

## Research Article

# Deep Learning Algorithms for Detection and Classification of Gastrointestinal Diseases

**Mosleh Hmoud Al-Adhaileh** <sup>1</sup>, **Ebrahim Mohammed Senan** <sup>2</sup>,  
**Fawaz Waselallah Alsaade** <sup>3</sup>, **Theyazn H. H Aldhyani** <sup>4</sup>, **Nizar Alsharif**,<sup>5</sup>  
**Ahmed Abdullah Alqarni** <sup>6</sup>, **M. Irfan Uddin** <sup>7</sup>, **Mohammed Y. Alzahrani** <sup>6</sup>,  
**Elham D. Alzain** <sup>4</sup>, and **Mukti E. Jadhav**<sup>8</sup>

<sup>1</sup>Deanship of E-Learning and Distance Education, King Faisal University, Al Hofuf, Saudi Arabia

<sup>2</sup>Department of Computer Science & Information Technology, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, India

<sup>3</sup>College of Computer Sciences and Information Technology, King Faisal University, Al Hofuf, Saudi Arabia

<sup>4</sup>Community College of Abqaiq, King Faisal University, P.O. Box 400, Al-Ahsa 31982, Saudi Arabia

<sup>5</sup>Department of Computer Engineering and Science, Albaha University, Al Bahah, Saudi Arabia

<sup>6</sup>Department of Computer Sciences and Information Technology, Albaha University, Al Bahah, Saudi Arabia

<sup>7</sup>Institute of Computing, Kohat University of Science and Technology, Kohat, Pakistan

<sup>8</sup>Shri Shivaji Science & Arts College, Chikhli Dist, Buldana, India

Correspondence should be addressed to Theyazn H. H Aldhyani; [taldhyani@kfu.edu.sa](mailto:taldhyani@kfu.edu.sa)

Received 30 April 2021; Accepted 22 September 2021; Published 18 October 2021

Academic Editor: Muhammad Javaid

Copyright © 2021 Mosleh Hmoud Al-Adhaileh et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Currently, nearly two million patients die of gastrointestinal diseases worldwide. Video endoscopy is one of the latest technologies in the medical imaging field for the diagnosis of gastrointestinal diseases, such as stomach ulcers, bleeding, and polyps. Medical video endoscopy generates many images, so doctors need considerable time to follow up all the images. This creates a challenge for manual diagnosis and has encouraged investigations into computer-aided techniques to diagnose all the generated images in a short period and with high accuracy. The novelty of the proposed methodology lies in developing a system for diagnosis of gastrointestinal diseases. This paper introduces three networks, GoogleNet, ResNet-50, and AlexNet, which are based on deep learning and evaluates them for their potential in diagnosing a dataset of lower gastrointestinal diseases. All images are enhanced, and the noise is removed before they are inputted into the deep learning networks. The Kvasir dataset contains 5,000 images divided equally into five types of lower gastrointestinal diseases (dyed-lifted polyps, normal cecum, normal pylorus, polyps, and ulcerative colitis). In the classification stage, pretrained convolutional neural network (CNN) models are tuned by transferring learning to perform new tasks. The softmax activation function receives the deep feature vector and classifies the input images into five classes. All CNN models achieved superior results. AlexNet achieved an accuracy of 97%, sensitivity of 96.8%, specificity of 99.20%, and AUC of 99.98%.

## 1. Introduction

Cancer is the most common cause of death in the world, and gastrointestinal cancer is the most frequently occurring type. The World Health Organization (WHO) estimates that 1.8 million people die annually from gastrointestinal diseases [1], and gastrointestinal cancer is the fourth cause of death in the

world. Gastrointestinal cancer grows from gastrointestinal polyps, which are abnormal tissue growths on the mucosa of the stomach and colon. The polyps grow slowly, and symptoms only appear when they are large [2]. However, polyps can be prevented and cured if detected at an early stage [3].

Video endoscopy plays an important role in increasing the early diagnosis of polyps in the gastrointestinal tract and

reducing the number of mortalities [4]. Endoscopy can determine the severity of ulcerative colitis by detecting mucosal patterns that include spatial differences in mucosal colour and texture (degree of roughness on the mucosal surface in gastrointestinal) [5]. Hundreds of images can be extracted from a gastrointestinal video, but disease appears in only a few images, and no medical person can devote the amount of time needed to monitor all the images. Consequently, the accuracy of the diagnosis depends mainly on the experience of the doctor; experts are able to diagnose polyps in up to 27% of cases [6]. Therefore, during the examination, polyps may remain undetected and lead to future malignancies.

The deficiencies of the radiologist and other human factors can lead to a false diagnosis, so a computer-aided automated method would be valuable for diagnosing polyps with high accuracy and at the early stages of cancer. Artificial intelligence techniques have shown massive potential in various medical fields for helping humans to visualize disease that cannot be discovered with the naked eye [7–9]. For example, artificial intelligence techniques can extract complex microimaged structures from endoscopy images and identify key features. Clinically, artificial intelligence techniques can distinguish between neoplastic and nonneoplastic tissues. Techniques are also available for extracting texture features to evaluate the risk of gastric cancers [10, 11]. Colonoscopy images have been used to classify colitis by extracting texture features [12, 13]. However, the challenges in extracting the features of gastrointestinal images limit the diagnostic accuracy [14].

Machine learning techniques have been used to extract colour, texture, and edge features from endoscopic images that depend on trial and error for disease diagnosis [15, 16]. Convolutional neural networks (CNNs) have begun to solve these feature engineering limitations, and the use of CNN in supervised learning has greatly improved medical image diagnostics [17]. CNN has proved its tremendous ability to extract features by transferring engineering to the learning process [18]. Deep learning algorithms have shown an outperformance of medical image diagnostics over the performance of experts [19]. Hence, computer-aided diagnostics using deep learning techniques for endoscopic images have the potential to achieve diagnostic accuracy that is better than that obtained by trained specialists [20].

Karkanis et al. [21] presented a technique for extracting colour features based on wavelet decomposition for diagnosing colon polyps. Many studies have applied machine learning techniques to diagnose gastrointestinal images, with features extracted through approaches that include polyp-based local binary, grey-level co-occurrence matrices (GLCMs), wavelets, context-based features, and edge shape and valley information [22–24]. The system proposed by Tajbakhsh et al. [25] has achieved better performance than other methods. Challenges remain for extracting hand-crafted features, such as light reflection, camera angle, and structural polyps. CNN techniques are powerful extractors of deep features, and CNN has achieved promising results in recent years in diagnosing medical images. Zhang et al. [26]

introduced a polyp detection system based on a Single Shot MultiBox Detector (SSD), where they reused missing features from the max pooling layers and added them to feature maps to increase the detection accuracy and classification. Godkhindi and Gowda [27] presented a CNN system for detecting polyps from CT colonography images. Their algorithm segments the colon from the CT image and isolates it from the rest of the organs. It then diagnoses a colon polyp by extracting the shape features. Ozawa et al. [28] presented a system for detecting colorectal polyps by SSD and reported promising results in diagnosing these polyps. Pozdeev et al. [29] presented a two-stage polyp segmentation and automated classification system. The first stage uses the global features of endoscopic images to classify the presence or absence of a tumor, while the second stage includes segmentation by CNN. Wan et al. [30] have introduced biomedical optical spectroscopy techniques to detect gastrointestinal cancers in early stages. This type of spectroscopy has the potential to provide structural and chemical information and has many advantages, including non-invasiveness, a reagent-free protocol, and a nondestructive procedure. Ribeiro et al. [31] have proposed the use of CNNs for diagnostics of the colon mucosa to uncover colon polyps for early-stage colon cancer classification.

