

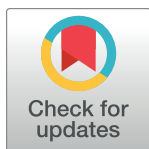
RESEARCH ARTICLE

Gastrointestinal tract disorders classification using ensemble of InceptionNet and proposed GITNet based deep feature with ant colony optimization

Muhammad Ramzan¹, Mudassar Raza^{2*}, Muhammad Irfan Sharif³, Faisal Azam¹, Jungeun Kim^{4*}, Seifedine Kadry^{5,6,7,8}

1 Department of Computer Science, COMSATS University Islamabad, Wah Campus, Pakistan, **2** Department of Computer Science, HITEC University Taxila, Taxila, Pakistan, **3** Department of Information Sciences, University of Education Lahore, Jauharabad Campus, Jauharabad, Pakistan, **4** Department of Software and CMPSI, Kongju National University, Cheonan, Korea, **5** Department of Applied Data Science, Noroff University College, Kristiansand, Norway, **6** Artificial Intelligence Research Center (AIRC), Ajman University, Ajman, United Arab Emirates, **7** Department of Electrical and Computer Engineering, Lebanese American University, Byblos, Lebanon, **8** MEU Research Unit, Middle East University, Amman, Jordan

* mudassarkazmi@yahoo.com (MR); jekim@kongju.ac.kr (JK)



OPEN ACCESS

Citation: Ramzan M, Raza M, Sharif MI, Azam F, Kim J, Kadry S (2023) Gastrointestinal tract disorders classification using ensemble of InceptionNet and proposed GITNet based deep feature with ant colony optimization. PLoS ONE 18(10): e0292601. <https://doi.org/10.1371/journal.pone.0292601>

Editor: Mujeeb Ur Rehman, York St John University, UNITED KINGDOM

Received: July 25, 2023

Accepted: September 24, 2023

Published: October 13, 2023

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0292601>

Copyright: © 2023 Ramzan et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: A minimal data set can be found via: Pogorelov, K., Randel, K. R., Griwodz, C., Eskeland, S. L., de Lange, T.,

Abstract

Computer-aided classification of diseases of the gastrointestinal tract (GIT) has become a crucial area of research. Medical science and artificial intelligence have helped medical experts find GIT diseases through endoscopic procedures. Wired endoscopy is a controlled procedure that helps the medical expert in disease diagnosis. Manual screening of the endoscopic frames is a challenging and time taking task for medical experts that also increases the missed rate of the GIT disease. An early diagnosis of GIT disease can save human beings from fatal diseases. An automatic deep feature learning-based system is proposed for GIT disease classification. The adaptive gamma correction and weighting distribution (AGCWD) preprocessing procedure is the first stage of the proposed work that is used for enhancing the intensity of the frames. The deep features are extracted from the frames by deep learning models including InceptionNetV3 and GITNet. Ant Colony Optimization (ACO) procedure is employed for feature optimization. Optimized features are fused serially. The classification operation is performed by variants of support vector machine (SVM) classifiers, including the Cubic SVM (CSVM), Coarse Gaussian SVM (CGSVM), Quadratic SVM (QSVM), and Linear SVM (LSVM) classifiers. The intended model is assessed on two challenging datasets including KVASIR and NERTHUS that consist of eight and four classes respectively. The intended model outperforms as compared with existing methods by achieving an accuracy of 99.32% over the KVASIR dataset and 99.89% accuracy using the NERTHUS dataset.

Johansen, D., . . . Halvorsen, P. (2023). KVASIR: A Multi-Class Image Dataset for Computer Aided Gastrointestinal Disease Detection. doi: [10.1145/3193289](https://doi.org/10.1145/3193289).

Funding: The author(s) received no specific funding for this work.

Competing interests: he authors have declared that no competing interests exist.

1. Introduction

The GIT is a vital part of the digestive system that is vulnerable to many infections. Infections can lead to serious illness if not detected and treated at early stages [1]. Colorectal cancer is a type that can be prevented in most cases with effective measures such as colorectal lymph nodes that can be detected, diagnosed, and treated in early stages [2]. Different kinds of cancer usually result in 70% of deaths in middle-income countries due to not being detected and treated in the early stages [3–5]. The countless distinct irregular mucosal outcomes may differ from moderate annoyances to deadly diseases in the GIT [6]. According to statistics compiled by the World Health Organization (WHO), a huge number (1.93 million) cases of colon and rectum cancer were suffering around the globe. The WHO reported a large number of death cases (916 000) and 1.09 million cases of stomach cancer in 2020 [7]. The insufficient understanding of the standardized classification of such cancers by endoscopic imagery contributes to the rise of the enormous number of fatalities worldwide. Endoscopic methods like wireless capsule endoscopy (WCE) and wired endoscopy (WE) are used to find illnesses in the GIT [8]. The WCE procedure is time-consuming and out of control for the operator, whereas the WE is a controlled procedure. An operator can focus on a specific region of the GIT by adopting the WE procedure [9]. Colonoscopy is used to scan the lower portion of the GIT, and WE procedures are used to examine both (the lower and upper) portions of the GIT [10]. Through the endoscopic procedure, videos are examined and captured by the medical expert for disease analysis and detection [11].

Multiple challenges are found in the GIT images, including huge amounts of clutter, occlusions, complex backgrounds, variable lighting conditions, huge inter-class similarities among objects of different classes, and huge intra-class variations among the objects in a single class. An experienced medical specialist is required to handle these challenges. Also, the challenging task for the medical expert is to examine the large number of video frames which increases the workload of the medical experts. An automatic system is required that can classify the frames of GIT diseases accurately and take less time as compared with the gastroenterologist. It is a challenging job to develop an automated procedure using machine learning algorithms. A precise Computer Aided Diagnostic System (CADx) can aid specialists in disease classification decision-making [12–14]. The accuracy of the disease classification is the most important factor [15]. In recent years, various artificial intelligence (AI) systems have come into being for diagnosing different kinds of diseases in GIT. CADx has become a very important part of research in medical and computer science for disease classification by using an imagery dataset. Deep learning methods are growing rapidly in the healthcare system for disease classification [16–18]. Automatic disease detection and classification approach is introduced [19]. The new emerging and popular deep learning models (InceptionNetV3, VGG, ResNet, and a googleNet) are being used in the research work that are already trained on the millions of imagery datasets for disease detection and classification. Convolutional neural networks (CNNs) are emerging tools in machine learning for deep feature learning and classification [20]. The main goal of this research is to introduce the computer-aided GIT disease classification system which assists the healthcare system in reducing death rates by diagnosing GI tract diseases at an early stage.

In this manuscript, the deep learning-based GIT disease classification system is proposed by using endoscopic frames. Based on our extensive experimentation, we finalized the proposed model in which the proposed preprocessing technique (AGCWD) is performed, and the resultant image is fed to the CNN-based networks. To obtain better results, out of two CNN networks, one (InceptionNetV3) is the pre-trained network, while the other (GITNet) is designed by ourselves to acquire features from both networks. After that features are optimized

by the ACO technique and fused serially. In the last stage, classification is performed by variants of SVM classifiers. The results show the effectiveness of the proposed solution. The AGCWD makes the frames more appealing. The typical procedure for feature extraction is performed by deep learning models, including InceptionNetV3 and GITNet. The deep learning models encode features automatically. Further features are optimized and fused serially. After that, the fused features are supplied to the classifiers. The classifiers including CGSVM, LSVM, CSVM, and QSVM are employed, whereas the QSVM classifier provides better results in terms of accuracy. The major contributions of the article are stated as follows:

- The technique AGCWD is employed as a preprocessing phase for enhancing the pixel intensity level, which helps to get better features for disease classification.
- The proposed deep learning model GITNet is trained and tested by using a third-party dataset i.e. CIFAR-10 [21]. The features are acquired from the fully connected layers of GITNet and an existing pre-trained model named as InceptionNet.
- A bio-inspired approach name ACO is applied for feature subset selection and features are fused serially and passed to the SVM-based classifiers for classification.

The organization of manuscript after the introduction section is organized as: The associated work is specified in Section 2. The material and methods are explained in Section 3. The results of the experiments are elaborated on in Section 4. Lastly, Section 6 concluded the manuscript and describes future work.

