Gastrointestinal tract disease recognition based on denoising capsule network

Yaw Afriyie¹,²*, Benjamin A. Weyori¹ and Alex A. Opoku³

Abstract: Today, cancer is one of the leading causes of death in humans in the world. Cancers affect different parts of the human anatomy in different ways. There are significantly more deaths associated with gastrointestinal cancers such as colorectal cancer than with all other cancers. With computer-aided diagnosis systems, medical practitioners can recognize a variety of illnesses more quickly than with manual procedures. It is possible to reduce mortality significantly by detecting and removing precancerous lesions early. Due to the time-consuming and challenging nature of manual diagnosis, researchers have developed computational algorithms that are used to identify and classify diseases such as gastrointestinal (GI) diseases. Computer-aided diagnosis relies heavily on the classification of medical images, which is a challenging task. Therefore, this study presents a less sophisticated but still effective pre-processing technique for identifying endoscopic images known as denoising capsule networks (Dn-CapsNets). Moreover, we constructed activation maps (AM) using the feature representations to visualize the results. As a result of these evaluations, the trained model achieved 94.16%, 83.1%, 86.7%, 96.1%, 86.6%, and +0.69 respectively in terms of accuracy, precision, sensitivity, specificity, F1-score, and Matthew’s correlation. Comparisons between the proposed method and the current state-of-the-art have shown improved accuracy.

Keywords: capsule network; colorectal cancer; CNNs; gastrointestinal diseases

1. Introduction
The automatic identification and classification of diseases such as skin cancer (Saba 2019)–(Khan et al., 2020), lung cancer (Muhmmad Irfan Sharif et al., 2019), brain tumor (M. I. Muhammed Irfan Sharif et al., 2019), stomach cancer (M Attique Khan, Rubab et al., 2020), and a few others have been major study subjects in medical imaging for decades (Muhmmad Irfan Sharif et al., 2019) — (M Attique Khan, Akram et al., 2020). Cancers of the colon and tumors of the gastrointestinal tract are two of the most common gastrointestinal health problems in the United States (Siegel et al., 2017), killing between 60 and 70 million people each year, affecting between 60 and 70 million people each year. Stomach cancer claimed the lives of over one million individuals worldwide in 2017, making it the third-leading cause of death (Hyung et al., 2019). Most Colorectal cancers (CRCs) are caused by previous adenomas, and the adenoma-carcinoma sequence can be used to test for CRC and prevent it (Brethauer, 2011). Colonoscopy combined with adenoma resection can reduce the risk of CRC by up to 80% and the related death by 50% (Vokes et al., 1993). The development of computer-aided diagnostic (CAD) technologies has proven immensely useful for doctors in their clinics (Akram et al., 2018), (Nasir et al., 2018), (M. Muhammad Sharif et al., 2021). CNNs, for instance, rely on pooling layers to handle the most significant information, which makes
the network translationally invariant. Consequently, CNNs need to be augmented with a variety of data augmentation methods in order to generalize to other perspectives. The time and difficulty involved in these augmenting methods are, on the other hand, considerable. With dynamic routing, capsule networks were developed to overcome CNNs’ drawbacks (Sara et al., 2017).

A capsule (Sara et al., 2017) is made up of a number of neurons whose activity vector indicates the parameters associated with the instantiation and whose length indicates the likelihood that a given feature is encoded by the capsule. In layer $l$, each lower-level capsule $i$ (where) $i$ uses its own activity vector in encoding spatial information (e.g., scale, position, orientation, and skewness) of the input features as instantiation parameters. Despite their ability to predict attributes from each object in an image, capsule networks struggle with complex images. This has led to the use of CIFAR 10 to train most capsule network models proposed in the literature for identifying images. To address this issue, researchers have started developing computational approaches. In these approaches, segmentation, classification and feature visualization are all crucial processes. In their work, Yuan & Meng (Yuan & Meng, 2014) combined a Bag of Features technique with a saliency map to create an integrated polyp detection system. To characterize local features in the first phase of the BoF technique, SIFT feature vectors are clustered using k-means. Saliency characteristics were then determined using the saliency map histogram. Lastly, employing both BoF and saliency variables, the SVM was employed to undertake classification. The authors in (Yuan et al., 2016) improved this technique by adding LBP and uniform LBP (ULBP), complete LBP (CLBP), and histograms of oriented gradients (HoG) features, along with SIFT features. Finally, SVM and Fisher’s linear discriminant analysis (FLDA) classifiers were employed to categorize these features using different combinations of local characteristics. The SVM classifier produced good classification accuracy when SIFT and CLBP characteristics were combined. The medical imaging technique of wireless capsule endoscopy is used to examine the gastrointestinal (GI) tract. As shown in Figure 1, this technique (Campus et al., 2018) is widely used in hospitals to detect gastrointestinal abnormalities such as ulcers, bleeding, and many others.

