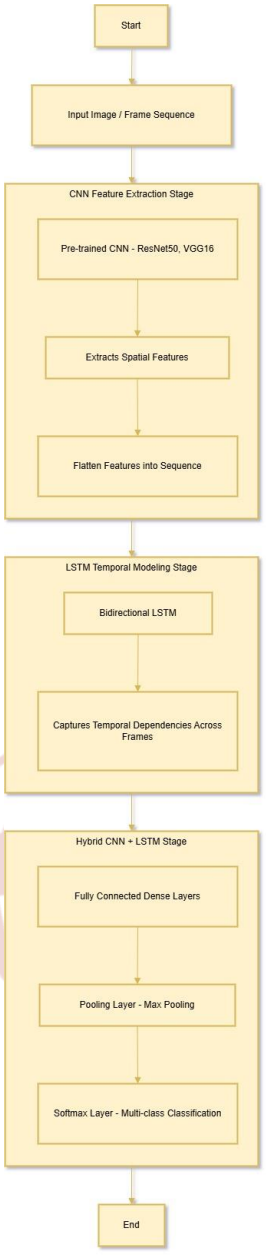

Fig. 1 Flow Diagram of Multi Classification in WCE

Temporal Modeling (LSTM):

- CNN-extracted flatten spatial features into a series.
- Pass the sequence into a bidirectional LSTM layer to capture sequential patterns across frames or image sets.
- To extract pertinent characteristics from every frame, use numerous convolutional layers with the right kernel sizes and filter numbers.

Pre-trained CNNs and Data Augmentation:

- Leverage pre-trained CNN models, utilizing knowledge from large datasets to improve performance.
- To improve model generalisation and diversify training data, use data augmentation approaches.

Classification Layer:

- Concatenate the final outputs of the LSTM.
- Use fully connected (dense) layers with ReLU activations.
- to reduce spatial dimensions and computational cost, we will apply pooling layers(Max Pooling)
- We the Use softmax activation in the last layer for multi-class classification.

3.3.3. Implementation Steps

1. **Input Layer:**

Input shape: (sequence_length, height, width, channels), e.g., (5, 224, 224, 3).

2. **CNN Backbone:**

- Determine the spatial characteristics of every picture in the series.
- Output shape: (sequence_length, feature_dim), where feature_dim is the number of CNN output features.

3. **LSTM Layer:**

- Process the temporal sequence using one or more bidirectional LSTM layers.Example configuration: 256 hidden units, dropout = 0.3.
- Feed the output of the CNN layers into LSTM layers to capture long-range dependencies between frames within a sequence.

4. **Fully Connected Layers:**

- Regularization is achieved by dropout after dense layers with 128 and 64 neurons.
- An output layer with as many neuron's as the Kvasir dataset's classes.
- To map the learnt features to the classification problem, connect the LSTM layers output to the fully connected layers for multi-class image classification, use the proper activation functions (softmax, for example).

5. **Loss Function:**

Use categorical cross-entropy for multi-class classification.

6. **Optimizer:**

We use the Adam optimizer with a learning rate 0.001 all the values of the hypermeter are showed in the above table-1.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 12 Issue 03 March 2025

3.3.4. Training Strategy

Batching:

Use mini-batches with sequences of images (e.g., sequences of 5 images from a video or neighboring image in the dataset).

Early Stopping:

After a few epochs, if there is no progress in validation accuracy, cease training.

Evaluation Metrics:

Precision, Recall, F1 Score, and Accuracy.

3.3.5. Loss Function and Optimization

Cross-Entropy Loss:

- Calculate the difference between the true and expected class probabilities.
- Optimize the model's parameters using an efficient optimizer like Adam or SGD with momentum.

3.3.6. Advantages

- **CNN:** Strong spatial feature extraction for individual images.
- **LSTM:** Captures sequential dependencies, improving classification for temporally correlated images (e.g., frames in a video).
- A CNN-LSTM hybrid architecture effectively captures both spatial and temporal information within the image sequences.

3.3.7. Training and Validation

Data Split:

- Separate the dataset into sets of testing, validation and training
- Use the training set to update model parameters.
- To avoid overfitting, keep an eye on performance on the validation set.
- Assess how well the finished model performs on the testing set.

Transfer Learning:

Leverage pre-trained models (e.g., RESNET, VGG) as feature extractors to improve performance and reduce training time.

