Automated Detection and Classification of Gastrointestinal Bleeding in Wireless Capsule Endoscopy

1st Sharika Punjabi
AI & DS Department
Indira Gandhi Delhi Technical University for Women
Delhi, India
sharika177btcseai23@igdtuw.ac.in

2nd Suhani Goyal
AI & DS Department
Indira Gandhi Delhi Technical University for Women
Delhi, India
suhani201btcseai23@igdtuw.ac.in

Abstract: WCE is a non-invasive innovative diagnostic tool of the GI tract, showing some advantages compared with traditional endoscopy, such as preventing discomfort and decreasing complications. However, manual analysis of WCE images is time-consuming and prone to errors. In this context, the paper proposes a deep learning approach of CNNs for automatic bleeding frame detection and classification from WCE images. Thus, the primary aim of this research is to improve this model through a combination of activation functions, layers, optimizers, andnormalizations. The objective is to achieve the highest accuracy along with its precision and recall level, the issues of large image set and lack of dependable and efficient automatic diagnosis system. Algorithms generated from experimental results show that the proposed model enhances classification performance, thus it is a potential solution for assisting healthcare professionals in accurate diagnosis of gastrointestinal disorders.

Keywords - wireless capsule endoscopy, machine learning, deep learning, computer vision, gastrointestinal tract infection, classification, convolutional neural network

I. Introduction

Endoscopy

It is the procedure generally carried out by long thin tubes and allows us to do imaging, surgery and other tasks. Using an endoscope, a flexible tube with a light and camera attached to it, your doctor can view pictures of your digestive tract on a colour TV monitor. During an upper endoscopy, an endoscope is passed through your mouth and throat and into your oesophagus, allowing the doctor to view the oesophagus, stomach, and upper part of the small intestine. Similarly, endoscopes can be passed into your large intestine (colon) through the rectum to examine this area of the intestine. This procedure is called sigmoidoscopy or colonoscopy depending on how far up the colon is examined. In some cases endoscopy is also performed using the small insertions in the abdomen. But there are drawbacks of endoscopy. The first one is contamination, infection, bleeding, perforation etc leading to discomfort and long recovery period. Second, making it a very hard procedure for people with lung disorders , heart disease and elderly. Third, it does not allow a thorough examination of the small intestine.

Small intestine examination can also be done using MRE (Magnetic resonance enterography). This is the non-invasive procedure which produces the images in a magnetic field when a contrasting drug is administered in the organ. Mri ids very helpful in imaging inflammations, bleeding, and obstructions. But this procedure poses a hindrance for the patients who have metallic implants or cardiac defibrillators. In addition it can't diagnose obscure lesions, as it is not sensitive enough to detect blood loss caused by small, undetectable lesions caused by iron-deficiency anaemia. This is where Wireless capsule endoscopy comes in. It is a modern technique which allows the diagnosis of small bowel very closely and precisely. By using CE (Capsule Endoscopy), one can see the small intestinal mucosa and diagnose lesions such as ulcers. Due to its ability to visualise the entire small bowel mucus, CE is very sensitive when diagnosing small bowel pathologies. WCE is performed using a pill sized capsule fitted with camera and LED,

being swallowed. This capsule passes out of the body after 2 to 3 bowels.

Wireless Capsule Endoscopy You swallow a capsule that contains a tiny camera, a transmitter and a light. As it passes through your stomach, intestines, colon and rectum, the capsule takes thousands of pictures and transmits them to a recorder that you wear outside of your body.

The quality of images obtained using capsule endoscopy has improved progressively to being comparable with those obtained using conventional endoscopy. The depth of view and luminosity control improves from one generation to another and for every brand of capsule.

The number of images seems to be important for assessing sensitivity, but increasing the number of images or using a double-tip capsule, such as the colonic capsule, leads automatically to an important increase in reading time, which is difficult for physicians.

One solution could be a motion sensor, probably because most failures of capsule endoscopy to identify significant lesions are related to rapid transit of the capsule through some digestive segments.

The future probably lies in automatic algorithms for specific lesion detection. A highly promising algorithm is the quick-view reading tool of Given Imaging. New algorithms will certainly be developed with a possibility that a computer will propose a series of 5 to 10 diagnoses, from among which the gastroenterologist will simply have to choose the right one. As we have established the importance of automated systems in diagnosing the images captured by capsule endoscope, the need of CNN model arises which processes large data of images.