CNN has lately been found to be quite useful in endoscopic procedures such as esophagogastro-duodenoscopy (EGD), colonoscopy, and capsule endoscopy. The anatomical position in EGD images (Takiyama et al., 2018), Helicobacter pylori (HP) infections (Shichijo et al., 2017), (Itoh et al., 2018) and gastric cancer were all challenges for a CNN-based diagnosis tool in EGD (Hirasawa et al., 2018). During a colonoscopy, a CNN-based diagnostic tool was employed to detect and characterize colorectal polyps (Komeda et al., 2017), (Byrne et al., 2019b), (R. Ruikai Zhang et al., 2017). In 2017, Komeda et al. (Komeda et al., 2017) reported a study that employed CNN to diagnose colorectal polyps utilizing 1,200 colonoscopy photographs and 10 additional video images of unlearned processes. A 10-fold cross-validation produces an accuracy of 0.751, where accuracy refers to the percent of answers that are accurate. In a study by Byrne et al. (Byrne et al., 2019a), the researchers developed a deep CNN model for real-time analysis of colorectal polyps within colonoscopic video images. Training and testing of the CNN model was done using only NBI videoframes (split evenly across relevant multi-classes) and unedited routine exam films that were not explicitly intended for CNN classification 106 sequentially encountered small polyps were used to validate the model using the second set of 125 movies. Using CNN to detect adenomas, the CNN

---

**Figure 1.** A sample frame of gastric abnormality and a healthy frame (Campus et al., 2018).
was accurate, sensitive, specific, negative predictive, and positive predictive, respectively, with 94\%, 98\%, 83\%, 97\%, and 90\%. Similar to Chang et al., Chang et al. (Y. Yuan Chang & Chen, 2019) achieved better categorization abilities across a wide range of categories using a deep attention neural network. A variety of techniques are employed to transform the data, including automatic data fusion, multi-epoch fusion, and adaptive threshold selection achieving an F1 score of 90.70\%. A polyp identification convolutional neural network based on a single shot multibox detector was described by (X. Xu Zhang et al., 2019). An upgraded SSD, according to the data, can increase mean average precision (mAP) from 88.5 to 90.4\%. The findings also show that when max-pooling layers are used, convolutional neural networks lose about 3/4 of their critical information. Ayidzoe et al. (Afriyie et al., 2021) suggested a capsule network variation that is less sophisticated but still robust and capable of extracting features. Their model took advantage of the Gabor filter and a proprietary preprocessing block to understand the structure and semantic information in an image, resulting in higher accuracy on the datasets studied. Several deep learning models based on CNNs have been proposed to diagnose gastrointestinal disorders. However, only a few studies have used gastrointestinal images to train capsule networks to our knowledge. We propose using CapsNets to construct a modified squash function for detecting gastrointestinal diseases as a result of our research. This paper also introduces improved methodologies that can be used as CapsNet performance indicators, enhancing the models’ dependability, explainability, and understandability.

The paper also underlines the need for CapsNet GI models that make judgments based on accuracies and activation map-related performance measures to be re-evaluated.

The identification of similar patterns of diseased regions is one of the most challenging components of gastrointestinal tract disease classification. Images of low-contrast ulcers, for instance, have similar hues to polyps, making classification difficult. Low contrast ulcer zones, on the other hand, are difficult to distinguish from healthy regions, making classification problematic. For successful classification, the selection of useable features is crucial, and this article focuses on the best features selection algorithm. By delivering more accurate classification results, robust features improve the overall efficiency of the CAD system.

We propose a modified framework to detect anomalies in GI images automatically by combining hybrid features and deep learning information. A denoising technique reduces noise in the images to help CapsNets learn better features from the endoscopic images and build a faster network. Summarily, the contributions of this paper are;

1. Based on the CapsNets, we present a novel and augmented framework for detecting malignant GI anomalies. To extract the required features, we create optimized CapsNets.
2. A proposed model is trained extensively tested on three datasets with GI anomalies. All datasets were successfully analyzed with our technique, and we demonstrated a generalized high performance with the newly built architecture.

2. Related work
Capsule network research is advancing rapidly in terms of its ability to recognize complex images. Capsules performed better on complicated images (S. Siwei Chang & Liu, 2020) when more layers were stacked, the number of capsule layers increased, and ensemble averaging was used. Many methods such as changing the reconstruction scaling factor, parallelizing capsule layer (Xiong et al., 2019), and others have been proposed. Detecting and recognizing gastrointestinal disorders is primarily done with supervised learning algorithms using CNNs and handmade features. A strategy for detecting esophageal cancer based on Gabor features and Faster Region-Based CNN is presented (Faster R-CNN). The handcrafted Gabor features were coupled with the CNN description in this manner (Ghatwary et al., 2019) Pogorelov et al. (Pogorelov et al., n.d.) published the “KVASIR” dataset, which used global features, deep CNNs, and deep transfer learning to
classify various diseases. For the classification of gastrointestinal illnesses, Nadeem et al. (Shees et al., 2018) used texture and deep learning features. Wimmer et al. (Wimmer et al., 2018) employed endoscopic image datasets to train three pre-trained CNN architectures, and SVM was then used to classify colonic polyps and celiac disease. They experimented with classification by concatenating and combining features from several levels. Other CNN-based techniques were outperformed by their method. Mossotto et al. (Mossotto et al., 2017) offered three unsupervised machine learning models that used only endoscopic data, only histology data, and mixed endoscopic data, with an accuracy of 71%, 76.9%, and 82.7%, respectively. Ozawa et al. (Ozawa et al., 2018) demonstrated the robustness of a GoogleLeNet CNN architecture based on a computer-aided diagnostic system (CAD) for predicting ulcerative colitis severity. According to Maeda et al. (Maeda et al., 2019) in a similar study persistent inflammation associated with ulcerative colitis can be predicted using CAD. Our study however presents a novel and integrated framework that is capable of detecting abnormalities from endoscopic images based on optimized CapsNets.