Data Balancing:

Use strategies like class weighting, undersampling, or oversampling to address issues of class imbalance.

- **Regularisation:**

Use strategies (such as dropout and L1/L2 regularisation) to enhance generalization and avoid overfitting.

- **Hyper Parameter Tuning:**

To maximize performance, experiment with various hyper parameters (such as learning rate, batch size, number of layers, and number of units).

By carefully considering these aspects and fine-tuning the architecture, the proposed CNN-LSTM hybrid model can achieve robust and accurate multi-image classification on the Kvasir dataset.[9]

IV. EXPERIMENTS AND EXPERIMENTAL RESULTS

Experiments demonstrate that CNN-LSTM models perform well on the Kvasir dataset, achieving higher classification accuracy than standalone CNN models. Metrics such as F1-score indicate that the LSTM layers contribute significantly to balancing accuracy across both common and rare classes, addressing challenges typical in medical image datasets.[7]

The evaluation of a model's performance involves several key metrics that provide insights into its effectiveness. [15][1]

The proposed method is executed on a system featuring an Intel Core i7-7700K CPU (4.2 GHz), 32 GB of DDR4 RAM, and an NVIDIA GeForce GTX 1080 graphics card. To assess the effectiveness of the suggested framework, three prominent CNN architectures—AlexNet, GoogLeNet, and ResNet50—are employed. [2]

These CNN models are applied through a transfer learning method to enhance the feature extraction component. Throughout the experiments, the three architectures are utilized with their standard settings and without altering the original layers. Only the specified LSTM blocks are incorporated. To evaluate how the sample size in the dataset impacts the proposed framework, three datasets of 2000, 4000, and 6000 samples are generated.[3][4]

The visuals in these datasets are chosen at random from the primary dataset. A total of 27 experiments were conducted using three different datasets along with three distinct CNN architectures, which include ANN, SVM, or our LSTM block. To assess our experimental results, we utilize six different metrics. The metrics measured are sensitivity, specificity, accuracy, precision, and F1-score.[5][6]

Table 2: Metrics Used in Proposed Model.

S.no	Metric	Equation
1	Accuracy (ACC)	(TP+TN)/(TP+TN+FN+FP)
2	Precision	TP/(TP+FP)
3	Recall	TP/(TP+FN)
4	F1-Score	$2 * \frac{Precision.Recall}{precision + Recall}$

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 12 Issue 03 March 2025

Table 3: Classification performance of proposed model with the other models

S.NO	MODEL	Precision	Recall	F1score	Accuracy
1.	Residual LSTM layered CNN for Classification of gastrointestinal tract diseases [1].	98.05	98.05	98.05	98.05
2.	Gastrointestinal tract classification using improved LSTM based CNN[2]	94.46	96.37	93.54	97.90
3.	Hybrid CNN-LSTM Model for the classification of wireless capsule endoscopy images for bleeding or Normal Diagnosis[3].	90	80	85	90
4.	Proposed Model	98.46	96.44	97.44	97.80