2. Related works

The most challenging problem for gastroenterologists is differentiating between healthy and harmful images, such as ulcers, bleeding, and polyps [22,23]. The endoscopic frames are classified using different methods that have been proposed. The most recent methods for endoscopic frame classification are addressed in this section. The image acquisition and preprocessing approaches are briefly described. The images of the GIT disease are acquired by the procedure of endoscopy. Most of the endoscopic frames contain low-intensity pixels and noise [24]. The noise is eliminated by the thresholding segments of the frames and image clipping methods. The image clipping divides the background into tiny noise spots like edge detection. Simple and median filters reduce the noise from the frames by designing an effective noise filter [25]. Many color processing methods use various color spaces including RGB [26], CIEXYZ [27], CIELAB [28], YIQ [29], HSV [30], YUV [31], and HIS [32]. Statistics of color and moments are used as color space. The preprocessed frames are processed further to extract features.

A variety of feature extraction approaches are employed in the existing work, including point features that represent the features of objects geographically [33–35], texture features that define the object surface as fine, coarse, smooth, and grained [36], HOG features focuses on shapes of the objects [37], and color features [38]. A combination of deep features enhances the performance of the system [39]. In one research work, deep features of two CNN models are also integrated [40]. CNN models extract the features from the deep layers [41], the model including AlexNet [42], VGG-16 [43], ResNet [44], and InceptionNetV3 [45]. The distributed deep learning method is employed for disease detection [46]. Diseases of GI tract are detected and classified by using a deep learning model for feature extraction [47]. Deep features extract the information and synthesize the information using skip connections from the previous layers [48]. The most appropriate features offer high accuracy in categorization outcomes. Mostly, the redundant features are present in the extracted feature set which affects the classification

results. To remove the redundancy of the features and reduce the computation cost of the classifiers, the ACO method is adopted to optimize the features. ACO optimizes the features for ulcerated lesion classification work [49]. The metaheuristic technique is employed for feature optimization [50]. There are three methods for feature selection and dimensionality reductions including the filtering method mapping relationships between input and target variables using a statistical approach, the wrapper method is the specific machine learning technique for feature selection, and the embedded method integrates the wrapper and filter method [51]. The feature fusion of automated systems in computer vision and image processing for illness classification tasks is crucial and supports the achievement of more accurate and effective results [40]. In the hybrid approach, several pieces of literature address handcrafted features and deep features being fused to provide better results. In this approach, deep features are fused for stomach disease classification [52]. In the existing work, the class of techniques that utilize the powers of hybrid methodologies is increasing. The different classifiers are trained by using fused features of handcrafted and deep features [53]. The KNN with SVM classifier is employed for GIT disease classification [54]. For disease classification, the k-Nearest Neighbor (KNN) with SVM classifier is employed [55]. The CNN with multi-scale feature fusion approach is applied for colon cancer identification [56]. VGG16 and VGG19 are employed with handcrafted methods in hybrid mode for GIT disease classification [16].

The existing literature reveals that researchers employ a hybrid approach for the classification of GIT diseases and obtained less accuracy. The technique that deals with the fusion of deep features is used very rarely. So, there is still a need to design a framework that can more accurately differentiate endoscopic frames of GIT diseases.

3. Proposed methodology

The proposed approach comprises five stages are presented in this section. In the first stage, the technique AGCWD is used for image enhancement. A new CNN-based GITNet model is designed in the second stage and deep features are obtained by the CNN-based pre-trained InceptionNetV3 and GITNet models. Features are optimized in the third stage by the ACO method. The serial feature fusion procedure is adopted in the fourth stage. In the last fifth stage, classifiers are used to classify the diseases of GIT using fused feature sets. An impression of the proposed GIT disease classification approach is depicted in Fig 1.

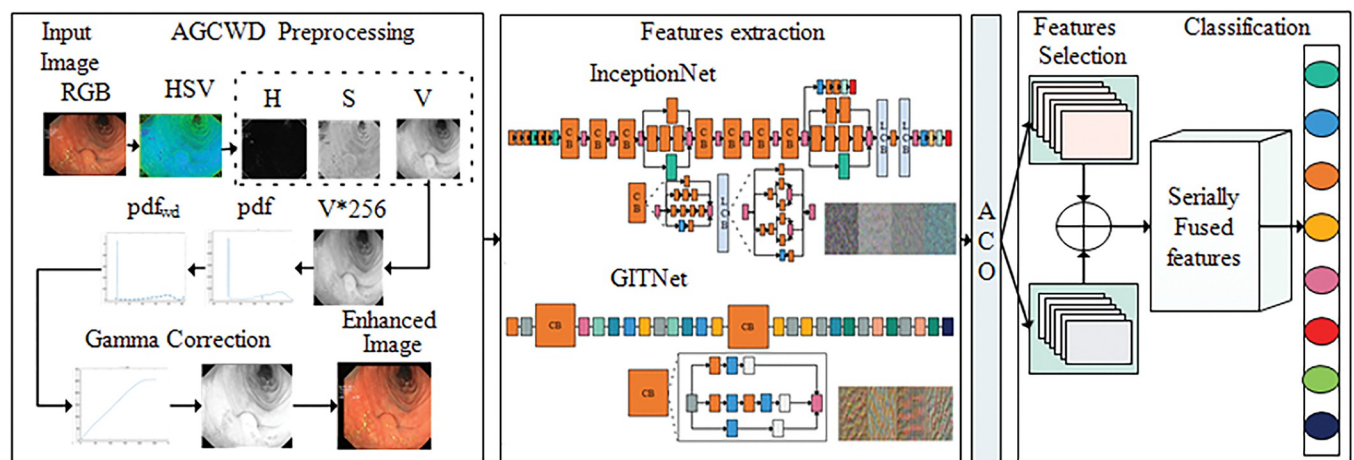


Fig 1. The proposed method of disease classification.

<https://doi.org/10.1371/journal.pone.0292601.g001>

Algorithm: The Proposed Approach for GIT disease classification

```

Input: Inp(x,y) // Endoscopic Input frame
          N: Total number of images in the dataset
Output: Out(x,y) // Classified Output Image
Step 1: START
Step 2: // AGCWD image method
          Enhance the dataset images using AGCWD method
          See Eqs (1) to (5)
Step 3: // Extract Deep CNN features of GITNet and InceptionNetV3
          Model
          for K to 1→N do
              FV1 = InceptionNetV3(K) //Features are extracted from fc7 layer
              FV2 = GITNet(K) //Features are extracted from fc12 layer
          End for
Step 4: //Feature selection using Ant colony optimization (ACO)
          for K to 1→N do
              Ofs1[m] = ACO(FV1)
              Ofs2[m] = ACO(FV2)
          End for
Step 5: //Feature Fusion
          fusedFeatureVector = Ofs1[m]+Ofs2[m]
Step 6: //Classify the output of Step 5
          Prediction = SVM(fusedFeatureVector)
Step 7: STOP

```

3.1 Preprocessing

Preprocessing is essential for improving the visualization of the frames of endoscopy. So, the enhancement of the contrast of images is a significant and vital stage for disease classification. This phase's major benefit is that it offers more powerful and relevant features for accurate classification. In this work, an Adaptive-Gamma-Correction (AGC) is employed with the weighting distribution (WD) function referred to as AGCWD, which enhances the intensity of the frames [57]. In the AGCWD method, the transformation of color space is performed from RGB to HSV color space. The color contents in the HSV are specified by the hue (H) and the saturation (S) that are preserved, whereas the luminance intensity (V) component is tweaked for color contrast enhancement as shown in Fig 2.