2.1. Proposed method
Dynamic routing (Sara et al., 2017) is used in the proposed method, which includes an optimized squash function and amplifier algorithm. In Figure 2, a three-layer convolutional capsule is shown along with four amplified layers that map onto the primary capsule to provide enhanced features. The images are resized to 28 × 28 to identify and categorize the Kvasir-v2 dataset, as shown in our proposed architecture. The input images are fed into the modified layers using 3 × 3 kernels to create 28 × 28 feature maps. As a result, the modified layer 1, which contains 96, 26 × 26 kernels and a Relu of 1, sends data to the Conv1. After the first modified layer produces 96 pieces, the second modified layer receives 26 dimensions. As input to the third modified layer, the second modified layer’s output is 96, 24 × 24 pixels, which are then fed into the Conv2, which has 128, 24 × 24 dimensions. Convolutional and modification layers are sandwiched between preprocessing layers and batch normalization layers. Concatenated layers were added to Conv4 to extract more semantic information and to improve the network’s learning ability.
2.2. Feature extraction and classification

Recently, extracting valuable learning features from a set of extracted features is an active area of research. The amount of irrelevant and redundant features in the originally extracted features may need to be ignored before final learning. It is critical to remove extraneous data to keep the suggested model consistent (Afriyie et al., 2021). Dynamic routing is used in this study because it has been widely adopted in the literature. This paper describes the encoder network using input, custom preprocessing (CP), convolutional layers, a primary capsule layer, and a class capsule layer. Activation maps are preprocessed (see Algorithm 1) to reduce noise, and features are extracted by convolutional and modified layers.

The trained images may contain noise which is expressed as \( \mu(x) = f(x) + n \) where \( n \) represents the white Gaussian additive noise.

\[
f(x) \approx \sum_y w(x,y)u(y)
\]

where

\[
w(x,y) = \frac{1}{W(x)} \exp \left( -\frac{1}{\sigma^2} \|u(N^d(x)) - u(N^d(y))\|^2 \right)
\]

where \( N^d(x) \) represents the square patch of size \((2d+1) \times (2d+1)\) centered at \( x \) and \( W \) is the normalizing term, \( W(x) = \sum_y w(x,y) \)

Eqns (1) and (2) demonstrate how the relevant features of each image are applied to the fastNLMMeansDenoisingColoredMulti function (Orchard et al., 2008) from open-cv2, which suppresses the noise. In the weighting scheme, the filter parameter (\( \alpha \)) controls how fast the exponential expression decays. The \( \alpha \)-parameter is typically controlled by algorithm 1. It is necessary to choose a very large \( \alpha \)-value to achieve a smoothing effect in the images, as a very small \( \alpha \)-value result in noise in the training images. Using the range \([30, 50]\) as an \( \alpha \)-value, we were able to achieve an overly-smooth image thereby discarding the irrelevant features.

Algorithm 1: DnCapsNets Algorithm

1: \( D^t_{m,n} = \{d^0_{m,n}, d^1_{m,n}, \ldots, d^{n-1}_{m,n}\} \leftarrow \text{Training Set} \)
2: \( P^t_{m,n} = \{p^0_{m,n}, p^1_{m,n}, \ldots, p^{n-1}_{m,n}\} \leftarrow \text{Test Set} \)
3: \( J^t_{m,n} \subset D^t_{m,n} \leftarrow \text{Features maps of Training set} \)
4: \( Q^t_{m,n} \subset P^t_{m,n} \leftarrow \text{Features maps of Test set} \)

Modifying feature maps generated from the training set

5: for \( i \) in \( J^t_{m,n} \) do:
6: \( K^t_{m,n} = J^t_{m,n} \ast \alpha \) where \( \alpha = [30, \ldots, 50] \)
7: fastNLMMeansDenoisingColoredMultiColored \( (K^t_{m,n}) \leftarrow \text{Eqn (2)} \)
8: return \( K^t_{m,n} \)
For all layer \( l + 1 \) capsules, the prediction vector \( \hat{u}_{ij} \) is generated, and it serves as the transformation of the entity represented by capsule \( j \) at a level \( l + 1 \) by capsule \( i \) at level \( l \). The vector \( \hat{u}_{ij} \) represents the proportion of the PC \( i \)'s contribution to the class capsule (CC) \( j \). In other words, the product \( \hat{u}_{ij} \) represents a vote for the class capsule \( j \) at layer \( l + 1 \) from the capsule \( i \) at the layer \( l \). If there should be top-down feedback to raise the coupling coefficient \( c_{ij} \) for that route, a scalar product between the prediction vector \( \hat{u}_{ij} \) and the output \( v_i \) of each class capsule is used. A big scalar product (from \( c_{ij} = \hat{v}_i \hat{u}_{ij} \)) for a lower-level capsule and a specific higher-level capsule indicates a strong agreement, with the \( c_{ij} \) for that connection increasing and the \( c_{ij} \) for all other links decreasing.

The function \( v_i \) works with the \( s_j \) input to a capsule \( j \), which is the weighted sum overall \( \hat{u}_{ij} \) from lower-level capsules. Coupling coefficients \( c_{ij} \) are trained during the iterative dynamic routing process and computed by SoftMax (Gao & Pavel, n.d.), with initial logits \( b_{ij} \) being the log prior probabilities that capsules \( i \) and \( j \) should be coupled. In other words, \( c_{ij} \)

equals \[ \frac{\exp(b_{ij})}{\sum \exp(b_{ij})} \] and is defined as \[ c_{ij} = \frac{\exp(b_{ij})}{\sum \exp(b_{ij})} \]. The logits are started with zero and updated using the old \( b_{ij} \) values and the agreement \( c_{ij} \) as \( b_{ij} = b_{ij} + v_i \hat{u}_{ij} = b_{ij} + a_{ij} \) in a for-loop. \[ \sum_{j=1}^{N} c_{ij} = 1 \] shows that at any point during the training, all the \( c_{ij} \) between a given capsule \( i \) and all other capsules in layer \( l + 1 \) must add up to one.