CNN extracts features by exploiting the input pixels to handle distortions caused by different light conditions. Min et al. [32] developed a computer-aided system to diagnose linked colour imaging by extracting colour features from lesions. The system classified images as adenomatous polyps and nonadenomatous polyps, and the system achieved satisfactory results. Song et al. [33] also developed a computer-aided system for diagnosing colorectal polyp histology by CNN techniques; their network classified the polyps into three types: serrated polyp, deep submucosal cancer, and benign adenoma mucosal or superficial submucosal cancer.

The main contribution of the present paper is the provision of a computer-aided detection method for lower gastrointestinal diseases with modified criteria for extracting deep shape, colour, and texture features and adapting them to a learning transfer technique for fine-tuned and contoured transfers. Extensive experiments were conducted to select pretrained models to diagnose lower gastrointestinal diseases. New models were developed for transferring features extracted from a nonmedical deep learning dataset and adapting them to the new dataset. The remainder of the paper is organised as follows. Section 2 describes the background and motivations. Section 3 discusses the materials and methods used to diagnose a dataset. Section 4 presents the analysis and results of the study that have been achieved and compares the proposed systems with those from previous studies. Finally, the conclusions are presented in Section 5.

## 2. Background and Motivations

This section provides the fundamentals of gastrointestinal diseases and an overview of deep learning for diagnosing medical images.

*2.1. Overview and Status of Gastrointestinal Diseases.* Computer-aided early detection of disease is an important research field that can improve healthcare systems and medical practice around the world. The Kvasir dataset, which contains gastrointestinal images, is classified into three clinically important findings, three significant anatomical landmarks, and two categories of endoscopic polyp removal. The gastrointestinal tract is affected by many diseases, with 2.8 million new cases and 1.8 million deaths caused annually by oesophageal and stomach cancers. The gold standard for gastrointestinal examination is endoscopy. The upper gastrointestinal examination, involving the stomach, oesophagus, and upper part of the small intestine is done by gastroscopy, while the colon and rectum are examined by a colonoscopy. Both of these examinations are done as real-time videos with high resolution. Endoscopy equipment is expensive and requires extensive experience and training. Endoscopic detection and removal of lesions in their early stages, followed by appropriate treatment, are important for preventing colorectal cancer. Doctors vary in their abilities to detect colorectal cancer, and this may affect colorectal cancer diagnosis if a doctor's ability to evaluate the images is limited. An accurate diagnosis of the type of disease is also important for treatment and follow-up. Therefore, automatic diagnostics would be very welcome. Automatic diagnosis of pathological findings could contribute to the evaluation and identification of gastrointestinal cancers, thereby improving the efficiency and use of medical resources.

*2.2. Deep Learning.* CNNs are computational systems designed for the purpose of pattern recognition. CNN has entered into a number of fields, including healthcare [34], and has an important role in the diagnostics of images obtained in early disease stages. Image recognition and diagnostic accuracy are two tasks in which CNN excels compared to human experts. CNN has three types of layers: convolutional layers, pooling layers, and fully connected layers [35]. CNN has more ability than traditional networks and other networks such as RNN to deal with images because it uses a combination of technologies with these layers [36, 37]. The basic idea underlying CNN is the use of two-dimensional images and application of two-dimensional filters, in addition to a learning transfer technique where models are trained using the best pretrained models and the last three layers are replaced to learn the weights of the problem to be solved. CNN features are extracted from the dataset that they are trained on, so experts do not need to manually extract the features [38]. CNN's strength comes from its ability to learn the representative features in its training dataset. Convolutional layers work like the human brain does in feedback, where each layer acts as feedback to the next layer and the process continues until precise features are obtained.

*2.2.1. GoogleNet.* GoogleNet, developed by Google researchers, is a model of CNNs, sometimes called Inception V1; it consists of 22 layers (27 layers including the pooling

layers). The GoogleNet architecture was the winner in the classification challenge for images at the ILSVRC 2014. GoogleNet is used in many fields, including computer vision tasks, as well as in medical image classification. Figure 1 illustrates the architecture of the GoogleNet used to classify 5,000 images into five diseases from the lower digestive system. The GoogleNet architecture consists of 27 layers, including layers that do not contain parameters. They are divided into the input layer, in which the images are inputted into an RGB system with a size of  $224 \times 224$  pixels. The first convolutional layer contains two  $7 \times 7$  filters; these are among the largest filters compared to the other layers. This layer reduces the size of the input image, followed by the max pool layer with  $3 \times 3$  filters, a convolutional layer with a  $3 \times 3$  filter, and then the max pool layer with  $3 \times 3$  filters. The output is inputted into a two-layer block inception module, followed by a max pool layer with  $3 \times 3$  filters and then another four-layer block for an inception module. This is followed by a max pool layer with  $3 \times 3$  filters, a two-layer block inception module, and then by a max pool layer with  $3 \times 3$  filters. The average pooling layer has a size of  $7 \times 7$  pixels. Stride is used to determine the amount of filter shift on the input image. The Dropout technique uses this to prevent overfitting. In our work, Dropout was set at 40%, which means that the neurons are stopped by 40% in each iteration, and different parameters are used in each iteration. The fully connected layer received 9216 features and produced 4,096 features. The softmax layer produces five classes: dyed-lifted polyps, normal cecum, normal pylorus, polyps, and ulcerative colitis.

*2.2.2. ResNet-50.* ResNet-50 is a residual CNN model consisting of 177 layers. ResNet-50 was the winner in the image classification challenge in 2015. ResNet-50 is the backbone of many computer vision tasks. Figure 2 illustrates the architecture of the ResNet-50 used to classify 5,000 images divided into five diseases from the lower digestive system. The ResNet-50 architecture consists of 16 blocks that contain 177 layers divided into the input layer that inputs RGB images with a size of  $224 \times 224$  pixels and 49 convolutional layers, which use different types of filters [39]. The convolutional layer extracts deep features from the input images and stores them in deep feature vector maps, with one pooling layer for both average and max. These two layers reduce the feature vector map dimensions. Batch normalisation then helps the network to choose the learning rate correctly. The Rectified Linear Activation function (ReLU) that follows the convolutional layers only passes positive outputs and converts the negative values to zero. The fully connected layer receives 9216 features and produces 4096 features, and the second connected layer produces 1000 features. The softmax layer produces the five classes: dyed-lifted polyps, normal cecum, normal pylorus, polyps, and ulcerative colitis.

*2.2.3. AlexNet.* AlexNet is a model of CNNs and consists of 25 layers. AlexNet was the winner in the ImageNet classification competition in 2012, with a top-5 error rate of 15.3%

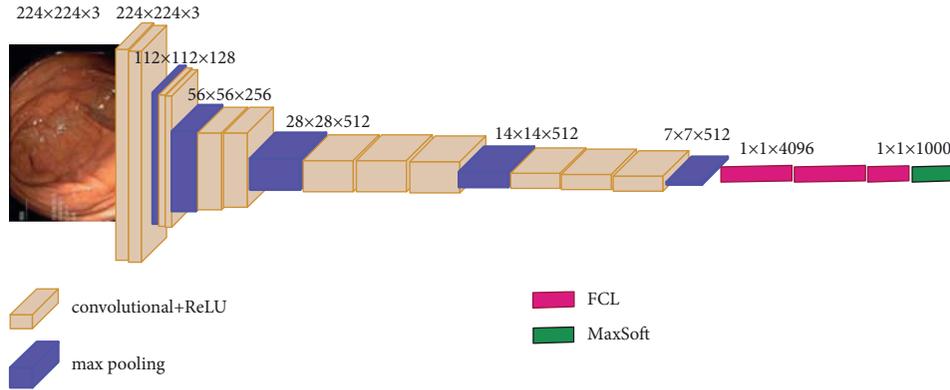


FIGURE 1: Structure of a GoogleNet for gastrointestinal disease diagnosis.