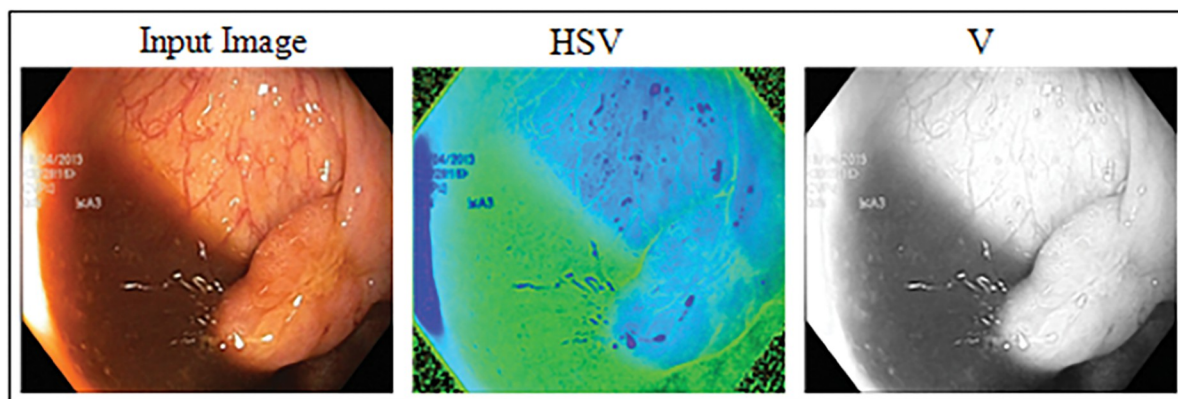


Fig 2. Transformation of the output from RGB to HSV color space and increasing intensity of V channel.

<https://doi.org/10.1371/journal.pone.0292601.g002>

The function of WD is to marginally alter the statistical histogram and reduce the development of negative consequences. The function of WD is formulated as:

$$pdf_{WD}(I) = pdf_{\max} \left(\frac{pdf(I) - pdf_{\min}}{pdf_{\max} - pdf_{\min}} \right)^{\alpha} \quad (1)$$

where $pdf_{WD}(I)$ is the probability-density-function (pdf) of the weighting distribution of the input image; A term α is the adjusted parameter. The $pdf_{\max}$ and $pdf_{\min}$ are the maximum and minimum pdf of the statistical histogram respectively. The cumulative-distribution-function (cdf) is dependent on pdf and is expressed as:

$$cdf_{WD}(I) = \sum_{I=0}^{I_{\max}} (pdf_{WD}(I) / \sum pdf_{WD}) \quad (2)$$

where the quantity pdf_{WD} is stated as follows:

$$\sum pdf_{WD} = \sum_{I=0}^{I_{\max}} pdf_{WD}(I) \quad (3)$$

The plot of the pdf and pdf_{WD} is illustrated in Fig 3.

AGC function uses cdf and pdf for the intensity transformation. The formulation of the AGC is given as:

$$P(I) = I_{\max} (I/I_{\max})^{\gamma} = I_{\max} (I/I_{\max})^{I-cdf(I)} \quad (4)$$

where $I_{\max}$ presents the maximum intensity of the input and $P(I)$ is the transformed intensity of each pixel of the image. The main advantage of the AGC is to gradually increase the intensity of low-intensity pixels while avoiding a major decrease in the intensity of high-intensity pixels. The gamma parameter is formulated as follows:

$$\gamma = I - cdf_{WD}(I) \quad (5)$$

The plot of the Luminance intensity, the output of the AGC function $P(I)$, and the enhanced image after the gamma operation is depicted in Fig 4.

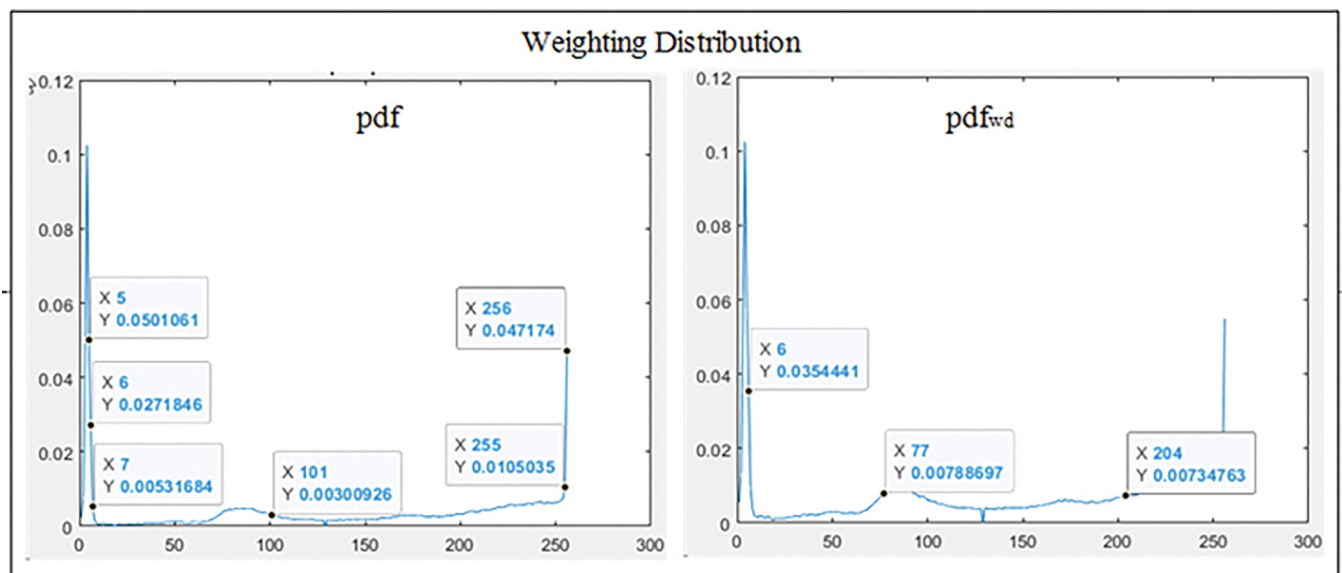


Fig 3. The plot of the pdf and weighting distribution pdf.

<https://doi.org/10.1371/journal.pone.0292601.g003>

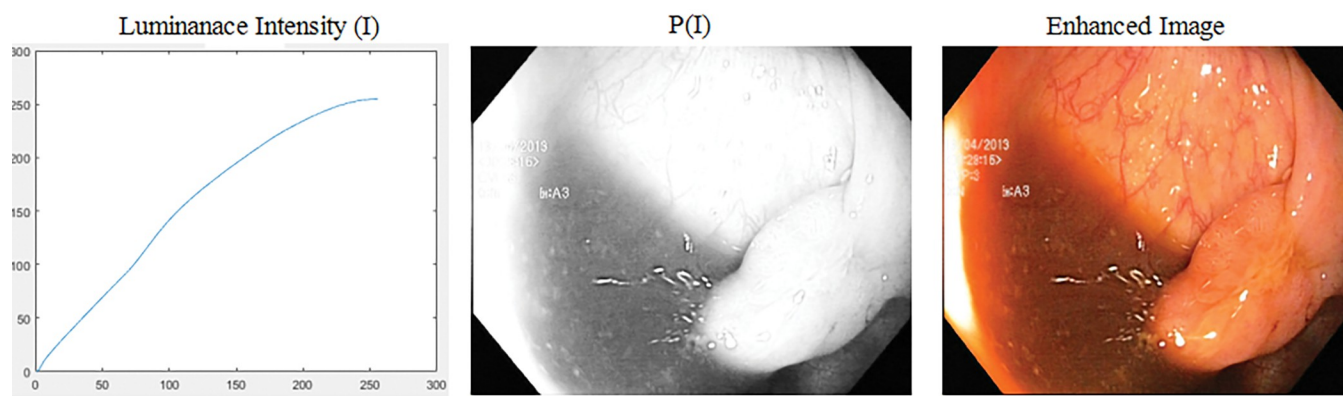


Fig 4. The output of the adaptive gamma Correction with WD.

<https://doi.org/10.1371/journal.pone.0292601.g004>

The AGCWD preprocessing method is utilized over the KVASIR dataset. Some enhanced frames of the KVASIR dataset are illustrated in Fig 5.

3.2 Deep feature extraction

Deep features are very important for computer vision tasks that are acquired from the endoscopic frames. Deep learning is an emerging technology that is participated with computer vision and image processing for disease classification [58]. A CNN consists of different layers (Input, Convolutional, Batch Normalization, Fully Connected, and ReLU layers). The convolutional layer gets the data from the CNN input layers and the weights of the network are calculated. The inactive neurons are detached, and the activation functions are used via the ReLU layer. In this work, two CNN models, InceptionNetV3 and GITNet, are employed for feature extraction. The following sections describe the deep learning architectures.