II. Related Work

Many efforts have been made towards automatic GI bleeding detection assisted by WCE, focusing on both sensitivity and the reduction of diagnostic errors. Several prior works utilised CNN architectures on feature extraction and classification tasks, which have rather modest performances. Prior related work, such as Luo et al. (2020), explored promotes generalised comparison with existing optimizations in CNN, whereas Zhang and Zheng (2016) focused on the designs of convolutional layers for image processing. Techniques like class balancing and min-max scaling have been widely discussed in literature for enhancing model performance on imbalanced datasets (Kotsiantis et al., 2006). Existing approaches often struggle with generalizability across varied datasets and the computational complexity of handling high-resolution WCE images. This study builds on these efforts, integrating state-of-the-art datasets, advanced CNN configurations, and loss function optimizations to address existing gaps, delivering improved diagnostic accuracy and efficiency. Figure 1 shows the wireless endoscopic capsule.

III. Proposed Methodology

In this we will discuss the detailed version of the steps taken to design the model. Many combinations of loss, activation function, and CNN layers were tested through the trial and error method to derive the current accuracy of the model. We have later compared our model with the accuracy of other models. This enabled us to analyse our model with different aspects.

I. Data set Selection

For this study we are using the WCEBleedGen dataset by Zenodo.

https://doi.org/10.5281/zenodo.7548319

This particular dataset was selected as it is a wireless capsule endoscopy dataset containing bleeding and non-bleeding frames. The dataset

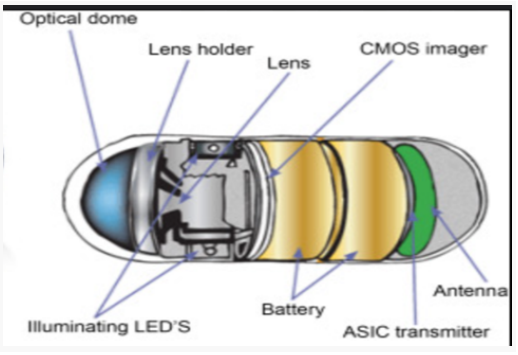

Fig1: Picture of a wireless endoscopic capsule and its various components

uses state-of-the-art methods, and contributes to better interpretability, and reproducibility of such automated systems. It contains two classes of bleeding and non-bleeding, as the bleeding annotated data of WCE is scarcely available; the ulcer, polyps, and red lesions are also included in it. WCEBleedGen is a collective framework of various other data sets. Here is the list of works compiled in WCEBleedGen data set:
Training dataset distribution is shown in Table 1.

Similarly, the test dataset is an independently collected WCE data containing bleeding and non-bleeding frames of more than 30 patients suffering from acute, chronic and occult GI bleeding referred to the Department of Gastroenterology and HNU, All India Institute of Medical Sciences, New Delhi, India.

Validation dataset distribution is shown in Table 2.

Type	Name(Data Source)	Img	Total
Bleeding	Set-2 data	152	1309
	Farah deeba data	50	
	KID	5	
	Kvasir capsule	443	
	Kvasir capsule (hematin blood)	10	
	Self-collected (YouTube video)	292	
	Gastrolab	357	
Non-Bleeding	Set-1 data	500	1309
	Set-2 data	80	
	Self-collected (YouTube video)	181	
	Gastrolab	548	

Table 1: Training Dataset Distribution

A sample of the normal and bleedy(infected) image is shown in Figure 2.

Type	Images	Description
Dataset 1	49	Randomly collected from 7 patients with marked and un-marked annotations
Dataset 2	514	Sequentially collected from 23 patients without marked annotations but known class labels

Table 2: Validation Dataset Distribution

II. Scaling Data

Data range of the images lie between 0-256 in RGB pixel values. We have used min-max scaling method to convert the values from 0-256 to 0-1. In reference with [1] This can have increase the accuracy as it eliminates large disparities in feature magnitudes,