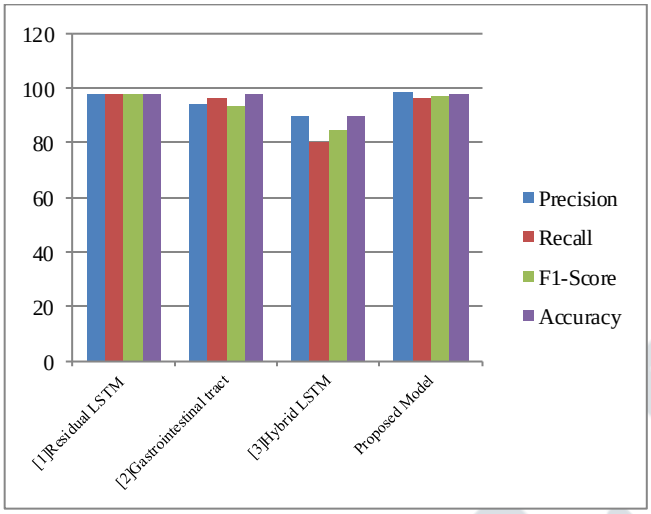

Fig. 2: Graphs for CNN+LSTM Proposed model with other models

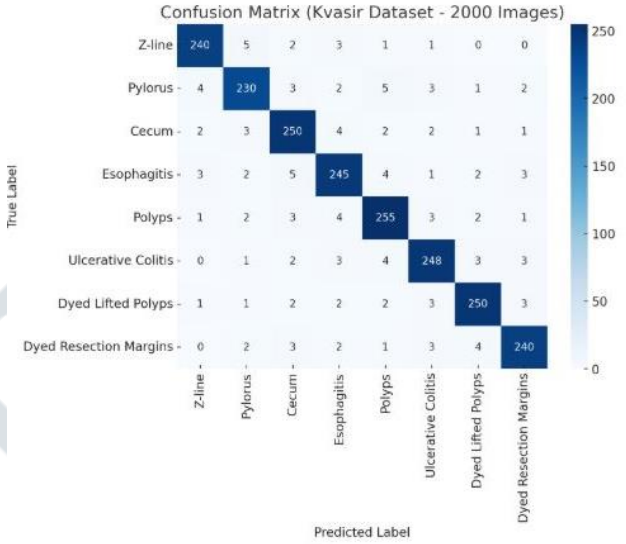

Fig. 3: Confusion Matrix For Proposed Model on Kvasir Dataset

V. CONCLUSIONS

By the use of this Hybrid CNN+LSTM for multi class classification in kvasir dataset, highlighting their capability of integrity of spatial and temporal data for improved accuracy. Using this Hybrid Deep Learning Model, we have seen the following features:

Class imbalance: While the paper addresses class imbalance through data augmentation and weighted loss functions, the inherent variability in the Kvasir dataset may still pose challenges. Certain categories may have significantly fewer samples, which can affect the model's performance on underrepresented classes **[1]**.

Limited Sample Size: The Kvasir dataset, although comprehensive, has limitations in sample size for some categories. A model with this limitation may not generalize well across all classes, especially for rare GI abnormalities.

Quality of images: A WCE image's quality and content variability can complicate disease detection. There is a possibility that the model's functionality could be affected by Low-quality or noisy images, which are common in real-world situations [22].

Fig. 4: Confusion Matrix For Proposed Model on Kvasir Dataset

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 12 Issue 03 March 2025

Static vs temporal analysis: Despite the hybrid model's effective combination of CNNs for spatial feature extraction and LSTMs for temporal dependencies, it may still struggle with sequential data. It is likely that the model will fall short of capturing all relevant patterns due to its reliance on the quality of the temporal data [23] [4].

Computational Complexity: Because the hybrid model combines CNN and LSTM, its architecture may require more computing power. This can be a hurdle to implementation in resource-constrained contexts, where faster, less resource-intensive models may be desirable.[21]

Generalizability: The Kvasir dataset may not adequately represent the range of WCE images found in clinical practice, limiting the model's generalisability. This raises issues over the model's capacity to generalise to different datasets or real-world applications.[15]

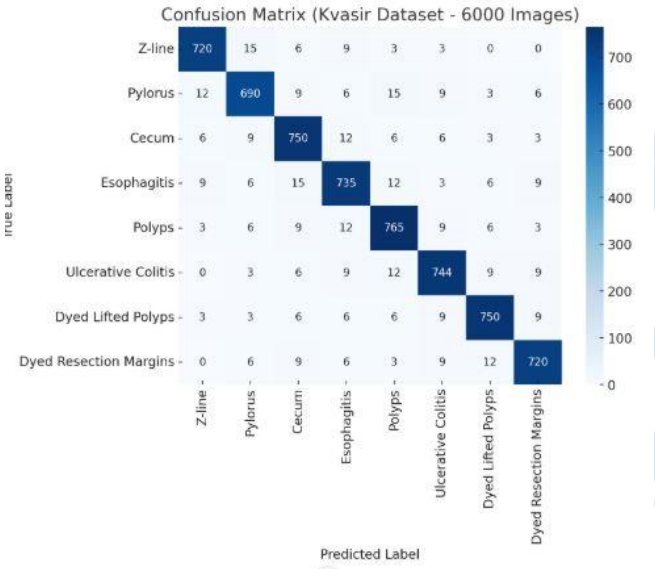

Fig. 5: Confusion Matrix For Proposed Model on Kvasir Dataset

- This study presents a novel deep learning approach for efficient WCE image analysis.the proposed hybrid CNN-LSTM model shows promise for increasing detection efficiency and accuracy of GI Tract Disease. [16][17]
- In addition to improving the effectiveness and efficiency of colorectal cancer diagnostic tools, this research makes them less invasive and more accurate.
- The prevalence of GI Tract Disease and the limitations of colonoscopy highlight the need for efficient diagnostic tools. Although Capsule endoscopy is a viable option, analyzing capsule endoscopy takes times. [18]
- This study suggests a method based on deep learning to address this challenge by leveraging hybrid CNN-LSTM architecture for WCE image classification on the Kvasir dataset. [19]
- The Proposed model combines the spatial feature extraction capabilities of CNN's with temporal dependence modeling of LSTM's offers a robust and efficient solution for improving the accuracy and efficiency of GI endoscopy image analysis.[20]