3. Datasets description and experimental setup

The experiments in this paper were carried out on a single 64-bit Windows PC with an NVIDIA GeForce GTX 1050 Graphic Processing Unit (GPU) running CUDA 10.1 and 8GB of dedicated memory, as well as Keras (TensorFlow backend). Using a 0.001 learning rate and a learning rate decay of 0.9, each model is trained for 200 epochs. The number of routing iterations for each dataset is changed between 1 and 6, and the margin loss function is used to train the models. Equation 2 depicts the margin loss. For all of the models, the datasets were divided into 9:1 ratio for training and testing. The margin and reconstruction losses make up the loss function needed to train the model.

In this implementation, the default values for \( m^+ \), \( m^- \), and the loss function (Sara et al., 2017) were kept. During the training of the Capsule models, three routing iterations were used. \[ L_k = T_k \max(0, m^+ - ||v_k||)^2 + \lambda (1 - T_k) \max(0, ||v_k|| - m^-)^2 \] (3)

where \[ T_k = \begin{cases} 1 & \text{if class } k \text{ is active} \\ 0 & \text{otherwise} \end{cases} \], \( \lambda = 0.5 \), \( m^+ = 0.9 \), and \( m^- = 0.1 \).

Kvasir-v2 Dataset: We used the Kvasir dataset for our tests (Pogorelov et al., n.d.). The Kvasir dataset was produced to improve applications that include automatic detection, classification, and localization of endoscopic pathological abnormalities in gastrointestinal tract images. A total of 4,000 colored images have been categorized and confirmed by medical endoscopists in this new collection. It is divided into eight classes that depict various diseases as well as typical anatomical markers. Each class has 500 examples, ensuring that the dataset is well balanced. We train all the datasets on the proposed model in this study. This experiment used five classes in total. Accordingly, 0 signifies esophagitis, 1 signifies normal-cecum, 2 signifies normal-pylorus, 3 signifies polyps, and 4 signifies ulcerative colitis. A polyp is a pathological condition of the cecum, whereas ulcerative colitis is a pathological condition of the stomach. Esophagitis is a condition characterized by abnormalities of the esophagus. Due to their duplication of the aforementioned classes, the other three datasets were not included in the experiment. A 90:10 leave-out strategy was applied to the dataset, which was scaled and resized to 28x28x3.

Figure 3. Samples from the Kvasir dataset.

(a) Esophagitis  (b) normal-cecum  (c) normal-pylorus  (d) polyps  (e) ulcerative colitis
F-MNIST: Xiao et al. (Vollgraf, n.d.) introduced the Fashion MNIST, which consists of 70,000 grayscale images divided among 10 classes, each with 7,000 images. Tests were conducted using 10,000 images, while training was conducted using 60,000 images.

CIFAR—10: In addition, we used Krizhevsky and Hinton’s CIFAR-10 (Krizhevsky, 2009), which consisted of 60,000 colored images that were divided into ten classes of 32 × 32 RBG images of the real world. There are 50,000 training images and 10,000 testing images in the dataset, with 90% serving as training images and 10% as testing images.

CIFAR—100: A total of 60,000 32x32x3 images make up this dataset. There are 100 classes in the dataset, made up of 50,000 training images and 10,000 test images.

As part of the training process, patience, the early stopping hyperparameter, is set to 10, and only the best models are saved.

4. Results and discussion
The experimental results of the model are evaluated using the methods proposed in this section. Samples of the different classes of the Kvasir dataset is shown in Figure 3.

4.1. Model’s flexibility and robustness
Models should be tested using the methods described here to ensure their stability, understandability, and credibility. Equation (4) gives the average accuracy of a model:

\[
\text{average accuracy} = \frac{\sum \text{of correct predictions}}{\text{total number of samples in testset}}
\]  

(4)

We adjusted the hyperparameters and/or intermediate layers of all the models randomly to see how responsive they were to the changes. It is difficult to detect how effectively a classifier differentiates one class from another when there is a class imbalance (Provost et al., 1998), the most common assessment metric for classification algorithms (Mensah et al., 2020). According to Patrick et al. (Ayidzoe, Yongbin, Kwabena et al., 2021), there was no significant difference in the performance of CapsNets models when momentum, batch size, learning rate, dropout, and learning rate decay were changed. One of the most significant hyperparameters that affected CapsNet models’ performance was the number of routing iterations. Three iterations generated the best results. Our study examines the proposed model’s accuracy and its validation findings (see, Figure 2). As shown in Table 1, our proposed model achieved comparable better results on all of the datasets used for testing.

In Figure 4, the confusion matrix shows how many images were correctly identified and how many were incorrectly identified. The precision, sensitivity, specificity, and accuracy values for each class were calculated to get useful information from this data. It can be shown that the proposed model performed well across all classes when the true positive and accuracy per class values are

| Dataset                  | Proposed Accuracy |
|--------------------------|-------------------|
| CapsNet (kvasir-v2)      | 94.16%            |
| CIFAR-10                 | 87.30%            |
| FMNIST                   | 94.93%            |
| CIFAR-100                | 69.14%            |
Compared to the proposed Capsule models, these adjustments negatively affected CNN models’ performance.