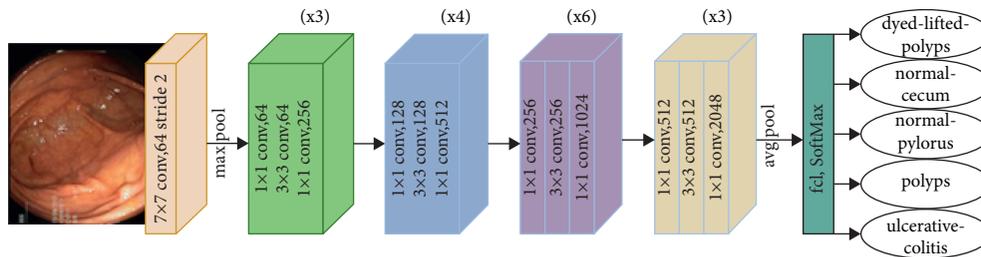


FIGURE 2: Structure of a ResNet-50 for gastrointestinal disease diagnosis.

[40]. Figure 3 illustrates the architecture of the AlexNet used to classify 5,000 images divided into five diseases from the lower digestive system. The architecture of AlexNet consists of 25 layers divided into the input layer that inputs RGB images at  $227 \times 227$  pixels size and five convolutional layers that use different types of filters. The convolutional layer extracts deep features from the input images. Three layers are used in max pooling; these layers reduce the dimensions of the feature vector maps. Two layers of cross-channel normalisation work on reparameterisation of the vector weights and choose the appropriate learning rate. Seven layers of the ReLU follow the convolutional layers. ReLU only outputs the positive values, while converting the negative values to zero. Three fully connected layers operate in series. The first connected layer receives 9216 features and produces 4096 features, the second connected layer produces 4096 features, and the third fully connected layer produces 1000 neurons (features). A softmax layer produces the five classes: dyed-lifted polyps, normal cecum, normal pylorus, polyps, and ulcerative colitis.

### 3. Materials and Methods

The computer-aided automatic detection of gastrointestinal diseases is an important research field. In this section, we describe the GoogleNet, ResNet-50, and AlexNet models of the CNN for early and accurate diagnosis of lower gastrointestinal disease. The general structure of the gastrointestinal detection system used in this work is shown in Figure 4. Preprocessing improves images and removes noise and artifacts, while the image augmentation technology improves training process. The convolutional layers extract the

deepest and most important features from each image. The fully connected layers diagnose and classify the gastrointestinal images.

**3.1. Dataset.** The dataset was collected from the Vestre Viken Health Trust (VV) in Norway from the gastroenterology department at the Baerum Hospital using endoscopic equipment. All images were described by experts from VV and the Cancer Registry of Norway (CRN). The CRN is the national body at the Oslo University Hospital that is in charge of screening and early detection of cancer to prevent spread. The Kvasir dataset consists of interpreted images by experts, including classes containing endoscopic procedures in the gastrointestinal tract and anatomical landmarks. The dataset contains hundreds of images that are sufficient for use in deep learning and transfer learning. The dataset is in RGB colour space and consists of images in resolution from  $720 \times 576$  up to  $1920 \times 1072$  pixels. In our work, the dataset contains 5,000 images equally divided into five diseases: dyed-lifted polyps, normal cecum, normal pylorus, polyps, and ulcerative colitis. Figure 5 shows samples from the Kvasir dataset. The data are available in this link: <https://datasets.simula.no/kvasir/#download>.

**3.2. Preprocessing and Augmentation Techniques.** Noise and artefacts arise from light reflections, photographic angles, and the mucous membranes surrounding the internal organs and reduce the performance of CNN due to the increased complexities of feature extraction. Therefore, optimisation processes have been of interest to researchers to improve image quality. In this paper, gastrointestinal

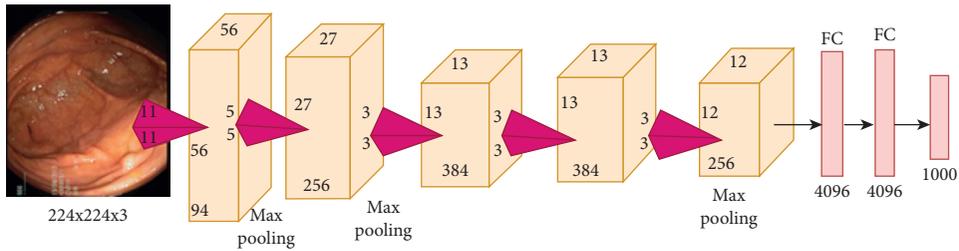


FIGURE 3: Structure of an AlexNet for gastrointestinal disease diagnosis.

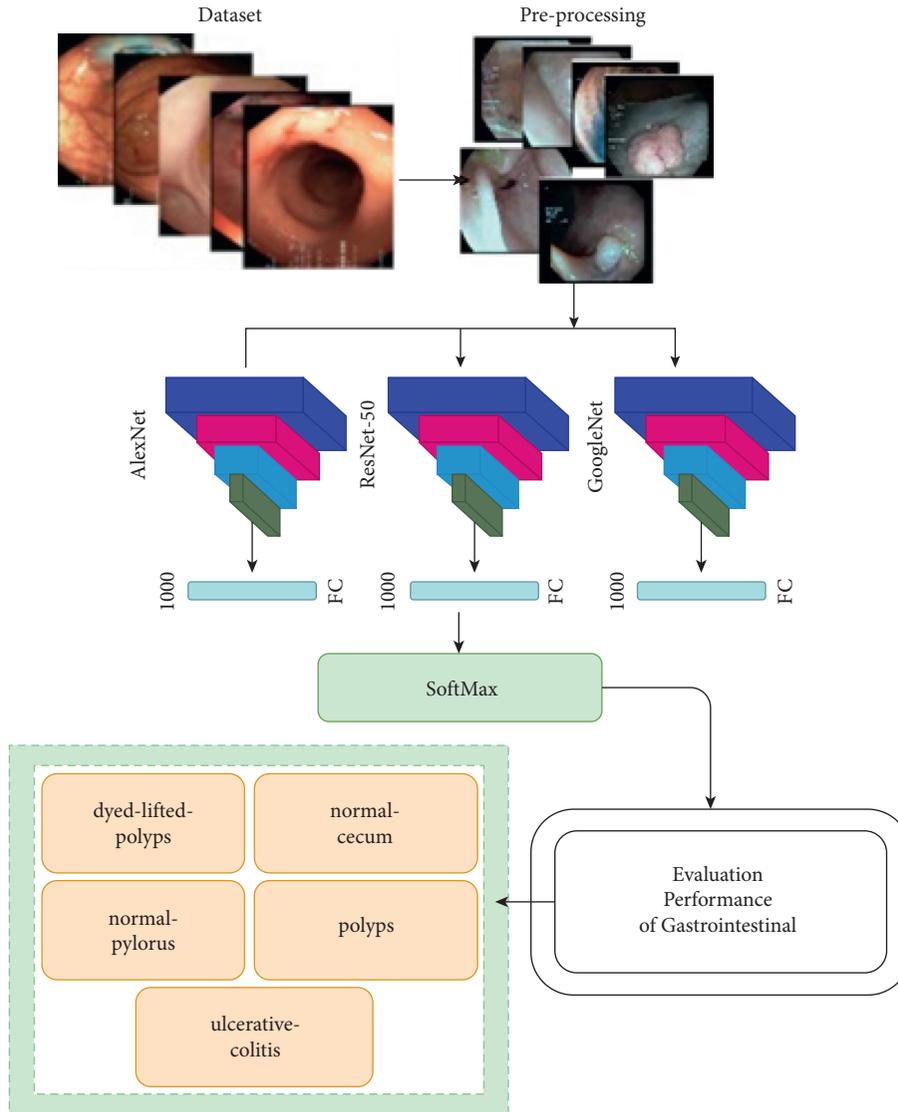


FIGURE 4: A general structure of gastrointestinal disease detection by deep learning.