3.2.1 InceptionNetV3 model. The InceptionNetV3 is a deep learning architecture that performs well for categorization. The architecture consists of a directed acyclic graph (DAG) having 316 layers, 350 connections, and 94 convolutional layers [59]. The different significant features are obtained by applying several masks on different layers of the model. As compared with conventional CNN models, InceptionNetV3 is a diverse model that allows the

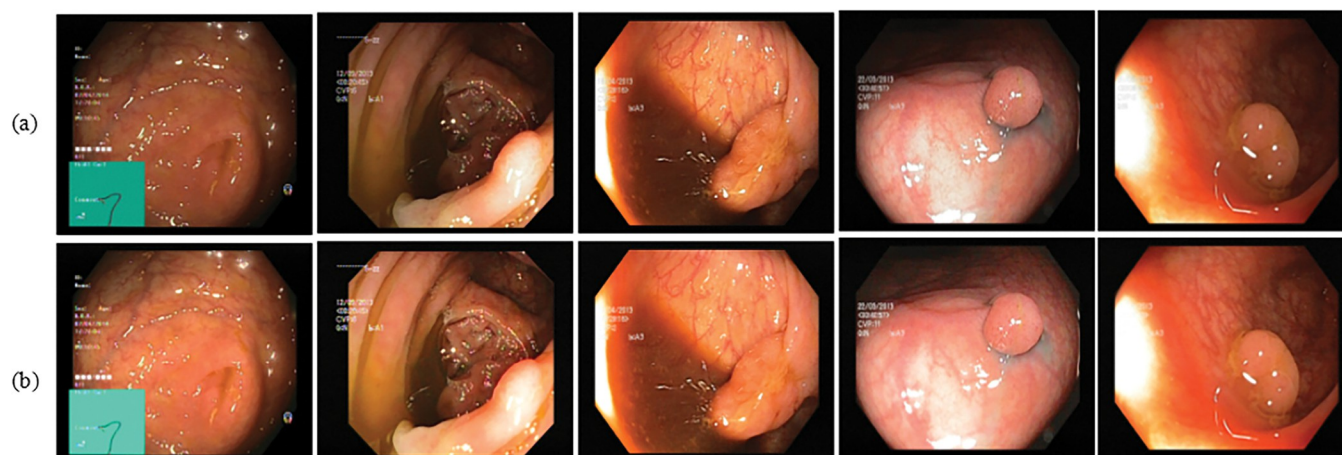


Fig 5. Preprocessed images (a) original images (b) Enhanced images.

<https://doi.org/10.1371/journal.pone.0292601.g005>

implementation of different sizes of masks on different layers. The InceptionNetV3 was trained on a challenging dataset such as ImageNet. It contains millions of images and over 1000 classes. The size of $299 \times 299 \times 3$ was the input of the InceptionNetV3 model. The whole network was composed of nine Inception modules that allow the operation of pooling and convolution with varying filter sizes.

In this manuscript, the approach of deep learning is employed. A new deep learning model takes a much time for training a model while a pre-trained deep learning model saves time for feature extraction. The input images are fed to the convolutional layer to get the feature maps in the training phase. Different filter sizes are employed in the model, including 1×1 , 3×3 , and 5×5 filter sizes. The large kernel size is considered suitable for collecting information that is distributed globally in the frames, while the small kernel size collects information locally. The activation functions are performed for scaling data and features are derived from the different blocks of the model. The InceptionNetV3 characterizes the convolutional block (CB) and large convolutional blocks (LCB) respectively, where feature maps from distinct paths are concatenated as the following module's input. The features with a dimension of (4000×2048) are taken from the fc7 layers. These features are optimized and fused before the training of the classifiers. By the end of the inception module, this network additionally employs global average pooling and fully connected structures. The softmax layer is used at the end of the model with a 1000-way global average pooling layer and fully connected layers. The correlation statistics are assessed step by step as the best network topology is built, resulting in highly correlated outputs when formulated as:

$$X_{InceptionNet} = \{s | s \in R^{MXN}\} \quad (6)$$

where $X_{InceptionNet}$ describes the feature set and the dimension of the feature set is kept MXN . Fig 6 illustrates the InceptionNetV3.

3.2.2 Proposed GITNet model. GITNet is the proposed model in our work that is designed based on the CNN method for deep feature extraction. The endoscopic frames are used as input for the model. The input image size for GITNet is kept at $227 \times 227 \times 3$. The model contains a 50-layer network which contains different layers such as convolutional (C) layers,

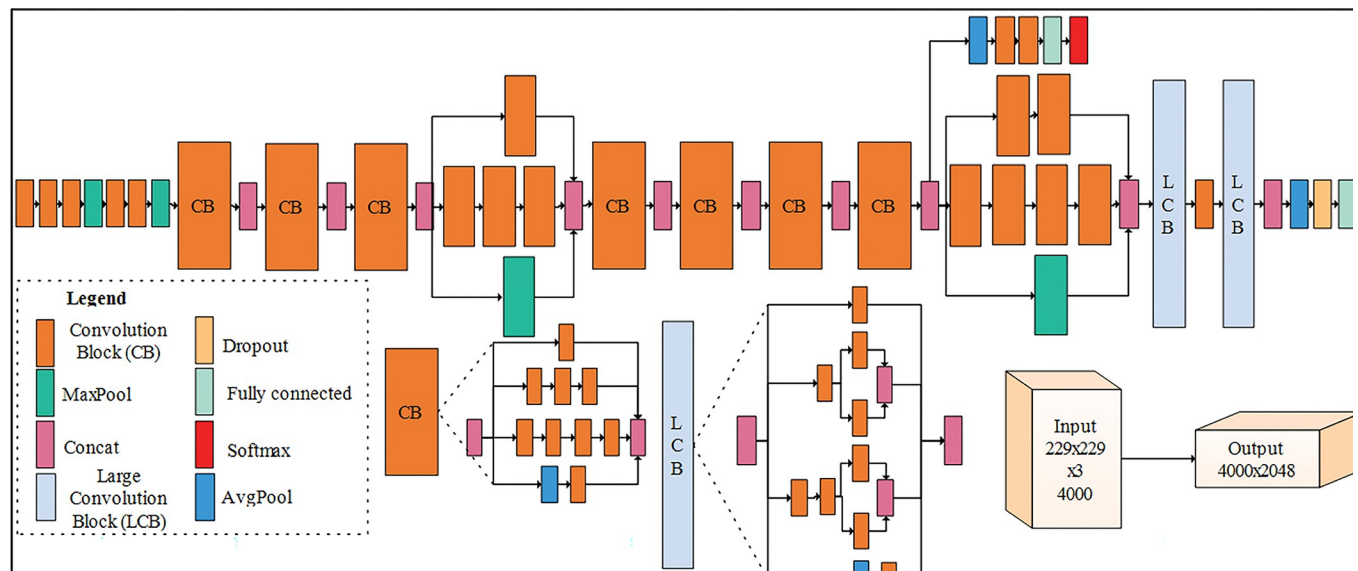


Fig 6. The architecture of the InceptionNetV3 Model (redrawn with a new style from [60]).

<https://doi.org/10.1371/journal.pone.0292601.g006>

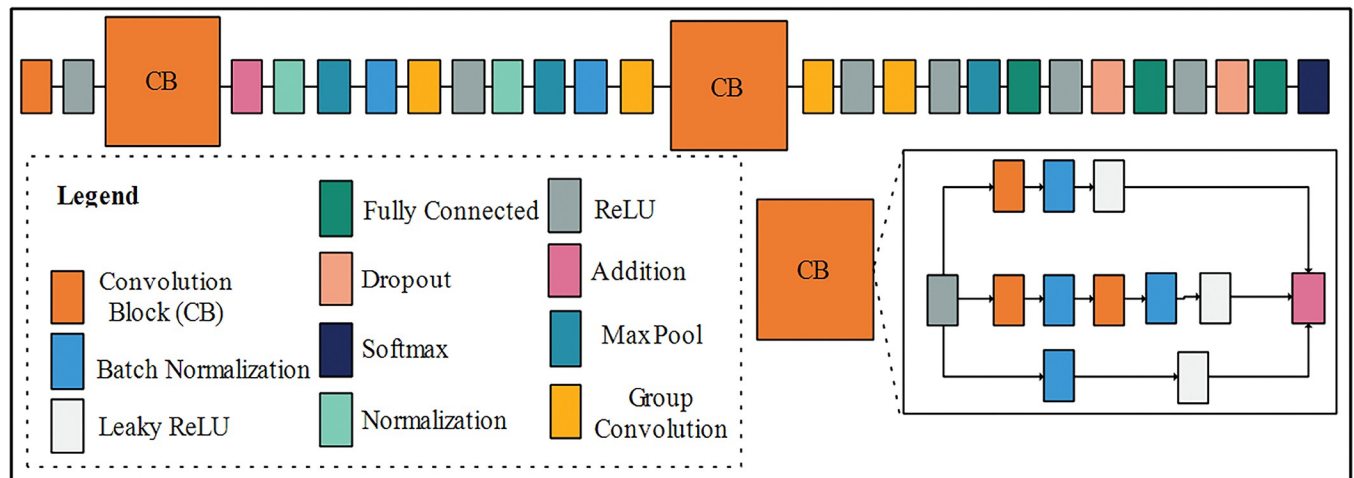


Fig 7. The proposed GITNet model.