4.2. Confusion matrix-based metrics

Based on multi-class confusion matrices, Figure 4 summarizes the performance of the models. The confusion matrix can also be used to deduce other performance metrics, such as Precision, Sensitivity, Specificity, and Accuracy-per class, to evaluate the model’s performance on specific classes. As part of the confusion matrix, true positives (TPs), true negatives (TNs), false positives (FPs), and false-negatives (FNs) are also generated, which are critical to decision making based on model performance, even if interpretations can be complex and situation-specific in some cases. It is not fatal to classify a healthy GI as unwell, which is the reason why a classifier should collect as many false positives (FP) as possible. In these conditions, however, many FN forecasts indicate a terrible outcome by the model, as it is hazardous to label an infected GI as healthy. We present other metrics from equations (5)—(9).

\[
\text{Precision} = \frac{TP}{TP + FP} \tag{5}
\]

\[
\text{Sensitivity/Recall} = \frac{TP}{TP + FN} \tag{6}
\]

\[
\text{Specificity} = \frac{TN}{TN + FN} \tag{7}
\]

\[
\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \tag{8}
\]

\[
\text{MCC} = \frac{TP \times TN}{\sqrt{(TP + TN)(TP + FP)(TN + FN)(TN + FP)}} \tag{9}
\]

Using the Matthew correlation coefficient (MCC; Matthews, 1975), which compares the true-positive rates with the false-positive rates, the fitness function is determined. MCC is within the range of \([-1, +1\)]. In general, a larger MCC value leads to better performance when it comes to
image classification. We achieved an average MCC value of +0.691 indicating a strong positive relationship between the true-positive rate and false-positive rate of the classes of GI disease. Also, the performance comparison shown in Table 2 depicts an overall accuracy, sensitivity, precision, F1-score, and specificity of 94.16%, 86.7%, 83.1%, 86.6%, and 96.1% respectively.

4.3. Model convergence
As can be seen in Figure 5, the proposed Dn-CapsNets model learns and converges faster. For example, the Dn-CapsNets start achieving higher accuracies models between epochs 75 and 200. As a result, the proposed Capsule network's final accuracy can be approximated during the first few epochs. Before the final precision of the training can be determined, one must wait for the complete time of the training. The optimized architecture's ability to encode the texture of the infectious region of the stomach is credited with this convergence. The collaboration of these layers results in rapid learning and convergence. These are particularly useful for prototyping and preliminary GI disease investigations.

When feature extraction is performed, examining whether sections of the network are missing is imperative, as this condition causes excessive oscillations and prolongs training time (Japkowicz & Shah, 2011).

Figure 6 shows that the Conv2 layer of the proposed model (see, Figure 6b) is an efficient extractor, possibly due to the fact that it is a higher-level layer that can sample features from lower-level layers (Conv1 and Conv2) to represent advanced image characteristics. Thus, the proposed PC layer possesses all of the important classification properties that the baseline model lacks (see, Figure 6a). As well as helping to make CapsNets more explainable and understandable, this method enables them to be used in real-world settings.

Due to its independence from the decision threshold and invariance to a priori probability distributions of the classes, area under the curve (AUC) is more preferred than accuracy. AUC is a superior metric to accuracy based on levels of consistency and discrimination (Ling et al., 2003). It is preferable to use a classifier with a large AUC over one with a lesser AUC. For both the Receiver Operating Characteristic Curve (ROC) and the Precision-Recall Curve, AUC may be calculated (PR). The ROC curve should not be utilized to analyze balanced datasets since it tends to be unduly optimistic in circumstances where the dataset class distribution has a big skew. For unbalanced datasets, the PR curve is appropriate (Singla & Domingos, 2005). When it is not possible to gather enough data for the model, the ROC curve should not be employed (Sokolova, n.d.). In the literature, this distinction is rarely addressed and adopted for CapsNet implementations. The proposed models' ROC and PR curves have huge regions beneath the curves, as shown in Figure 8. In ROC and PR spaces, a class's objective is to be in the top left and upper right corners, respectively. As shown in Figure 7, the proposed model classes achieve this purpose better.

4.4. Comparing of results
To demonstrate the performance of our proposed model, we compare it with the state-of-the-art in literature. In light of the limited number of works on GI diseases using deep learning, we compare our proposed methods to those used on similar tasks. The proposed method for the classification of the datasets with labeled data is quite effective concerning the medical image dataset. Also, data sets with a numerical imbalance between classes are expected to benefit from the proposed method.

We further compare the average training time on the Kvasir-v2 dataset for the proposed and the Sabour’s model as seen in Table 4. It can be seen from Table 4 that the proposed converges faster
| Class         | TP  | TN  | FP  | FN  | F1-score | Sens | Prec | Spec | MCC  | Acc./Class |
|---------------|-----|-----|-----|-----|----------|------|------|------|------|------------|
| Esophagitis   | 198 | 793 | 7   | 2   | 0.978    | 0.99 | 0.966| 0.991| +0.972| 0.991      |
| Normal-Cecum  | 181 | 595 | 36  | 19  | 0.868    | 0.905| 0.834| 0.943| +0.137| 0.934      |
| Normal pylorus| 190 | 803 | 4   | 10  | 0.964    | 0.950| 0.979| 0.995| +0.956| 0.986      |
| Polyps        | 138 | 807 | 35  | 62  | 0.740    | 0.69 | 0.798| 0.958| +0.686| 0.907      |
| Ulcerative    | 160 | 573 | 51  | 40  | 0.779    | 0.800| 0.758| 0.918| +0.706| 0.890      |
| Averages      |     |     |     |     | 0.866    | 0.867| 0.831| 0.961| +0.691| 0.9416     |
than the baseline model. We further show our proposed algorithm with states-of-the-art schemes as shown in Table 3.