images were preprocessed before being inputted into the CNN models. First, the image was scaled for colour constancy, and the image sizes were changed to  $244 \times 244$  pixels for both GoogleNet and ResNet-50 models and to  $227 \times 227$  pixels for the AlexNet model. The mean for the three RGB channels was then calculated for the gastrointestinal images. Finally, the enhancement process was conducted through

the average filter, which calculates the average for each pixel with its neighbours and replaces it; this process continues for all pixels of the image [41, 42].

CNN techniques depend mainly on the volume of data. A larger set of training data generates more promising results for the model. Because of the lack of medical images, data augmentation techniques improve CNN models for

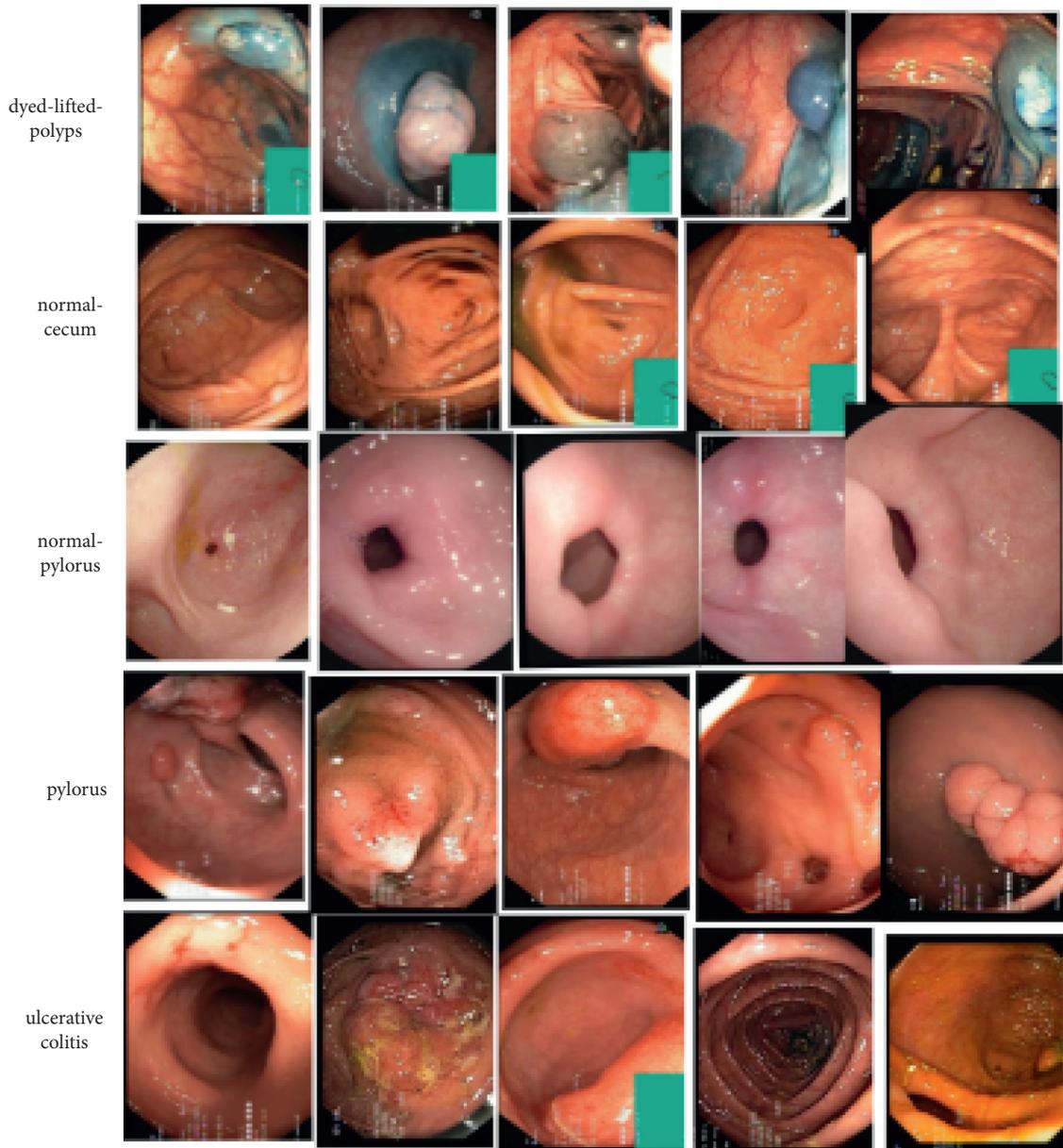


FIGURE 5: Samples from the Kvasir dataset.

accurate classification [37, 43]. The data augmentation technology also works to balance the dataset when the number of images differs between classes. In this paper, images of the training data were augmented through the operations of flipping, zooming, shifting, and  $\pm$ rotation [44].

**3.3. Convolutional Layers.** The gastrointestinal dataset contains many features, such as shape, texture, and colour. The manual extraction of features requires substantial experience, especially when extracting images from a video, where many images do not include the disease and the disease features appear in a few images that may be missed by the radiologist and the specialists. CNN algorithms work by extracting representative features of each disease through convolutional layers. GoogleNet contains many

convolutional layers and nine inception layers, ResNet-50 has 49 convolutional layers, and AlexNet has five convolutional layers. These layers apply a set of filters and adjust the weights during the training phase to address the deep features and pass them on to the next layer. The average and max pooling layers also reduce the size of the feature maps and represent a group of pixels either by means of the average or the max value between the groups of pixels. The convolutional layers extract representative features of each image for a total of 9216 features per image and represent them in feature maps to feed them to the classification layers.

**3.4. Normalisation of the Image.** Normalisation is one of the techniques used in training deep neural networks to normalise images for the purpose of accelerating training

processes. Normalisation aids in the appropriate choice of the learning rate through gradient descent converging. Without the normalisation process, the learning rate is more difficult and takes longer. In our work, the image normalisation process was done by subtracting the mean of the complete training set for each pixel. A variance of the dataset was calculated and divided by every pixel, resulting in data centring and making the variance of each feature equal to one.

**3.5. Dropout Technology.** CNNs produce millions of parameters, leading to overfitting. For this reason, CNN applies a Dropout technique to reduce overfitting. In our work, the Dropout technique was applied to GoogleNet, ResNet-50, and AlexNet at 40%, which means that 40% of the neurons are stopped in each iteration. Therefore, the networks use different parameters for each iteration. However, the Dropout technique doubles the training time.