<https://doi.org/10.1371/journal.pone.0292601.g007>

ReLU, Batch Normalization (BN) layer, Leaky ReLU layer, addition (A) layer, and two convolutional blocks (CB). Each CB of the GITNet model comprises three parallel networks that are composed of C layer, BN layer, Leaky ReLU layer, and an A layer. Each layered network in the CB gets the same input and derives dissimilar feature windows. The feature maps are obtained from the CB and summed up by the adding layer and forwarded to the following layers and CB for deriving more deep features. Feature maps determine disease information about GIT. The features are acquired from the fc12 layers of the GITNet. The proposed GITNet model is depicted in Fig 7.

The dimensionality of the extracted feature from the fc12 layer is achieved as 4000x4096. The complete feature set obtained from the GITNet is formulated as follows:

$$Y_{GITNet} = \{t | t \in R^{MXN}\} \quad (7)$$

where Y_{GITNet} describes the feature set with MXN dimension. The variants of filter sizes, including filter depth, max-pooling, and stride size over the network are used for dealing with a large variety of objects having features in the frames. The GITNet explores semantic information and mutual information by employing varying depths of filter and filter size. The layered summary of the complete GITNet model is depicted in Table 1. The different filter sizes of 5x5, 1x1, and 3x3 are used with changing padding and stride size such as [4 4], [1 1], [2 2] and padding [0], [same], [2] and [1]. Similarly, the number of filters (filter depth) is changed by convolution layers such as 96, 48,256, 384, and 256. Fully connected layers (fc12, fc13) comprise 1x1x4096 and the fc14 dimension is 1x1x100. Pooling window size or other values are set to scale 0.01 and Max pooling 3x3 is set. The complete GITNet model is composed of 50 layers network.

The visualization of features in different layers of the GITNet model is shown below. The features are obtained from group convolution and convolution layers for visualization. The layers are stated as GC2 (C6), C7, C9, and GC3 (C10) of the GITNet model. The visualization of features of the InceptionNetV3 is illustrated below, as the list of layers (conv2d_90, conv2d_75, conv2d_93, and conv2d_94 layers). Fig 8 shows the picturing of the features of both deep learning models, including GITNet and InceptionNetV3.

The gradient-weighted class activation mapping (Grad-CAM) is employed for the evaluation of the deep learning model. The Grad-CAM shows how well the GITNet model has been

Table 1. The detailed layered information of GITNet Architecture.

Layer No	Layers	Details of Feature maps	Filter information	Stride Size	Padding size	Pooling layers size
1	Input	227×227×3	11×11×3×96	[4 4]	[0 0 0 0]	-
2	C1	55×55×96	-	-	-	-
3	R1	55×55×96	-	-	-	-
4	C4	55×55×96	5×5×96×96	[1 1]	Same	-
5	C2	55×55×48	1×1×96×48	[1 1]	Same	-
6	BN4	55×55×96	-	-	-	-
7	LR2	55×55×96	-	-	-	Scale 0.01
8	BN1	55×55×48	-	-	-	-
9	C3	55×55×96	11×11×48×96	[1 1]	Same	-
10	BN3	55×55×96	-	-	-	-
11	LR1	55×55×96	-	-	-	-
12	BN2	55×55×96	-	-	-	-
13	LR3	55×55×96	-	-	-	Scale 0.01
14	ADD1	55×55×96	-	-	-	-
15	CN1	55×55×96	-	-	-	-
16	P1	27×27×96	-	[2 2]	[0 0 0 0]	Max pooling 3×3
17	BN5	27×27×96	-	-	-	-
18	GC1(C5)	27×27×256	Two groups of 5×5×48×128	[1 1]	[2 2 2 2]	-
19	R2	27×27×256	-	-	-	-
20	CN2	27×27×256	-	-	-	-
21	P2	13×13×256	-	[2 2]	[0 0 0 0]	Max pooling 3×3
22	BN6	13×13×256	-	-	-	-
23	GC2(C6)	13×13×384	3×3×256×384	[1 1]	[1 1 1 1]	-
24	R3	13×13×384	-	-	-	-
25	C7	13×13×192	1×1×384×192	[1 1]	Same	-
26	C9	13×13×384	3×3×384×384	[1 1]	Same	-
27	BN8	13×13×384	-	-	-	-
28	LR6	13×13×384	-	-	-	Scale 0.01
29	BN10	13×13×384	-	-	-	-
30	LR5	13×13×384	-	-	-	Scale 0.01
31	BN7	13×13×384	-	-	-	-
32	C8	13×13×384	5×5×192×384	[1 1]	Same	-
33	BN9	13×13×384	-	-	-	-
34	LR4	13×13×384	-	-	-	Scale 0.01
35	ADD2	13×13×384	-	-	-	-
36	GC3(C10)	13×13×384	Two groups of 3×3×192×192	[1 1]	[1 1 1 1]	-
37	R4	13×13×384	-	-	-	-
38	GC4(C11)	13×13×256	Two groups of 3×3×192×128	[1 1]	[1 1 1 1]	-
39	R5	13×13×256	-	-	-	-
40	P3	6×6×256	-	[2 2]	[0 0 0 0]	Max pooling 3×3
41	BN11	6×6×256	-	-	-	-
42	FC12	1×1×4096	-	-	-	-
43	R6	1×1×4096	-	-	-	-
44	D1	1×1×4096	-	-	-	50% Dropout
45	FC13	1×1×4096	-	-	-	-

(Continued)

Table 1. (Continued)

Layer No	Layers	Details of Feature maps	Filter information	Stride Size	Padding size	Pooling layers size
46	R7	1×1×4096	-	-	-	-
47	D2	1×1×4096	-	-	-	-
48	FC14	1×1×100	-	-	-	-
49	SOFTMAX	1×1×100	-	-	-	-
50	OUTPUT	-	-	-	-	-

<https://doi.org/10.1371/journal.pone.0292601.t001>

learned by the feature pattern. It is visually confirmed by observing the network and ensuring that it has the right patterns in the image and activating around them. The gradients of the prediction score concerning the final convolutional feature map that is used in the Grad-CAM interpretability technique. Parts of an image with a high Grad-CAM map value have the most impact on the network score for that class. Fig 9 shows the high Grad-CAM map value in the images.