As shown in Figure 9, the secondary capsule layer clusters for the Dn-CapsNets model are produced. As can be seen from Figure 9, the first column shows clusters that are derived from the proposed model. Despite some overlap in classes, the proposed models for these datasets form separable clusters. A few outliers appear in the proposed model, but they are not too far from their

![Figure 5. The proposed model’s training accuracy.](image-url)
Figure 6. Training and test losses for the proposed model.

Figure 7. Feature maps for the proposed model.
respective clusters. This indicates that the proposed model can discriminate well between GI images when it comes to classifying them.

5. Conclusion
In this study, we present a robust framework (Dn-CapsNets) for identifying gastrointestinal tract diseases in the Kvasir-v2 dataset. A deep learning technique can help reduce the likelihood of developing malignant diseases and even decrease the number of benign tumors that need to be

![Figure 8. Receiver operating characteristics and precision-recall curves of the proposed model](image)

(a) P-R curve of the proposed model  
(b) ROC of the proposed model

| Table 4. Convergence/training time |
|-----------------------------------|
| Model               | Average Time |
|----------------------|--------------|
| Baseline             | 283s 28 ms   |
| Proposed             | 249s 17 ms   |

![Figure 9. A proposed Dn-CapsNet model based on class capsules and digit capsules.](image)
removed. To improve classification accuracy, the images were optimized to remove noise, and the convolutional and modified layers were applied to extract relevant features. The Dn-CapsNets we proposed showed superior performance on the KVASIR dataset when applied to endoscopic image classification. Resulting in improved classification, we discovered the proposed model was able to capture more detailed differences between similar-looking images. As well, we believe that our framework can be applied to other diseases that can be detected through computer-aided diagnosis. In the future, we hope to implement an improved model on HyperKvasir which is the largest, unbalanced, and possibly most complex digestive tract visual data set.

Funding
This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Author details
Yaw Afriyie1,2
E-mail: yafriyie@ubids.edu.gh
ORCID ID: http://orcid.org/0000-0003-4821-5804
Benjamin A. Weyori1
ORCID ID: http://orcid.org/0000-0001-5422-4251
Alex A. Opoku1
ORCID ID: http://orcid.org/0000-0001-6776-9736

1 Department of Computer Science and Informatics, School of Sciences, University of Energy and Natural Resources, Sunyani, Ghana.
2 Department of Computer Science, Faculty of Information and Communication Technology, SD Dumbo University of Business and Integrated Dev. Studies, Wa, Ghana.
3 Department of Mathematics & Statistics, School of Sciences, University of Energy and Natural Resources, Sunyani, Ghana.

Availability of data and material
All the datasets used in this study are publicly available.

Code availability
The code used for this study is a modification of the code at https://www.github.com/Xifeng-Guo/CapsNet-Keras

Authors’ contributions
All Authors contributed equally to the preparation of this study.

Disclosure statement
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.

Data availability statement
All data generated or analysed during this study are included in this manuscript (and its supplementary information files).

Citation information
Cite this article as: Gastrointestinal tract disease detection with deep attention neural network. 2568–2572. https://doi.org/10.1109/ACCESS.2020.2990700

Gao, B., & Pavel, L. (n.d.). On the properties of the softmax function with application in game theory and reinforcement learning. 1–10. https://arxiv.org/abs/1704.00805

Chang, Y., & Chen, W. (2019). Gastrointestinal tract disease detection with deep attention neural network. 2568–2572. https://doi.org/10.1109/ACCESS.2020.2990700

Chang, S., & Liu, J. I. N. (2020). Multi-lane capsule network for classifying images with complex background. IEEE Access. https://doi.org/10.1109/ACCESS.2020.2909700

Chang, Y., & Chen, W. (2019). Gastrointestinal tract disease detection with deep attention neural network. 2568–2572. https://doi.org/10.1109/ACCESS.2020.2990700

Gao, B., & Pavel, L. (n.d.). On the properties of the softmax function with application in game theory and reinforcement learning. 1–10. https://arxiv.org/abs/1704.00805

Ghatawary, N., Ye, X., & Zolgharni, M. (2019). Esophageal abnormality detection using densestnet based faster R-CNN With Gabor Features. IEEE Access, 7, 84374–84385. https://doi.org/10.1109/ACCESS.2019.2925585

Hirasawa, T., Aoyama, K., Tanimoto, T., Ishihara, S., Shichijo, S., Ozawa, T., Ohnishi, T., Fujishiro, M., Matsuo, K., Fujisaki, J., & Tada, T. (2018). Application of artificial intelligence using a convolutional neural network for detecting gastric cancer in endoscopic
images. *Gastric Cancer*, 21(4), 653–660. https://doi.org/10.1007/s10120-018-0793-2

Itoh, T., Kawahira, H., Nakashima, H., Yata, N., Kawahira, H., & Medical Engineering, F. (2018). Deep learning analyzes helicobacter pylori infection by upper gastrointestinal endoscopy images. 4–9. https://doi.org/10.1055/s-0043-120830

JangHyung, L., Young Jae, K., Yoon Woo, K., Sungjin, P., Youn-i, C., Yoon Joe, K., Dong Kyun, P., Kwang Gi, K., & Jun-Won, C. (2019). Spotting MALIGNANCIES FROM GASTRIC ENDOSCOPIC IMAGES USING DEEP LEARNING. *Surgical Endoscopy*. https://doi.org/10.1007/s00464-019-06677-2

Japkowicz, N., & Shah, M. (2011). Evaluating learning algorithms: a classification perspective. Cambridge University Press. https://doi.org/10.5555/1964882