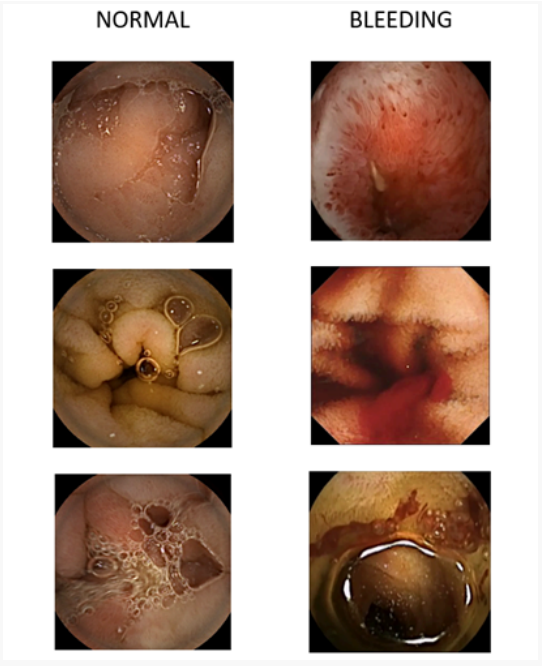

Fig 2: Sample of data for both normal and bleeding class.

allowing the CNN to learn uniformly across all features without bias toward higher-valued ones.Min max scaling also increases the speed of convergence,by ensuring the gradient computed in backpropagation is well scaled, it ensures that the values of gradients do not become overly large or small. Since we will be using CNN model on our given data using min max scaling also prevents the dead neuron or saturation,as CNN layers are sensitive to the input data, and normalised values between 0 to 1 comes under the definition range of various activation functions like Relu, sigmoid,etc. This in turn provide training and computational stability to model as there is no large difference in the values being fed in the computational cycle. Thus min max scaling enhance the model's ability to generalise the unseen data and extract the meaningful features.

III. Model Building

A code snippet is represented in Figure 3.

We have used the Conv2d layer which is the basic driving force behind the CNN model. This gives the dot product of the small region of the input image and the filter (kernel) , this filter slides across the images. The first convo 2D layer in this model extract the basic features like texture and edges from the model.

We are using the stride of 1 , this means our model moves one pixel at a time, and 16 layers of filters is used, in reference to the paper [3], using smaller amount of filters in basic (starting layers) avoids redundancy in feature extraction and reduces the computational complexity and using too many filters at early stages causes the over-fitting in the model, especially when we have a small dataset,

```
model.add(Conv2D(16, (3,3), 1, activation='relu', input_sh
model.add(MaxPooling2D())
model.add(Conv2D(32, (3,3), 1, activation='relu'))
model.add(MaxPooling2D())
model.add(Conv2D(16, (3,3), 1, activation='relu'))
model.add(MaxPooling2D())
model.add(Flatten())
model.add(Dense(256, activation='relu'))
model.add(Dense(1, activation='sigmoid'))
model.compile('adam', loss=tf.losses.BinaryCrossentropy(),
```

Fig3: code snippet for the model architecture

various other CNN like AlexNet also used less filters in their starting layers. We are using Relu as an activation function since it introduces the non- linearity in the relation.

ReLU(x)= max(0, x). It is much simpler and gives faster computation than other activation functions like sigmoid or tanh. ReLU function also avoid vanishing gradient (i.e learning slope for some input values becomes very small, which results in stalled learning or more computation). Figure 4 represents the graph of ReLU function.

The MaxPooling layer is beneficial in a few convolutional layers and aids in the registration of the most important learned features. It helps the network to focus on some higher levels such as shapes, patterns and or objects but not pixel value. It picks up the maximum value of a set of values in a pooling window (for example, 2 x 2, 3 x 3, …). For example if the window being used is 2 * 2 then it picks the maximum pixel intensity value from the 4 pixels that the window covers and outputs this[4]. When using convolutional and pooling layers the end product is a multi dimensional tensor (for example 2D or 3D). The Flatten layer flattens this tensor into a 1D array which is useful to be fed into the fully connected networks (Dense)[5]. Then a dense layer is used where Each neuron computes a weighted sum of the inputs it receives from the previous