**3.6. Transfer Learning.** Transfer learning is one of the most important steps in CNN [45]. In our study, transfer learning and fine-tuning were applied to GoogleNet, ResNet-50, and AlexNet networks pretrained to the ImageNet dataset [46]. Transfer learning is based on training a dataset to solve a specific problem and then transferring that learning to solving another related dataset problem [47, 48]. Transfer learning works by choosing the pretrained model and the size of the problem and using what has been learned to transfer the generalisation to another task. Transfer learning also avoids overfitting. In this work, transfer learning was applied to GoogleNet, ResNet-50, and AlexNet, where the weights were fine-tuned. The GoogleNet, ResNet-50, and AlexNet models were trained on the ImageNet dataset, and the learning was then transferred to the gastrointestinal dataset. The last three layers of the patterns were deleted and replaced with a fully connected layer. The first connected layer received 9,216 neurons and outputted 4,096 neurons, while the second connected layer received 4,096 neurons and outputted those. The softmax layer produced the five classes: dye-lifted polyps, normal cecum, normal pylorus, polyps, and ulcerative colitis.

**3.7. Optimizers (Adam).** Optimizers are used to change and tune parameters of neural networks such as weights, biases, and learning rate to reduce loss. Optimizer methods are considered to improve the deep learning classifier, which helps to speed up the performance of models. Adaptive Moment Estimation (Adam) is one of the best deep learning optimizers. Adam is a compilation of both RMSProp and momentum [49]. The adaptive learning rate for each parameter is calculated by Adam. It keeps average past gradient like momentum  $m_t$ , and it keeps squared gradients by storing past decaying average  $v_t$ . The following equation describes how Adam works to tune parameters, learning rate, etc.

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t, \quad (1)$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2, \quad (2)$$

where  $m_t$  refers to the first moment in the gradient,  $v_t$  refers to the second moment in the gradient, and  $\beta_1$  and  $\beta_2$  indicate the decay rate.

## 4. Experimental Results

Weights and parameters were adjusted for GoogleNet, ResNet-50, and AlexNet CNNs in the training phase to evaluate the dataset of gastrointestinal diseases. Table 1 shows the training options for the three networks and the execution time in the MATLAB environment. The resources are Core i5 Gen 6 with 4G NVIDIA GPU. In this paper, three experiments were conducted to evaluate a gastroenterology dataset containing 5,000 images divided equally among five diseases. The same dataset was applied in all three experiments. The dataset was divided into 80% for training and 20% for selection and validation.

Figure 6 shows the confusion matrix and AUC obtained from GoogleNet, ResNet-50, and AlexNet. The confusion matrix reviews all test images that are correctly classified (true negative (TN) and true positive (TP)) and incorrectly classified (false positive (FP) and false negative (FN)). The AUC also shows the ratio of TP vs. FP. Table 2 and Figure 6 show the evaluation of the dataset for three CNN models. Accuracy, sensitivity, specificity, and AUC are calculated according to equations (3)–(6). All networks showed promising results, as indicated in Table 2. The systems achieved an accuracy of 96.7%, 95%, and 97%, a sensitivity of 96.60%, 94.80%, and 96.80%, a specificity of 99%, 98.80%, and 99.20%, a validation accuracy of 96.70%, 97%, and 94.88%, and an AUC of 99.99%, 99.69%, and 99.98%, for GoogleNet, ResNet-50, and AlexNet, respectively. Note that the results in all experiments were approximately equal.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} * 100\%, \quad (3)$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} * 100\%, \quad (4)$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} * 100\%, \quad (5)$$

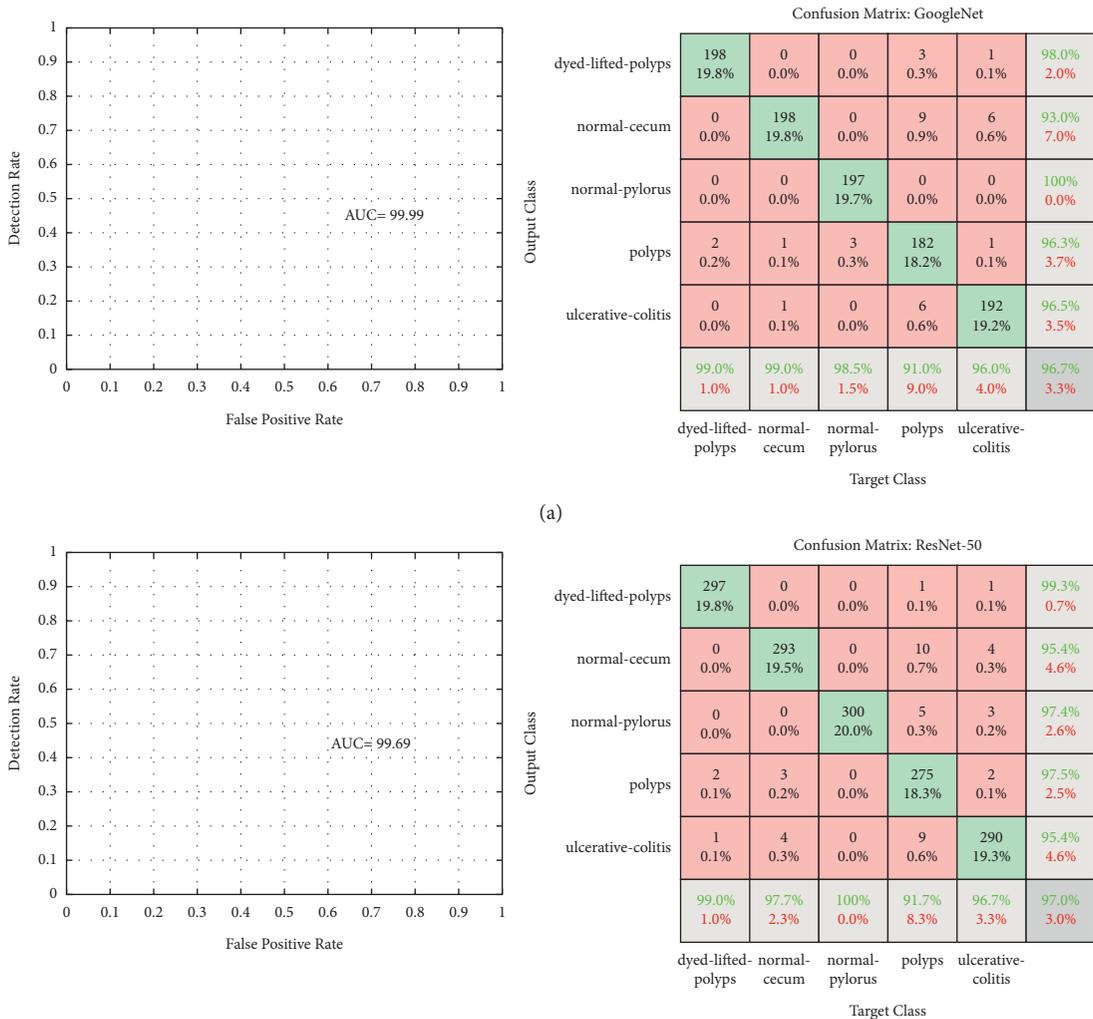
$$\text{AUC} = \frac{\text{true positive rate}}{\text{false positive rate}} = \frac{\text{Sensitivity}}{\text{Specificity}} * 100\%. \quad (6)$$

TP represents the number of positive samples correctly classified. TN represents the number of negative samples correctly classified. FP represents the number of benign samples classified as malignant. FN represents the number of malignant samples classified as benign.