Kahsaygebreslassie, A., Yaeocgbirmayezaheng, Tsighe Hagos, M., Achimihentai, P., & Pooja, J. (2019). “Automated gastrointestinal disease recognition for endoscopy images.” *Proceedings - 2019 International Conference on Computing, Communication, and Intelligent Systems*, ICCCIS 2019. January 312–316. https://doi.org/10.1109/ICCCCS.2019.8734458

Khan, M. A., Akram, T., Sharif, M., Ahmad, S., & Bukhari, C. (2020). An integrated framework of skin lesion detection and recognition through saliency method and optimal deep neural network features selection. 15929–15948. https://doi.org/10.1007/s00521-019-04534-0

Khan, M. A., Lali, I. U., Rehman, A., Ishaq, M., Sharif, M., & Saba, T. (2019). “brain tumor detection and classification: a framework of marker-based watershed algorithm and Multilevel priority features selection,” no. October, 909–922. https://doi.org/10.1002/jemt.23238

Khan, M. A., Rubab, S., Kashif, A., Imran, M., Muhmmad, N., Hussain, J., Zhang, Y.-D., & Chandra, S. (2020). Lungs cancer classification from CT Images: An integrated design of contrast based classical features fusion and selection. *Pattern Recognition Letters*, 129, 77–85.

Khan, M. A., Sharif, M., Akram, T., Ahmad Chan Bukhari, S., & Sunder Nayak, R. (2020). Developed newton-raphson based deep features selection framework for skin lesion recognition. *Pattern Recognition Letters*, 129, 293–303. https://doi.org/10.1016/j.patrec.2019.11.034

Khan, M. A., Sharif, M., Akram, T., Yasmin, M., & Sunder Nayak, R. (2019). Stomach deformities recognition using rank-based deep features selection. *Journal of Medical Systems*, 43(12). https://doi.org/10.1007/s10916-019-1466-3

Komeda, Y., Handa, H., Watanabe, T., Nomura, T., Kitashishi, M., Sakurai, T., Okamoto, A. (2017). Computer-aided diagnosis based on convolutional neural network system for colorectal polyp classification: preliminary experience. *Oncology (Switzerland)*, 93(1), 30–34. https://doi.org/10.1159/000481227

Krzhizhevsky, A. (2009). “Learning multiple layers of features from tiny images.” *http://www.cs.toronto.edu/~kriz/learning-features-2009-TR.pdf*

Ling, C. X., Huang, J., & Zhang, H. (2003). AUC: A better measure than accuracy in comparing learning algorithms. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2671, 329–341. https://doi.org/10.1007/3-540-44886-1_25

Maeda, Y., Kudo, S.-E., Mori, Y., Misawa, M., Ogata, N., Sasanuma, S., Wakamura, K., Oda, M., Mori, K., & Ohtsuka, K. (2019). Fully automated diagnostic system with artificial intelligence using endoscopy to identify the presence of histologic inflammation associated with ulcerative colitis (with Video). https://doi.org/10.1001/jama.2019.1572657

Matthews, B. W. (1975). Comparison of the Predicted and Observed Secondary Structure of T4 Phage Lysozyme. *BBA - Protein Structure*, 405(2), 442–451. https://doi.org/10.1016/0005-2795(75)90109-9

Mensah, P., Asubarn, B., & Abra, A. (2020). Exploring the Performance of LBP-capsule networks with k-means routing on complex images. In *Journal of King Saud University - Computer and Information Sciences* (pp. 2574–2588). https://doi.org/10.1016/j.jsuci.2020.10.006

Mossotta, E., Ashto, J. J., Coelho, T., Beattie, R. M., Mootha, B. D., & Ennis, S. (2017). *Classification of paediatric inflammatory bowel disease using machine learning*. Scientific reports. https://doi.org/10.1038/s41598-017-02606-2

Nasir, M., Attique, M., Ullah, I., Tonzilo, L., Tassawar, S., & Sharif, M. (2018). An improved strategy for skin lesion detection and classification using uniform segmentation and feature selection based approach. no. December 2017. Microscopy research and technique. https://doi.org/10.1002/jemt.23009

Orchard, J., Ebrahimi, M., Wong, A., & Canada, N. (2008). Efficient nonlocal-means denoising using the SVD University of waterloo Cheitn school of computer science, 2 Dept. of Applied Mathematics, 3 Systems Design Engineering Dept . Design Engineering, 1732–1735. https://ieeexplore.ieee.org/abstract/document/4712109

Ozawa, T., Ishihara, S., & Fujishiro, M. (2018) Novel Computer-Assisted Diagnosis System for Endoscopic Disease Activity in Patients with Ulcerative Colitis. *Gastrointestinal Endoscopy*. https://doi.org/10.1016/j.gie.2018.10.020

Pogorelov, K., De Lange, T., Johansen, D., Thelin Schmidt, P., Rieger, M., Lux, M., & Halvorsen, P. (n.d.). *Kvasir: A multi-class image dataset for computer aided gastrointestinal disease detection*. https://doi.org/10.1109/3083187.3083212

Poudel, S., & Woong Lee, S. (2021). Deep Multi-Scale Attentional features for medical image segmentation. *Applied Soft Computing*, 109, 107445. https://doi.org/10.1016/j.asoc.2021.107445

Provost, F., Fawcett, T., & Kohavi, R. 1998. “The case against accuracy estimation for comparing induction algorithms.” *International Conference on Machine Learning*, 445.