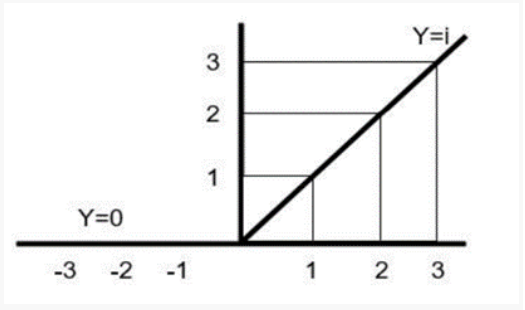

Fig 4 : ReLU function

layer, applies a bias term and passes this value through an activation function (such as ReLU or Sigmoid) to produce an output.It integrates and enhances the features detected by previous layers of convolutional and pooling layers. Since all the neurons of the Dense layer are connected with all the neurons of the previous layer, it can learn more complex representations which can help in decision making for classification of the given input data.The final Dense layer is used to make the final prediction or classification based on the features learned by the previous layers.

We are using Adam optimizer in the compilation of the model Adam as a result of combining two other very effective formulations of weights optimization, AdaGrad and RMSProp. It is the gradient of the momenta and squared gradients in order to control the learning rate.Adam helps to accelerate convergence and makes the process effective when working with big data and deep structures. It is especially beneficial in intricate structures such as CNNs because the learning rate is variable[6]. The loss function is key to training the model as it guides the optimization process. Minimising the binary cross-entropy loss ensures the model's predictions become more accurate over time.

IV. Model Training

The calculation of class weights is something that is very pertinent, especially at the training stage of the model because it modifies the loss function placing greater importance on classes that were less represented ensuring that there is balanced learning. Some classes tend to have fewer samples than others and therefore class imbalance occurs. In situations like this, the training could easily become biassed whereby the model achieves good performance on the majority class input, but very poor performance on the minority class inputs.

class weight=total samples/(number of samples in class)×number of classes

In the case of class weights applied, loss function pays more penalty for the model's mistake towards classifying samples of low-data class [7]. The entire training process is conducted over 15 epochs where in another validation dataset is used to help keep track of how well the model is able to generalise and help avoid overfitting as well. A TensorBoard callback is also added to the trainer in order to log metrics related to progress, and to make the training and validation performance over time more informative and visually pleasing respectively.

V. Model Training

We are employing a distinct testing dataset in order to test the model. A trained model is assessed using a test dataset by traversing its batches and modifying the following performance metrics: precision, recall and accuracy. It uses TensorFlow's as_numpy_iterator() method to perform test data in memory efficient clean up and processes. For each batch, predictions (yhat) are produced by the model using model.predict() and labels (y) are checked against these predictions. After all the batches have been processed, the cumulative metrics indicate how well the model is expected to perform on data that it has not seen during training.

IV. Evaluation Metrics

We are using four types of values in our evaluation metrics let us understand their meaning and values:

- *Accuracy*

The ratio of correctly predicted samples to the total number of samples.High accuracy indicates the model performs well overall. Equation 1 shows the formula:

$$Accuracy = \frac{correct\ predictions}{total\ predictions}$$

We can also define it as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

- *Precision*

The proportion of correctly predicted positive observations to all observations predicted as positive.High precision reduces false alarms but doesn't account for missed positives. Equation 2 shows the formula:

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

- *Recall*

The proportion of actual positives correctly identified by the model.High recall ensures fewer positive cases are missed but doesn't consider false positives. Equation 3 shows the formula:

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

- *F1 Score*

Harmonic mean of precision and recall, balancing the trade-off between the two.F1 is particularly useful when dealing with imbalanced datasets, as it considers both false positives and false negatives. Equation 4 shows the formula:

$$F_1 score = 2 * \frac{Precision * Recall}{Precision + Recall} \quad (4)$$

Here,

● TP is the value when the model predicts the image as Normal and the actual label of the image is also Normal.

● TN is the value when the model predicts the image as Infected and the actual label of the image is also Infected.

● FP is the value when the model predicts the image as Normal but the actual label of the image is Infected.

● FN is the value when the model predicts the image as Infected but the actual label of the image is Normal.

V. Result

Here is the evaluation metric of the model trained. The outcomes of the study are analysed in the context of model assessment metrics. Since we have a balanced data set having high accuracy of the model signifies the excellence of the model. Even though the accuracy is high it is very important to consider the precision and recall especially in the medical diagnosis where false negative and false positive could have significant consequences. The results of the proposed approach with the balanced dataset are shown in Table 3.