3.3 Feature selection and fusion

The extracted features from the GITNet and InceptionNetV3 are optimized by the ACO method which diminished the redundant features in the $X_{InceptionNet}$ Y_{GITNet} feature sets. The extracted features contain a lot of redundant information which causes a reduction in the model performance. Features are optimized by the ACO algorithm which eliminates the unnecessary information in the sample set by using the probabilistic technique. The advantage

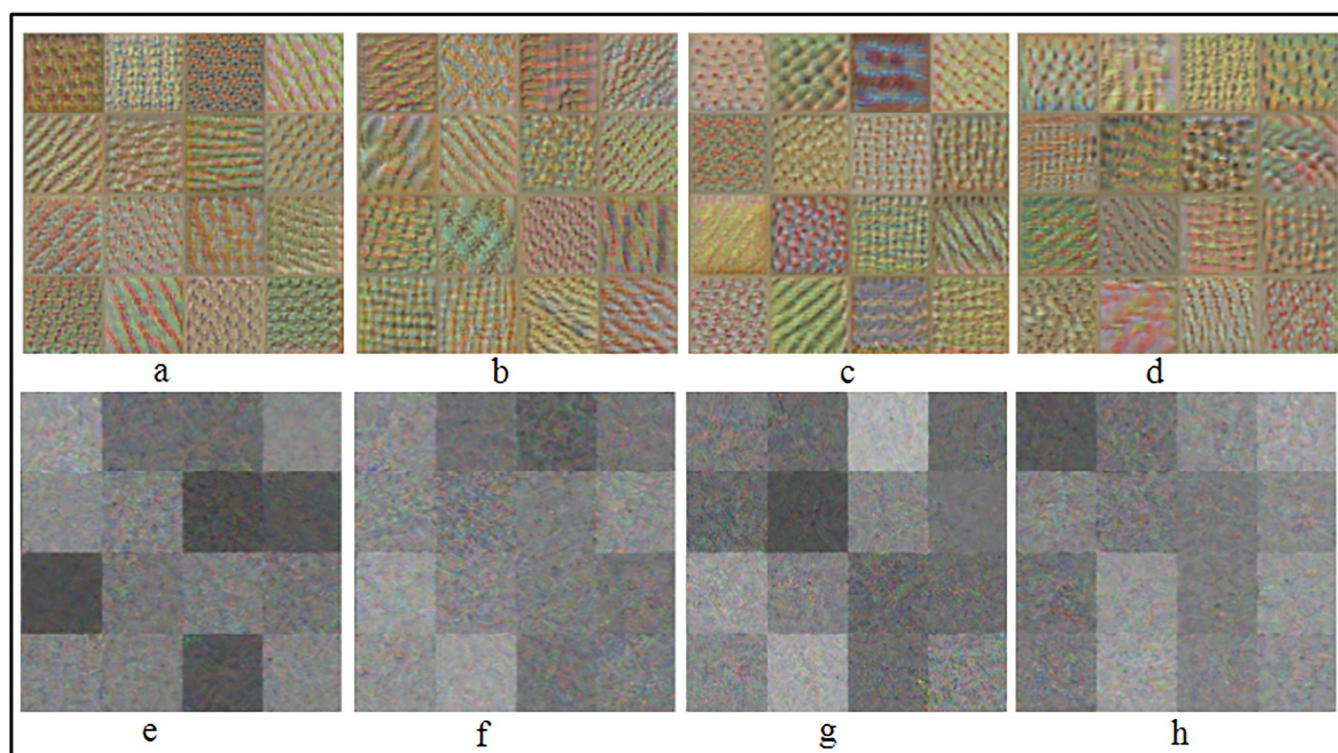


Fig 8. Deep features visualization of GITNet model from a to d with layers name (a) GC2(C6) (b) C7 (c) C9 (d) GC3(C10) and deep features visualization of InceptionNetV3 model from e to h with layer name (e) conv2d_75 (f) conv2d_90 (g) conv2d_93 (h) conv2d_94.

<https://doi.org/10.1371/journal.pone.0292601.g008>

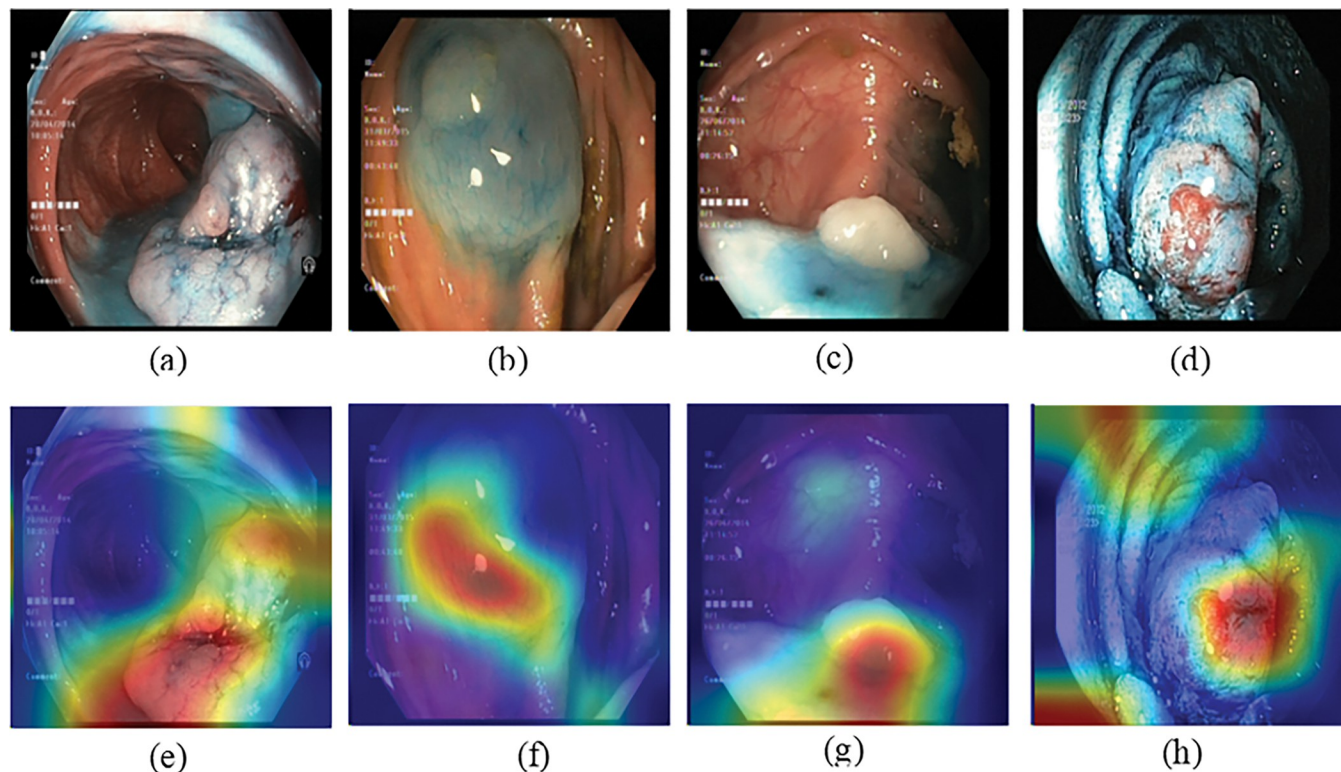


Fig 9. Grad-CAM map value of KVASIR dataset frames (a-d) Original and (e-h) Output.

<https://doi.org/10.1371/journal.pone.0292601.g009>

of using an optimizer is that it reduces the computation cost of the classifier. The mathematical formulation of the two-deep learning models is as follows:

$$\varsigma_{InceptionNet} = ACO(X_{InceptionNet}) \quad (8)$$

$$\varsigma_{GITNet} = ACO(Y_{GITNet}) \quad (9)$$

where the notation $\varsigma_{InceptionNet}$ and ς_{GITNet} are the optimized features of InceptionNetV3 and GITNet that are obtained from the ACO method. After that, the optimized features are fused serially and mathematically as follows:

$$\Phi = \varsigma_{InceptionNet} + \varsigma_{GITNet} \quad (10)$$

$$\Phi = \{\tau | \tau \in R^{MXN}\} \quad (11)$$

where the symbol Φ describes the combined form of the features of two-deep learning models with the dimensions MXN . The detailed form of the feature fusion method is depicted in Fig 10. The same input image with a different size is given to both deep learning models. The GITNet and the InceptionNetV3 explore the deep features by using the CNN method. The extracted features are optimized by the ACO method and fused serially. The classifiers are trained on these features.