Table 3 and Figure 7 show the classification performance of GoogleNet, ResNet-50, and AlexNet at the level of each disease. ResNet-50 and AlexNet achieved 99% accuracy in classifying dyed-lifted polyps, while ResNet-50 and AlexNet achieved 95% accuracy in classifying normal cecum disease. GoogleNet achieved the best classification (100%) for normal pylorus disease. Polyps were classified at 98% by ResNet-50,

TABLE 1: Options for configuring the training parameters for deep learning networks.

| Options               | GooleNet       | ResNet-50      | AlexNet        |
|-----------------------|----------------|----------------|----------------|
| Training options      | Adam           | Adam           | Adam           |
| Minibatch size        | 20             | 20             | 128            |
| Max epochs            | 10             | 4              | 10             |
| Initial learn rate    | 0.0003         | 0.0001         | 0.0001         |
| Validation frequency  | 3              | 5              | 50             |
| Training time         | 608 min 21 sec | 693 min 25 sec | 100 min 37 sec |
| Execution environment | GPU            | GPU            | GPU            |



(b) FIGURE 6: Continued.

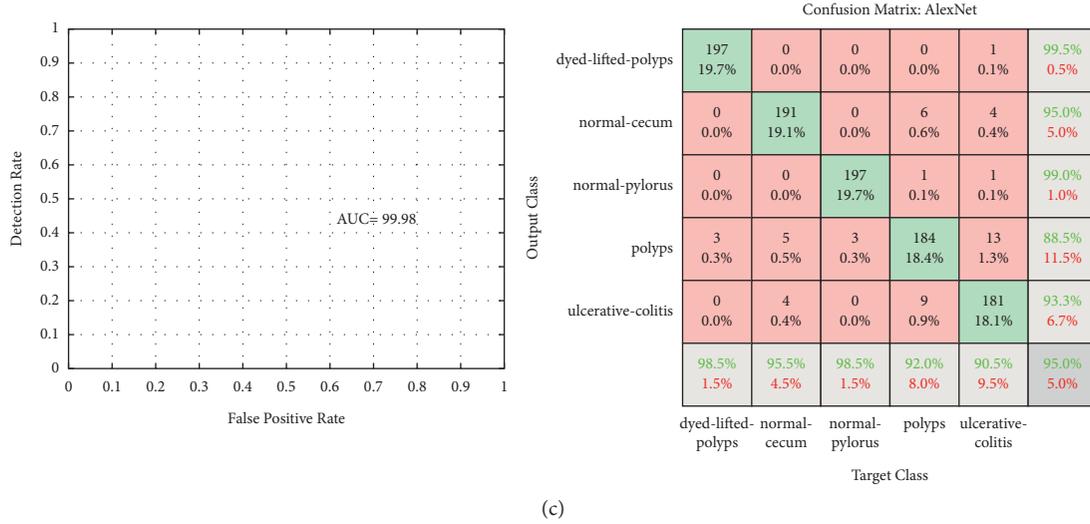


FIGURE 6: Confusion matrix and AUC for (a) GoogleNet, (b) ResNet-50, and (c) AlexNet.

TABLE 2: Results of diagnosing gastrointestinal diseases using deep learning models.

| Measurement                | GoogleNet | ResNet-50 | AlexNet |
|----------------------------|-----------|-----------|---------|
| Accuracy (%)               | 96.70     | 95.00     | 97.00   |
| Sensitivity (%)            | 96.60     | 94.80     | 96.80   |
| Specificity (%)            | 99.00     | 98.80     | 99.20   |
| Accuracy of validation (%) | 96.70     | 97.00     | 94.88   |
| AUC (%)                    | 99.99     | 99.69     | 99.98   |

TABLE 3: Performance evaluation results for the gastrointestinal disease datasets.

| Disease types      | GoogleNet | ResNet-50 | AlexNet |
|--------------------|-----------|-----------|---------|
| Dyed-lifted polyps | 98.00     | 99.00     | 99.00   |
| Normal cecum       | 93.00     | 95.00     | 95.00   |
| Normal pylorus     | 100.00    | 97.00     | 99.00   |
| Polyps             | 96.00     | 98.00     | 88.00   |
| Ulcerative colitis | 96.00     | 95.00     | 93.00   |

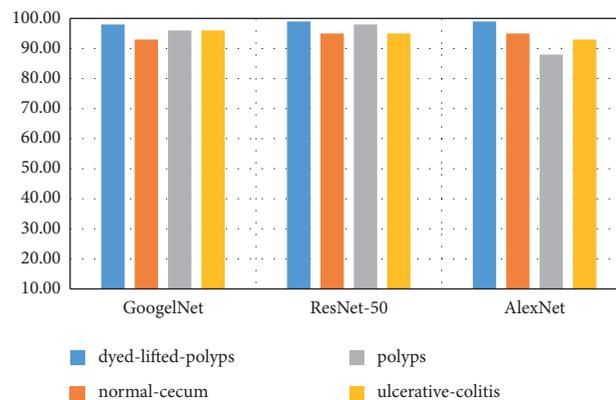


FIGURE 7: Performance of the three models for the detection of each gastrointestinal disease.

while GoogleNet achieved the best performance for classifying ulcerative colitis (96%).

The proposed CNNs were evaluated through several measures previously examined in the literature, as shown in Table 4. All relevant literature reported accuracy between

70.40% and 90.20%, while our proposed system reached an accuracy of 97%. Related previous studies achieved sensitivity between 70.40% and 95.16%, while our proposed system achieved a sensitivity of 96.80%. The specificity in previous studies ranged between 70.90% and 93%, while our

TABLE 4: Comparison of the performance of our proposed system with models of previous studies.

| Previous studies                  | Accuracy (%) | Sensitivity (%) | Specificity (%) | AUC (%)      |
|-----------------------------------|--------------|-----------------|-----------------|--------------|
| Godkhindi and Gowda [27]          | 88.56        | 88.77           | 87.35           | —            |
| Pozdeev et al. [29]               | 88.00        | 93.00           | 82.00           | —            |
| Bour et al. [50]                  | 87.10        | 87.10           | 93.00           | —            |
| Zhang et al. [51]                 | 85.90        | 87.60           | —               | 86.00        |
| Ribeiro et al. [31]               | 90.96        | 95.16           | 74.19           | —            |
| Min et al. [32]                   | 78.40        | 83.30           | 70.10           | —            |
| Zhu et al. [52]                   | 85.00        | 83.00           | 82.54           | —            |
| Fonollá et al. [53]               | 90.20        | 90.10           | 90.30           | 97.00        |
| Zhang et al. [51]                 | 70.40        | 70.40           | 70.90           | 90.00        |
| <b>Proposed model (AlexNet)</b>   | <b>97.00</b> | <b>96.80</b>    | <b>99.20</b>    | <b>99.98</b> |
| <b>Proposed model (GoogleNet)</b> | <b>96.70</b> | <b>96.60</b>    | <b>99.00</b>    | <b>99.99</b> |
| <b>Proposed model (ResNet-50)</b> | <b>95.00</b> | <b>94.80</b>    | <b>98.80</b>    | <b>99.69</b> |

proposed system reached a specificity of 99%. The proposed system outperformed all previous studies with regard to AUC, as our proposed system reached an AUC of 99.99%. The comparison of the proposed system against the existing models is presented in Figure 7.