Our model has a high precision value which minimises the chances of false positives (misclassifying normal images as bleeding), which is crucial in reducing unnecessary follow-up tests and patient anxiety. Whereas high recall ensures that very few bleeding cases are missed, addressing the risk of false negatives. Missing such cases could delay critical medical intervention, making recall vital in medical applications. A high F1-score (close to 1) indicates the model is reliable in this context, ensuring the right balance between identifying true cases and avoiding overdiagnosis[8].

Evaluation metrics	Score
Accuracy	0.98
Precision	0.984
Recall	0.98
F1 Score	0.982

Table 3: The evaluation metrics of the model

Let us also discuss the learning curve of our model. Learning curve of our model through the epochs provides the insight to the model like accuracy should increase and stabilise over time otherwise the model might be overfitted(i.e. it is only memorising rather than learning to the unseen data).

Table 4 provides the data for loss and accuracy of the model throughout the ten epochs for the model.

One potential solution to the data imbalance problem is to follow the weighted cross-entropy approach where more weight is put on the class with fewer samples. Experiments are performed to evaluate the performance of the model loss function with the dataset. Weighted Cross Entropy is a type of cross entropy loss in which the weights of each class are taken into consideration with a view of levelling the scale of each class.

To train, validate and test the automated detection system of the gastrointestinal bleeding through the WCE images analysis, the following two datasets were employed in the experiments. These datasets were reviewed and marked by specialist gastroenterologists and expanded to increase the model's stability.

Previous research on automatic detection utilises machine learning, as well as, deep learning approaches, yet, the provided accuracy can still be enhanced.

Epochs	Loss	Accuracy
Epoch 1	0.48	0.74
Epoch 2	0.20	0.92
Epoch 3	0.14	0.94
Epoch 4	0.08	0.96
Epoch 5	0.05	0.98
Epoch 6	0.02	0.992
Epoch 7	0.0148	0.995
Epoch 8	0.0103	0.997
Epoch 9	0.0048	0.999
Epoch 10	0.0016	1.000

Table 4 : Epoch vise values of loss and accuracy

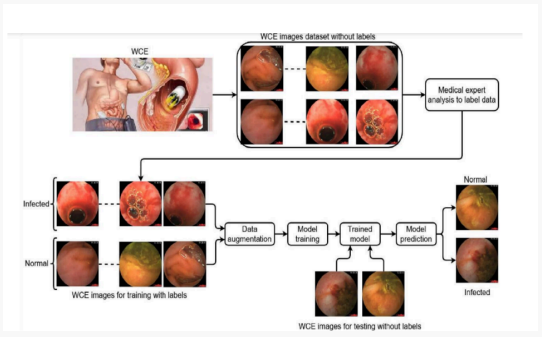

Fig 5 : . Approach of this study for bleedy image classification [8]

We contribute to this classification problem by enhancing the accuracy of the classification model. For this purpose, this desired model is proposed that performs very well for bleedy image classification. Figure 6 represents the loss-accuracy graph of the values mentioned in Table 4. The accuracy and the loss curves that were derived through the codes during model building are represented in Figure 7(a) and 7(b).