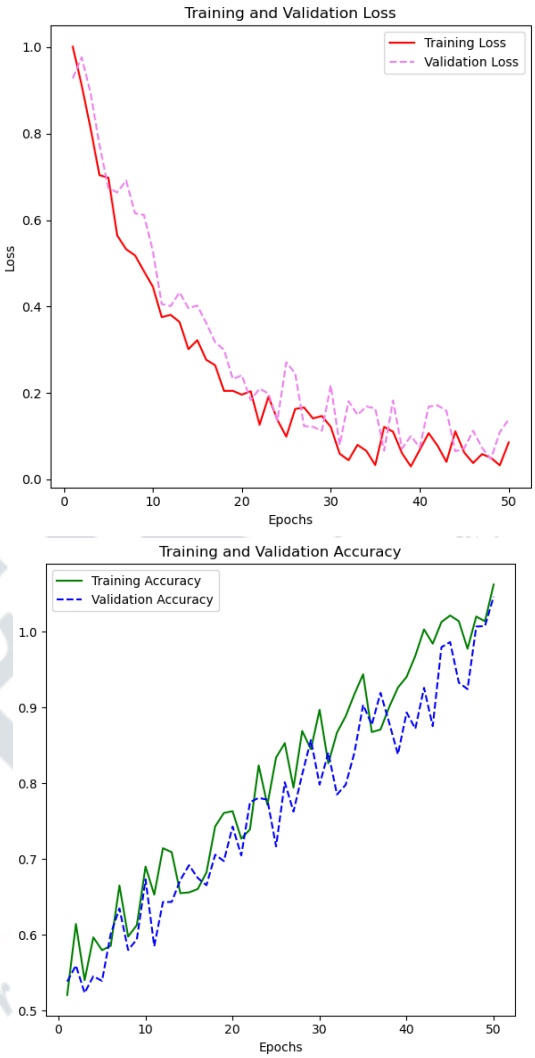

Fig.6: Training and Validation Accuracy and Loss for Proposed Method on Kvasir Dataset

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES

[1] Residual LSTM layered CNN for classification of gastrointestinal tract diseases S¸ aban Oztürk ¨ a,*, Umut Ozkaya ¨ b a Amasya University, Technology Faculty, Electrical and Electronics Engineering, Amasya 05100, Turkey b Konya Technical University, Engineering and Natural Science Faculty, Electrical and Electronics Engineering, Konya, Turkeyhttps://doi.org/10.1016/j.jbi.2020.103638Journal of Biomedical Informatics 113 (2021) 103638

[2] Ozturk S, Ozkaya U: Gastrointestinal tract classification using improved c LSTM based CNN. Multimed. Tools Appl. 2020; 79(39–40): 28825–28840. Publisher Full Text

[3] Hybrid CNN-LSTM Model For The Classification Of Wireless Capsule Endoscopy Images For Bleeding Or Normal Diagnosisdivya Bharathi.P1, Ramachandran.M2, Rajkumar.M3, Rajkumar.R. K4International Research Journal of Engineering and Technology (IRJET) eVolume: 11 Issue: 04 | Apr 2024p-ISSN: 2395-0072

[4] Comparative Study of CNN and LSTM based Attention Neural Networks for Aspect-Level Opinion Mining Wei Quan∗, Zheng Chen∗, Jianliang Gao† Xiaohua Tony Hu∗2018 IEEE International Conference on Big Data (Big Data)

[5] Adewole S, Fernandes P, Jablonski J, et al.: Graph Convolutional Neural Network For Weakly Supervised Abnormality Localization In Long Capsule Endoscopy Videos. Piscataway: IEEE; 2021; pp. 388–399. Reference Source