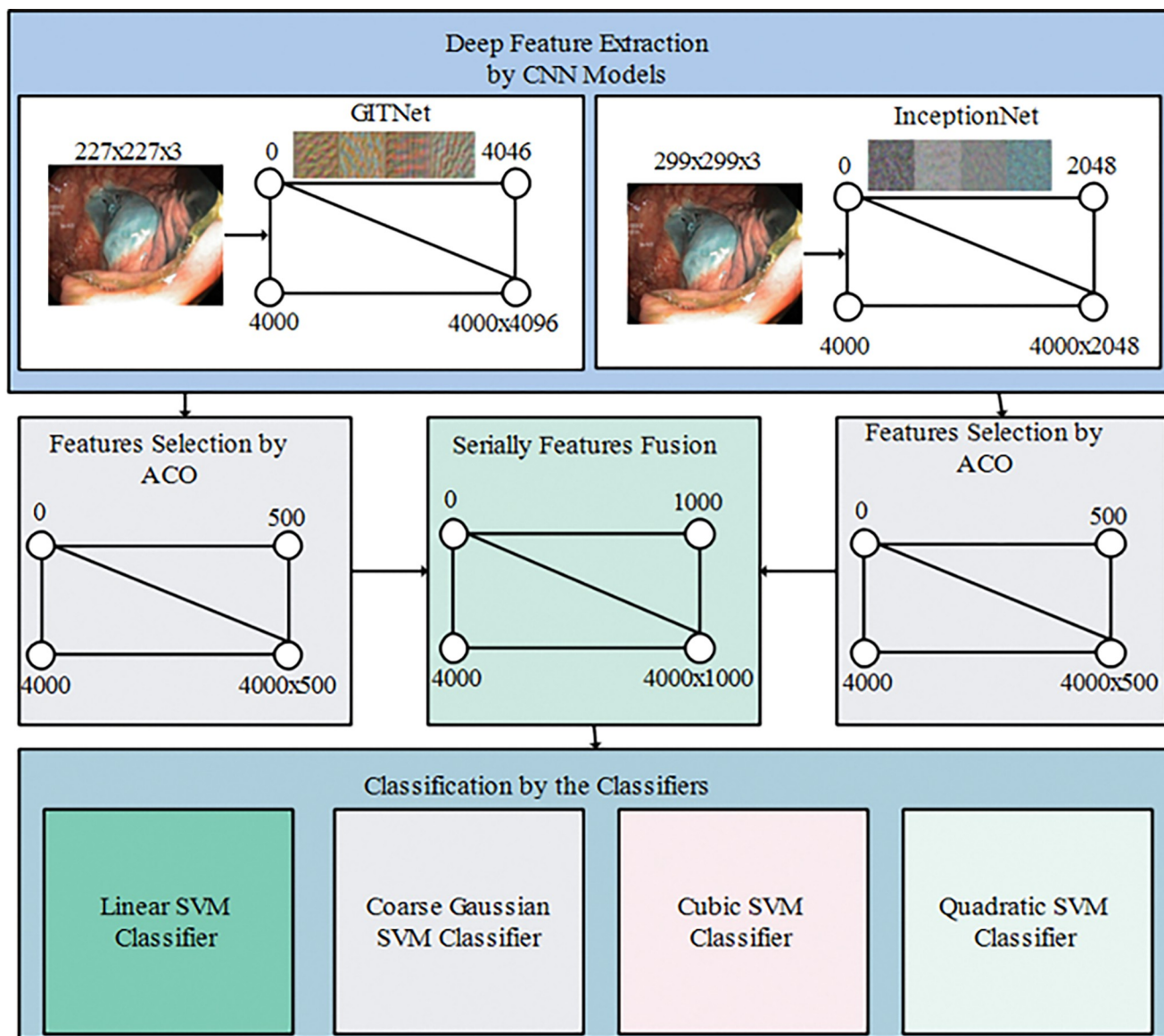


Fig 10. Proposed deep model for GIT disease classification using KVASIR dataset.

<https://doi.org/10.1371/journal.pone.0292601.g010>

3.4 Classification

The classifiers are trained on the optimized feature set and classify the diseases of GIT. The four classifiers that are considered in this research work are CSVM, CGSVM, QSVM, and LSVM. The multiple-dimensional space support vectors are generated by the SVM which classify the disease in the frames of endoscopy using 5-fold cross-validation. The best performance

Table 2. The detail of the KVASIR and NERTHUS datasets.

Ref/Dataset/year	Origin	Image modalities	No of frames	No of Classes	Frames per class	File Access	File type
[61] KVASIR (2017)	Simula Research Laboratory Norway	Endoscopy	4000	8	500	Free	jpg
[63] NERTHUS (2017)	Simula Research Laboratory Norway	Colonoscopy	5525	4	500, 2700, 975,1350	Free	jpg

<https://doi.org/10.1371/journal.pone.0292601.t002>

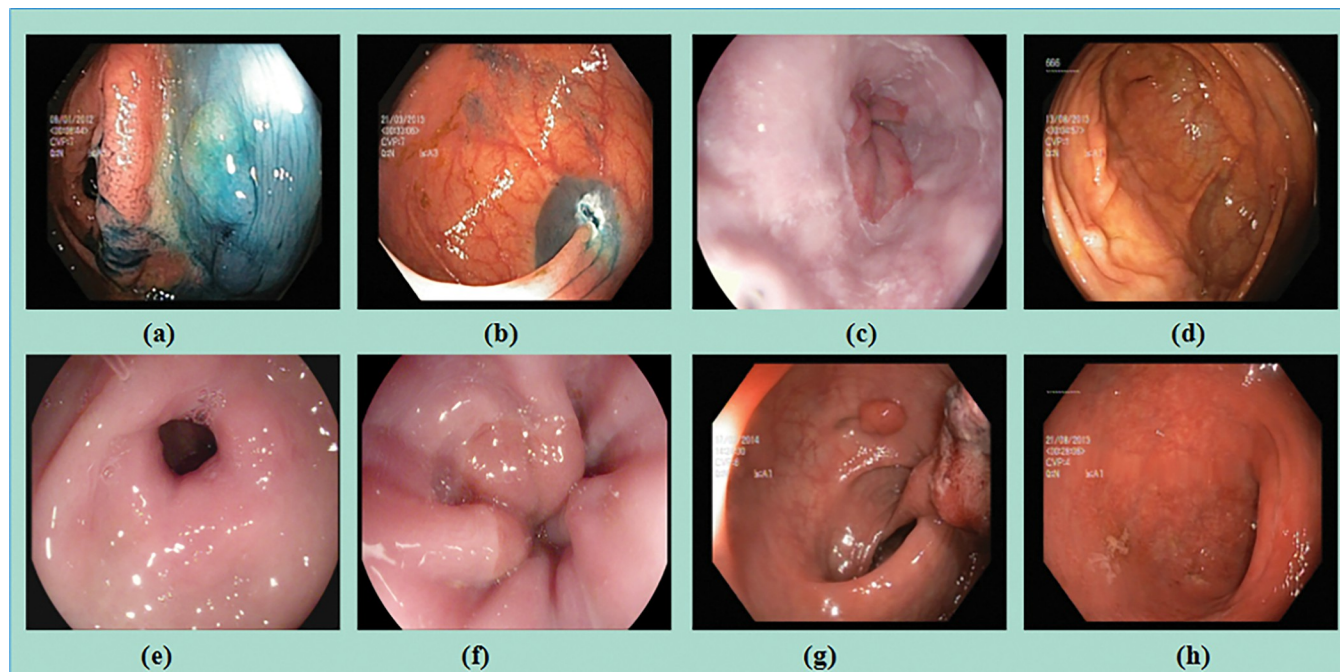


Fig 11. The samples of the KVASIR dataset with 8 classes (a) Dyed-Lifted-Polyp (b) Dyed-Resection-Margins (c) Esophagitis (d) Normal-Cecum (e) Normal-Pylorus (f) Normal-z-Line (g) Polyps (h) Ulcerative-Colitis.

<https://doi.org/10.1371/journal.pone.0292601.g011>

is acquired by the QSVM in the form of 99.32% accuracy using KVASIR dataset. The parameters of all classifiers are arranged such that the kernel is adjusted automatically, where box constraint level 1 is adjusted with the true value of the standardized data, and the multi-class method of each classifier is set to one-vs-one.

4. Results and discussion

The best three experiments are illustrated in this manuscript from several experiments. A detailed description of the KVASIR and NERTHUS datasets is given in section 4.1. The different methods of SVM are used for classification, where QSVM and CSVM provide better results as compared with other SVM classifiers. The classifiers are trained and tested by using 5-fold cross-validation. The experiments are carried out using the model NVIDIA GTX 1070 GPU. The Core i5 machine has 8GB of inbuilt RAM with the Windows 10 platform. MATLAB 2021a is exercised for overall experimentation. This section describes the detailed description of the datasets and performance evaluation method with numerical results and visualizations.