## 5. Conclusion

This work provides a robust framework for classifying the gastrointestinal tract diseases in the Kvasir dataset. Deep learning techniques can reduce the probability of developing malignant diseases by aiding in early detection, while also reducing the unnecessary removal of benign tumors. Video endoscopy is the most widely used diagnostic method for diagnosing gastrointestinal polyps, but many human factors lead to improper diagnosis of gastrointestinal diseases. This paper presents three deep learning models, GoogleNet, ResNet-50, and AlexNet, that can direct the doctor's focus to the most important regions that may have been missed. The dataset was divided into 80% for training and 20% for testing and validation. The images were optimised to remove noise and artifacts. The data augmentation technology works by multiplying the images during the training phase for high accuracy. Convolutional layers extract features of shape, colour, and texture. Overall, 9216 features were extracted and passed into the fully connected layers that produced 1000 neurons. The softmax layer produces five classes, classifying each image into one of the five types of gastrointestinal diseases. All three models achieved equally promising results. Advanced deep learning algorithm will be applied in future.

## Data Availability

The data are available in the following link: <https://datasets.simula.no/kvasir/#download>.

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

The authors extend their appreciation to the Deanship of Scientific Research at King Faisal University for funding this research work through the project number NA00095.

## References

- [1] "Latest global cancer data," 2018, [https://www.iarc.fr/wpcontent/uploads/2018/09/pr263\\_E.pdf](https://www.iarc.fr/wpcontent/uploads/2018/09/pr263_E.pdf).
- [2] J.-B. Liu and X.-F. Pan, "Minimizing Kirchhoff index among graphs with a given vertex bipartiteness," *Applied Mathematics and Computation*, vol. 291, pp. 84–88, 2016.
- [3] J.-B. Liu and J. Cao, "Applications of Laplacian spectra for  $n$ -prism networks," *Neurocomputing*, vol. 198, pp. 69–73, 2016.
- [4] A. Khan, M. A. Gul, A. Alharbi, M. I. Uddin, S. Ali, and B. Alouffi, "Impact of lexical features on answer detection model in discussion forums," *Complexity*, vol. 2021, no. 4, 8 pages, 2021.
- [5] Y. Sasaki, R. Hada, and A. Munakata, "Computer-aided grading system for endoscopic severity in patients with ulcerative colitis," *Digestive Endoscopy*, vol. 15, no. 3, pp. 206–209, 2003.
- [6] P. Wang, T. M. Berzin, J. R. Glissen Brown et al., "Real-time automatic detection system increases colonoscopic polyp and adenoma detection rates: a prospective randomised controlled study," *Gut*, vol. 68, no. 10, pp. 1813–1819, 2019.
- [7] J.-B. Liu, X.-F. Pan, J. Cao, and F.-F. Hu, "A note on "some physical and chemical indices of clique-inserted lattices"," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2014, no. 6, Article ID P06006, 2014.
- [8] J.-B. Liu, X.-F. Pan, F.-T. Hu, and F.-F. Hu, "Asymptotic Laplacian-energy-like invariant of lattices," *Applied Mathematics and Computation*, vol. 253, pp. 205–214, 2015.
- [9] J.-B. Liu and X.-F. Pan, "A unified approach to the asymptotic topological indices of various lattices," *Applied Mathematics and Computation*, vol. 270, pp. 62–73, 2015.
- [10] J.-B. Liu and X.-F. Pan, "Asymptotic incidence energy of lattices," *Physica A: Statistical Mechanics and its Applications*, vol. 422, pp. 193–202, 2015.
- [11] Y. Sasaki, R. Hada, T. Yoshimura, N. Hanabata, T. Mikami, and S. Fukuda, "Computer-aided estimation for the risk of development of gastric cancer by image processing," in