[6] Mohammed A, Farup I, Pedersen M, et al.: PS-DeVCEM: Pathologysensitive deep learning model for video capsule endoscopy based on weakly labeled data. Comput. Vis. Image Underst. 2020; 201: 103062. Publisher Full Text|Reference Source

[7] Review of Deep Learning Performance in Wireless Capsule Endoscopy Images for GI Disease Classification Tsedeke Temesgen Habe, Keijo Haataja, Pekka Toivanen https://doi.org/10.12688/f1000research.145950.2 F1000Research 2024

[8] Wireless Capsule Endoscope Localization Using LSTM Network,umma Hany,/nafe Muhtasim Hye,Lutfa Akter,IEEE Sensors Letters, VOl 7 Issue 12, 2023 , DOI: 10.1109/LSENS.2023.3330401

[9] Goel N, Kaur S, Gunjan D, et al.: Investigating the significance of color space for abnormality detection in wireless capsule endoscopy images. Biomed. Signal Process Control. 2022; 75: 103624. Publisher Full Text

[10] Time-based self-supervised learning for Wireless Capsule EndoscopyGuillem Pascual a,* , Pablo Laiz a , Albert García a , Hagen Wenzek b , Jordi Vitria` a , Santi Seguí ahttps://doi.org/10.1016/j.compbiomed.2022.105631 Computers in Biology and Medicine 146 (2022) 105631

[11] Dilated CNN for abnormality detection in wireless capsule endoscopy images, Data analytics and machine learning , jan-2022,VOl 26, pg 1231-1247, Nidhi Goel,SamarjeetKaur, DeepakGunjan &S.J.Mahapatra https://link.springer.com/article/10.1007/s00500-021-06546-y

[12] Pascual G, Laiz P, García A, et al.: Time-based self-supervised learning for Wireless Capsule Endoscopy. Comput. Biol. Med. 2022; 146: 105631. PubMed

[13] X. Zhang, F. Chen, T. Yu, J. An, Z. Huang, J. Liu, W. Hu, L. Wang, H. Duan, J. Si, Real-time gastric polyp detection using convolutional neural networks, PLoS ONE 14 (2019) e0214133.

[14] Y. Ma, X. Chen, B. Sun, Polyp Detection in Colonoscopy Videos by Bootstrapping Via Temporal Consistency, in: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), 2020, pp. 1360–1363.

[15] J.-J. Wan, T.-Y. Chen, B.-L. Chen, Y.-T. Yu, Y.-Y. Sheng, X.-G. Ma, A polyp detection method based on FBnet, Comput. Mater. Continua 63 (2020) 1263–1272. [16] A.A. Pozdeev, N.A. Obukhova, A.A. Motyko, Automatic analysis of endoscopic images for polyps detection and segmentation, IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus) 2019 (2019) 1216–1220.

[16] Y. Shin, H.A. Qadir, L. Aabakken, J. Bergsland, I. Balasingham, Automatic colon polyp detection using region based deep CNN and post learning approaches, IEEE Access 6 (2018) 40950–40962.

[17] J.M. Fitzpatrick, Z. Wang, M. Sonka, L. Li, J. Anderson, D.P. Harrington, Z. Liang, Computer-aided detection and diagnosis of colon polyps with morphological and texture features, Med. Imag. 2004: Image Processing (2004).

[18] S. Hwang, J. Oh, W. Tavanapong, J. Wong, P.C. de Groen, Polyp Detection in Colonoscopy Video using Elliptical Shape Feature, in: 2007 IEEE International Conference on Image Processing, 2007, pp. II - 465-II - 468.

[19] L. Zhao, C. Botha, J. Bescos, R. Truyen, F. Vos, F. Post, Lines of curvature for polyp detection in virtual colonoscopy, IEEE Trans. Visual Comput. Graphics 12 (2006) 885–892.

[20] T.A. Chowdhury, O. Ghita, P.F. Whelan, A statistical approach for robust polyp detection in CT colonography, in: 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference, 2005, pp. 2523–2526.

[21] Z. Qian, M.Q.H. Meng, Polyp detection in wireless capsule endoscopy images using novel color texture features, in: 2011 9th World Congress on Intelligent Control and Automation, 2011, pp. 948-952.

[22] B. Li, M.Q.H. Meng, Automatic polyp detection for wireless capsule endoscopy images, Expert Syst. Appl. 39 (2012) 10952–10958.