Sara, S., Frost, N., & Hinton, G. E. (2017, December). Dynamic Routing between Capsules. *Advances in Neural Information Processing Systems*, Nips, 3857–3867. https://proceedings.neurips.cc/paper/2017/file/2cad8f0a7b7bfe7822fadb8be65374b894-Paper.pdf

Sharif, M., Attique Khan, M., Rashid, M., Yasmin, M., Afza, F., John Tanik, U., Sharif, M., Attique Khan, M., & Rashid, M. (2021). Deep CNN and Geometric Features-Based Gastrointestinal Tract Diseases Detection and Classification from Wireless Capsule Endoscopy Images. *Journal of Experimental & Theoretical Artificial Intelligence*, 33(4), 577–600. https://doi.org/10.1080/0952813X.2019.1572657

Sharif, M. I., Jian Ping, L., Attique Khan, M., & Asim Saleem, M. (2019) Active Deep Neural Network Features Selection for Segmentation and Recognition of Brain Tumors Using MRI Images. *Pattern Recognition Letters*. https://doi.org/10.1016/j.patrec.2019.11.019
Shees, N., Atif Tahir B., M., Sadiq, S., Naqvi, A., & Zaid, M. (2018). Ensemble of texture and deep learning features for finding abnormalities in the gastro-intestinal tract. Springer International Publishing. https://doi.org/10.1007/978-3-319-98446-9

Shichijo, S., Nomuro, S., Aoyama, K., Nishikawa, Y., Miura, M., Shinagawa, T., Takiyama, H., Tonimoto, T., Ishihara, S., Matsuo, K., & Toda, T. (2017). Application of convolutional neural networks in the diagnosis of Helicobacter Pylori infection based on endoscopic images. EBioMedicine, 25, 106–111. https://doi.org/10.1016/j.ebiom.2017.10.014

Siegel, R. L., Fedewa, S. A., Anderson, W. F., Miller, K. D., Jemal, A., & Noone, A. M. (2017). Colorectal Cancer Incidence Patterns in the United States, 1974–2013. Journal of the National Cancer Institute, 109(8), 72–82. https://doi.org/10.1093/jnci/djw322

Singla, P., & Domingos, P. 2005. “Discriminative training of Markov logic networks.” Proceedings of the National Conference on Artificial Intelligence 2: 868–873.

Sokolova, M. (n.d.). Beyond Accuracy, F-Score and ROC: A family of discriminant measures for performance evaluation. https://doi.org/10.1007/11941439_114

Takiyama, H., Ozawa, T., Ishihara, S., Fujishiro, M., de Alcantara, B. N., Araujo, F. M. C., Queiroz, M. G. L., Cruz, A. C. R., Vasconcelos, B. H. B., Chiang, J. O., Martins, L. C., Casseb, L. M. N., da Silva, E. V., Carvalho, V. L., Vasconcelos, B. C. B., Rodrigues, S. G., Oliveira, C. S., Quaresma, J. A. S., & Vasconcelos, P. F. C. (2018). Automatic Anatomical Classification of Esophagogastroduodenoscopy Images Using Deep Convolutional Neural Networks. Scientific Reports, 8(April), 1–8. https://doi.org/10.1038/s41598-018-25842-6

Taruna, A., Rahul, G., & Shrikanth, N. 2019. “ON EVALUATING CNN REPRESENTATIONS FOR LOW RESOURCE MEDICAL IMAGE CLASSIFICATION” ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 1363–1367. Vokes, E. E., Weischelbaum, R. R., Lippman, S. M., & Ki Hong, W. (1993). The New England journal of medicine. no other uses without permission. copyright © 1993 Massachusetts medical society. all rights reserved. The New England Journal of Medicine, 29 (2), 1230–1235. https://doi.org/10.1056/NEJM1995205222462101

Vollgraf, R. (n.d.). Fashion-MNIST: A novel image dataset for benchmarking machine learning algorithms. 1–6. https://arxiv.org/abs/1708.07747

Wimmer, G., Véceli, A., Höfler, M., & Uhl, A. (2018). Fisher Encoding of Convolutional Neural Network Features For Endoscopic Image Classification. Journal of Medical Imaging, 5(1), 1. https://doi.org/10.1117/1. jmi.5.3.034504

Xiong, Y., Guijing, S., Shiwe, Y., Sun, Y., & Sun, Y. (2019). Deeper Capsule Network For Complex Data. July, 1–8. https://ieeexplore.ieee.org/iel7/8840768/8851681/08850200.pdf

Yuan, Y., Baopu, L., & Meng, M. Q. (2016). Improved bag of feature for automatic polyp detection in wireless capsule endoscopy images. 13(2), 529–535. https://ieeexplore.ieee.org/iel7/8856/7447676/07052426.pdf

Yuan, Y., & Meng, M. Q. (2016). Polyp classification based on bag of features and saliency in wireless capsule endoscopy’. 3930–3935. https://ieeexplore.ieee.org/abstract/document/6907429

Zhang, X., Chen, F., Yu Id, T., Jiye, A., Huang, Z., Jiquan Liu, I., Weiling, H., Wang, L., Duan, H., & Jianmin, S. (2019). Real-time gastric polyp detection using convolutional neural networks. 1–16. https://doi.org/10. 1371/journal.pone.0214133

Zhang, R., Zheng, Y., Wing Chung Mak, T., Ruoxi, Y., Wong, S. H., Lou, J. Y. W., & Poon, C. C. Y. (2017). Automatic detection and classification of colorectal polyps by transferring low-level CNN features from nonmedical domain. IEEE Journal of Biomedical and Health Informatics, 21(1), 41–47. https://ieeexplore.ieee.org/abstract/document/7769237