- Artificial Intelligence in Theory and Practice III*, pp. 197–204, Springer, Berlin, Germany, 2010.
- [12] S. Amin, M. Irfan Uddin, H. Al-Baity, M. Ali Zeb, M. Abrar Khan, and M. Abrar Khan, “Machine learning approach for covid-19 detection on twitter,” *Computers, Materials & Continua*, vol. 68, no. 2, pp. 2231–2247, 2021.
  - [13] M. P. Tjoa and S. M. Krishnan, “Feature extraction for the analysis of colon status from the endoscopic images,” *Bio-Medical Engineering Online*, vol. 2, no. 1, pp. 9–17, 2003.
  - [14] J. I. Bangash, A. W. Khan, A. Khan, A. Khan, M. I. Uddin, and Q. Hua, “Multiconstraint-aware routing mechanism for wireless body sensor networks,” *Journal of Healthcare Engineering*, vol. 2021, no. 4, 15 pages, 2021.
  - [15] J.-B. Liu and J. Cao, “The resistance distances of electrical networks based on Laplacian generalized inverse,” *Neurocomputing*, vol. 167, pp. 306–313, 2015.
  - [16] J.-B. Liu, X.-F. Pan, L. Yu, and D. Li, “Complete characterization of bicyclic graphs with minimal Kirchhoff index,” *Discrete Applied Mathematics*, vol. 200, pp. 95–107, 2016.
  - [17] J.-B. Liu, W.-R. Wang, Y.-M. Zhang, and X.-F. Pan, “On degree resistance distance of cacti,” *Discrete Applied Mathematics*, vol. 203, pp. 217–225, 2016.
  - [18] T. Hirasawa, K. Aoyama, T. Tanimoto et al., “Application of artificial intelligence using a convolutional neural network for detecting gastric cancer in endoscopic images,” *Gastric Cancer*, vol. 21, no. 4, pp. 653–660, 2018.
  - [19] T. N. Sainath, B. Kingsbury, G. Saon et al., “Deep convolutional neural networks for large-scale speech tasks,” *Neural Networks*, vol. 64, pp. 39–48, 2015.
  - [20] V. Mnih, K. Kavukcuoglu, D. Silver et al., “Human-level control through deep reinforcement learning,” *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.
  - [21] S. A. Karkanis, D. K. Iakovidis, D. E. Maroulis, D. A. Karras, and M. Tzivras, “Computer-aided tumor detection in endoscopic video using color wavelet features,” *IEEE Transactions on Information Technology in Biomedicine*, vol. 7, no. 3, pp. 141–152, 2003.
  - [22] J. Bernal, N. Tajbaksh, F. J. Sánchez et al., “Comparative validation of polyp detection methods in video colonoscopy: results from the MICCAI 2015 endoscopic vision challenge,” *IEEE Transactions on Medical Imaging*, vol. 36, no. 6, pp. 1231–1249, 2017.
  - [23] D. K. Iakovidis, D. E. Maroulis, and S. A. Karkanis, “An intelligent system for automatic detection of gastrointestinal adenomas in video endoscopy,” *Computers in Biology and Medicine*, vol. 36, no. 10, pp. 1084–1103, 2006.
  - [24] L. A. Alexandre, N. Nobre, and J. Casteleiro, “Color and position versus texture features for endoscopic polyp detection,” in *Proceedings of the 2008 International Conference on BioMedical Engineering and Informatics*, May 2008.
  - [25] N. Tajbaksh, S. R. Gurudu, and J. Liang, “Automated polyp detection in colonoscopy videos using shape and context information,” *IEEE Transactions on Medical Imaging*, vol. 35, no. 2, pp. 630–644, 2015.
  - [26] X. Zhang, F. Chen, T. Yu et al., “Real-time gastric polyp detection using convolutional neural networks,” *PLoS One*, vol. 14, no. 3, p. e0214133, 2019.
  - [27] A. M. Godkhindi and R. M. Gowda, “Automated detection of polyps in CT colonography images using deep learning algorithms in colon cancer diagnosis,” in *Proceedings of the 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS)*, pp. 1722–1728, IEEE, Chennai, India, August 2017.
  - [28] T. Ozawa, S. Ishihara, M. Fujishiro, Y. Kumagai, S. Shichijo, and T. Tada, “Automated endoscopic detection and classification of colorectal polyps using convolutional neural networks,” *Therapeutic advances in gastroenterology*, vol. 13, p. 1756284820910659, 2020.
  - [29] A. A. Pozdeev, N. A. Obukhova, and A. A. Motyko, “Automatic analysis of endoscopic images for polyps detection and segmentation,” in *Proceedings of the 2019 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus)*, pp. 1216–1220, IEEE, Moscow, Russia, January 2019.
  - [30] Q. S. Wan, T. Wang, and K. H. Zhang, “Biomedical optical spectroscopy for the early diagnosis of gastrointestinal neoplasms,” *Tumour biology: the Journal of the International Society for Oncodevelopmental Biology and Medicine*, vol. 39, no. 7, p. 1010428317717984, 2017.
  - [31] E. Ribeiro, A. Uhl, and M. Häfner, “Colonic polyp classification with convolutional neural networks,” in *Proceedings of the 2016 IEEE 29 International Symposium on Computer-Based Medical Systems (CBMS)*, pp. 253–258, IEEE, Dublin, Ireland, June 2016.
  - [32] M. Min, S. Su, W. He, Y. Bi, Z. Ma, and Y. Liu, “Computer-aided diagnosis of colorectal polyps using linked color imaging colonoscopy to predict histology,” *Scientific Reports*, vol. 9, no. 1, pp. 2881–2888, 2019.
  - [33] E. M. Song, B. Park, C. A. Ha et al., “Endoscopic diagnosis and treatment planning for colorectal polyps using a deep-learning model,” *Scientific Reports*, vol. 10, no. 1, pp. 30–10, 2020.
  - [34] M. I. Razzak, S. Naz, and A. Zaib, “Deep learning for medical image processing: overview, challenges and the future,” in *Classification in Bioapps: Automation of Decision Making*, N. Dey, A. S. Ashour, and S. Borra, Eds., Springer, Berlin, Germany, pp. 323–350, 2018.
  - [35] F. W. Alsaade, T. H. H. Aldhyani, and M. H. Al-Adhaileh, “Developing a recognition system for classifying covid-19 using a convolutional neural network algorithm,” *Computers, Materials & Continua*, vol. 68, no. 1, pp. 805–819, 2021.
  - [36] A. Chakrabarty, N. Mansoor, M. I. Uddin, M. H. Al-adaileh, N. Alsharif, and F. W. Alsaade, “Prediction approaches for smart cultivation: a comparative study,” *Complexity*, vol. 2021, no. 4, 16 pages, Article ID 5534379, 2021.
  - [37] E. M. Senan, M. E. Jadhav, and A. Kadam, “Classification of PH2 images for early detection of skin diseases,” in *Proceedings of the 2021 6th International Conference for Convergence in Technology (I2CT)*, pp. 1–7, IEEE, Mumbai, India, April 2021.
  - [38] M. Roman, A. Shahid, M. I. Uddin, Q. Hua, and S. Maqsood, “Exploiting contextual word embedding of authorship and title of articles for discovering citation intent classification,” *Complexity*, vol. 2021, no. 4, 13 pages, Article ID 5554874, 2021.
  - [39] A. Aggarwal, V. Sharma, A. Trivedi et al., “Two-way feature extraction using sequential and multimodal approach for hateful meme classification,” *Complexity*, vol. 2021, Article ID 5510253, 2021.
  - [40] D. Singh, V. Kumar, V. Yadav, and M. Kaur, “Deep neural network-based screening model for COVID-19-infected patients using chest X-ray images,” *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 35, no. 3, p. 2151004, 2021.
  - [41] E. M. Senan and M. E. Jadhav, “Classification of dermoscopy images for early detection of skin cancer—a review,”

- International Journal of Computer Application*, vol. 975, p. 8887, 2019.
- [42] M. Ibrahim, A. Al-Mashhadani, T. H. H. Aldhyani et al., "Human-animal affective robot touch classification using deep neural network," *Computer Systems Science and Engineering*, vol. 38, no. 1, pp. 25–37, 2021.
- [43] E. M. Senan, F. W. Alsaade, M. I. A. Al-Mashhadani, H. H. Theyazn, and M. H. Al-Adhaileh, "Classification of histopathological images for early detection of breast cancer using deep learning," *Journal of Applied Science and Engineering*, vol. 24, no. 3, pp. 323–329, 2021.
- [44] E. M. Senan and M. E. Jadhav, "Techniques for the detection of skin lesions in PH2 dermoscopy images using local binary pattern (LBP)," in *Recent Trends in Image Processing and Pattern Recognition. RTIP2R 2020*, K. C. Santosh and B. Gawali, Eds., vol. 1381 Singapore, Singapore, Springer, 2021.
- [45] R. Awan, N. A. Koohbanani, M. Shaban, A. Lisowska, and N. Rajpoot, "Context-aware learning using transferable features for classification of breast cancer histology images," in *Proceedings of the International Conference Image Analysis and Recognition*, pp. 788–795, Póvoa de Varzim, Portugal, June 2018.
- [46] A. Baccouche, B. Garcia-Zapirain, C. Castillo Olea, and A. Elmaghraby, "Ensemble deep learning models for heart disease classification: a case study from Mexico," *Information*, vol. 11, no. 4, p. 207, 2020.
- [47] E. M. Senan and M. E. Jadhav, "Analysis of dermoscopy images by using ABCD rule for early detection of skin cancer," *Global Transitions Proceedings*, vol. 2, no. 1, 2021.
- [48] F. W. Alsaade, T. H. H. Aldhyani, and M. H. Al-Adhaileh, "Developing a recognition system for diagnosing melanoma skin lesions using artificial intelligence algorithms," *Computational and Mathematical Methods in Medicine*, vol. 2021, Article ID 9998379, 20 pages, 2021.
- [49] M. N. Y. Ali, M. G. Sarowar, M. L. Rahman, J. Chaki, N. Dey, and J. M. R. S. Tavares, "Adam deep learning with SOM for human sentiment classification," *International Journal of Ambient Computing and Intelligence*, vol. 10, no. 3, pp. 92–116, 2019.
- [50] A. Bour, C. Castillo-Olea, B. Garcia-Zapirain, and S. Zahia, "Automatic colon polyp classification using convolutional neural network: a case study at Basque country," in *Proceedings of the 2019 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*, pp. 1–5, IEEE, Ajman, UAE, December 2019.
- [51] R. Zhang, Y. Zheng, T. W. C. Mak et al., "Automatic detection and classification of colorectal polyps by transferring low-level CNN features from nonmedical domain," *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 1, pp. 41–47, 2016.
- [52] R. Zhu, R. Zhang, and D. Xue, "Lesion detection of endoscopy images based on convolutional neural network features," in *Proceedings of the 2015 8th International Congress on Image and Signal Processing (CISP)*, October 2015.
- [53] R. Fonollá, F. Van Der Sommen, R. M. Schreuder, E. J. Schoon, and P. H. De With, "Multi-modal classification of polyp malignancy using CNN features with balanced class augmentation," in *Proceedings of the 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pp. 74–78, IEEE, Venice, Italy, April 2019.