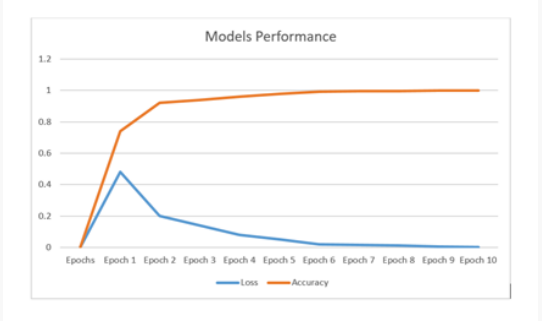

Fig 6 Graphical representation of model performance in form of line chart

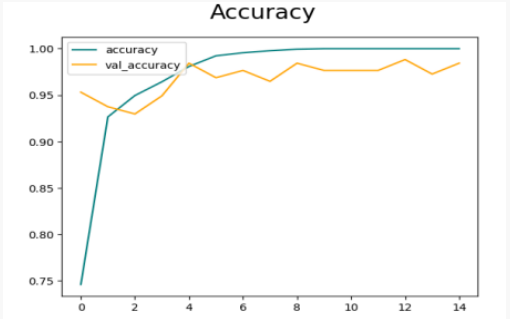

Fig 7(a) Accuracy curve

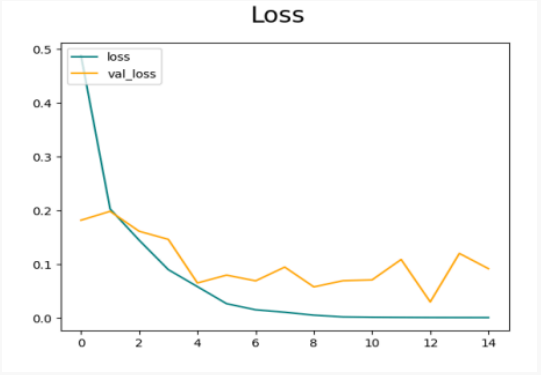

Fig 7(b) Loss curve

Figure 8, shows the confusion matrix of the model. The model predicts all the infected (bleeding) images correctly.

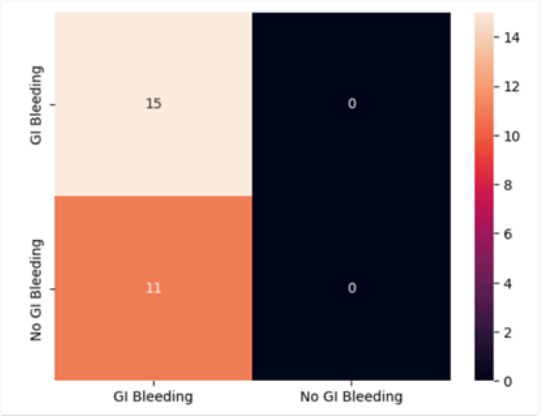

Fig 8 Confusion matrix

VI. Acknowledgement

We would like to express our sincere gratitude to Mr. Govind Gupta sir who was our mentor in this internship journey for his incredible support and guidance. We would also like to extend our gratitude to Prof. Anu Singh Lather, Vice Chancellor and Dr. Ritu Rani, our coordinator for providing us with this great opportunity.

VII. Conclusion

This research envisages a high-performance, CNN-based framework for automatic gastrointestinal bleeding detection in WCE images, which is one of the major steps toward raising diagnostic efficiency in gastroenterology. The model achieves near-perfect performance metrics through effective use of advanced preprocessing techniques and CNN layer optimizations of the WCEBleedGen dataset. This integration incorporates activation functions with pooling layers and the Adam optimizer, thus enhancing feature extraction and improving the accuracy of decision-making. These results further signify the importance of using tailored deep learning architectures in medical image analysis. Other future directions may involve model validation across diverse datasets, enhancing the adaptability of the model to different imaging protocols, and exploring multi-modal learning incorporating clinical metadata. This is indicative of the new role AI can play in transforming medical diagnostics into quicker, more reliable, and globally accessible healthcare solutions.

VIII. References

[1] Improving CNN Performance with Min-Max Objective

[2] 1-A Comprehensive Review on CNN Architectures and Techniques" by Zhiwei Luo, Li Sun, and Yifan Zhang (2020)

[3] LeCun et al. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE.

[4] Zhang, X., & Zheng, Y. (2016). Convolutional neural networks for image processing: An overview. IEEE Access.

[5] Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. Proceedings of the 3rd International 4- Conference on Learning Representations (ICLR).

[6] Kotsiantis, S. B. et al. (2006): Handling imbalanced datasets: A review. Discusses various methods, including class weighting, for dealing with imbalanced datasets.

[7] ConVision Benchmark: A Contemporary Framework to Benchmark CNN and ViT Models

[8] FURQAN RUSTAM 1 , MUHAMMAD ABUBAKAR SIDDIQUE1 , HAFEEZ UR REHMAN SIDDIQUI1 , SALEEM ULLAH 1 , ARIF MEHMOOD 2 , IMRAN ASHRAF 3 , AND GYU SANG CHOI 3 Wireless Capsule Endoscopy Bleeding Images Classification Using CNN Based Model