4.1 Datasets

In this study, KVASIR and NERTHUS datasets are employed for assessment. The KVASIR dataset consists of 4000 images comprising 500 images of each class [61]. The labeling of these images was carried out by an expert endoscopist. Each labeled class consists of 500 images, which makes for a balanced dataset. The deep models required a large dataset for better feature extraction. The data augmentation approach is carried out in the proposed work. The classes that are part of the KVASIR dataset include "Z-line", which is the area where the esophagus transits to the stomach; "Pylorus," which refers to the site of the stomach-duodenum entry; and the proximal segment of the large intestine is known as "Cecum". These three classes

belong to the anatomical landmarks section. Accurate classification of these classes provides substantial help in materializing efficient navigation inside the GIT. Three more classes labeled "Esophagitis" (an abnormal condition of the esophagus), "Polyps" (lesions affecting the bowel), and "Ulcerative Colitis" (inflammatory conditions of the large bowel) consist of images related to pathological findings. Besides these six classes, two more classes belong to Endoscopic Mucosal Resection (EMR) related conditions. These are labeled as "dyed and lifted polyps" (lesions before removal and injected with blue-colored saline) and the "dyed-resection-margins." The resolution of the frames varies from 720 x 576 pixels to 1920 x 1072 pixels in the KVASIR dataset. NERTHUS dataset is produced by colonoscopy (endoscopic examination of the bowel). NERTHUS comprises four classes with 5525 frames of bowel from 21 videos [62]. The detailed information on the KVASIR and NERTHUS datasets is specified in Table 2. The sample of the KVASIR dataset comprises eight classes, as illustrated in Fig 11.

4.2 Performance evaluation protocols

This study deals with a multiclass classification problem where the models are assessed by various standard evaluation metrics. The accuracy quantifies the correct prediction over all samples. The precision determines the correct positive identified findings. Sensitivity or recall returns the information of correctly identified real positive numbers. Similarly, the F1 is correlated between Precision and recall. Geometric means indicate the central tendency. The mathematical expression of each evaluation protocol is given below.

$$Accuracy = \frac{(T_{positive} + T_{negative})}{(T_{positive} + T_{negative}) + (F_{positive} + F_{negative})} \quad (12)$$

$$Precision = \frac{T_{positive}}{(T_{positive} + F_{positive})} \quad (13)$$

$$Sensitivity/recall = \frac{T_{positive}}{(T_{positive} + F_{negative})} \quad (14)$$

$$F_1 score = \frac{T_{positive}}{T_{positive} + 1/2(F_{positive} + F_{negative})} \quad (15)$$

$$Geometric Mean = \left(\prod_{i=1}^n f_i \right)^{1/n} \quad (16)$$

$$Specificity = \frac{T_{negative}}{(T_{negative} + F_{positive})} \quad (17)$$

The rate of the correctly predicted positive class is depicted as true positive ($T_{positive}$). The rate of the correctly identified as the negative class is attributed as true negative ($T_{negative}$). The wrong predictions for the positive class are termed "false positives" ($F_{positive}$). The inaccurate estimation probability of the negative class is a false negative ($F_{negative}$).

4.3 Experimental setup

The convolutional networks InceptionNetV3 and GITNet are used for feature learning. The deep models are trained by using two datasets including KVASIR and NERTHUS containing 4000 and 5525 images respectively. Extracted features from deep models are optimized and

Table 3. Classifiers and feature combination with achieved accuracy over KVASIR and NERTHUS datasets.

Experiments Over KVASIR	Classifiers	Inception Features	GITNet Features	Total Features	Accuracy	Prediction Speed(obs/sec)
Experiment 1	CSVM	100	100	200	91.10%	~2300
Experiment 2	QSVM	250	250	500	98.00%	~630
Experiment 3	QSVM	500	500	1000	99.32%	~160
Experiments Over NERTHUS	Classifiers	Inception Features	GITNet Features	Total Features	Accuracy	Prediction Speed(obs/sec)
Experiment 1	CSVM	100	100	200	99.84%	~2050
Experiment 2	CSVM	250	250	500	99.82%	~780
Experiment 3	CSVM	500	500	1000	99.89%	~623

<https://doi.org/10.1371/journal.pone.0292601.t003>

fused serially. In this manuscript, several experiments are performed, but only the best three experiments' results are illustrated with 200, 500, and 1000 feature sets. Four classifiers are considered: LSVM, CGSVM, CSVM, and QSVM provide better results in terms of Acc. The results of models achieving high Acc are depicted in Table 3, such as 91.10%, 98%, and 99.32% Acc. The prediction speed in terms of observation per second (obs/sec) is provided of the best classifiers in the three experiments.

4.4 Experimental detail using KVASIR and NERTHUS datasets

The four classifiers that are considered in this research and the results of individual classifiers are illustrated in sub-sections. The CSVM, CGSVM, QSVM, and LSVM are trained by setting a 5-fold testing arrangement. The performance of the model is shown in the following sections, Accuracy (Acc), Sensitivity (Sens), Specificity (Spec), Precision (Prec), F1 score (F1.S), and Geometric means (G.M) are performance measures.

4.4.1 Experiment 1: Results with 200 features on KVASIR dataset. In test experiment 1, classifiers are trained and tested by using 200 optimized features. The CSVM classifier provides

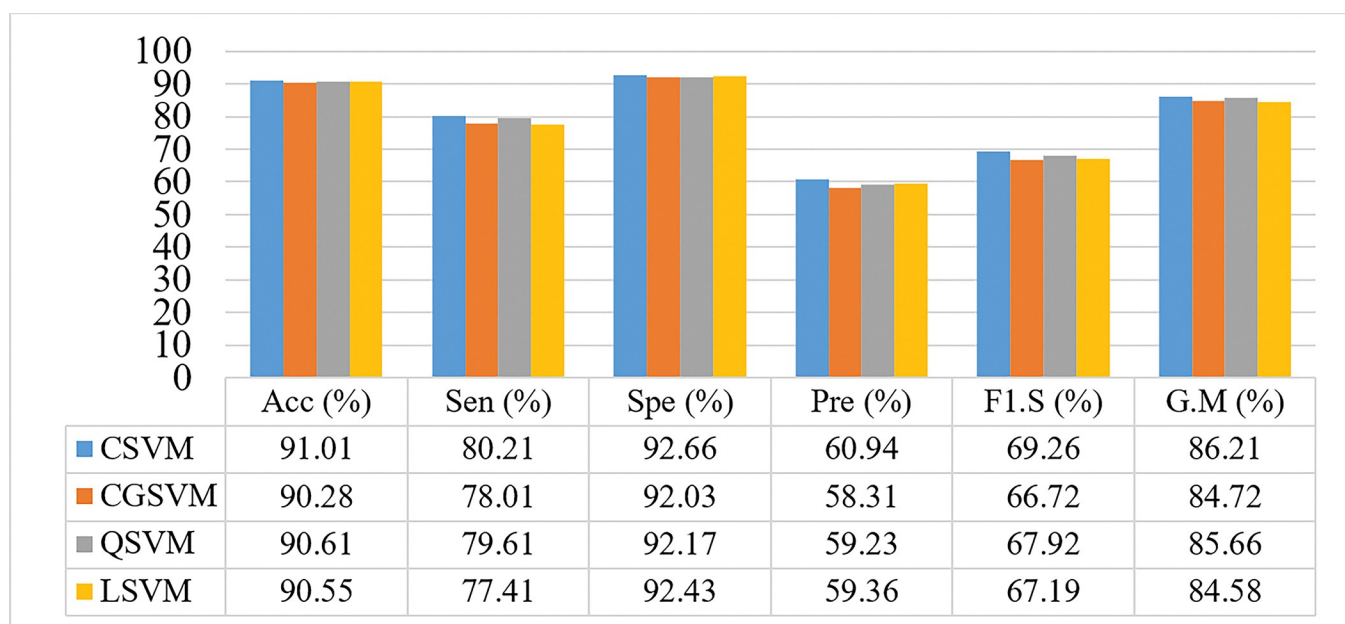


Fig 12. Evaluation of the model using 200 features.

<https://doi.org/10.1371/journal.pone.0292601.g012>

high accuracy in experiment 1 as compared with other classifiers, as shown in Fig 12. The results of the CSVM are illustrated as the Acc is measured as 91.01%, Sens is calculated as 80.21%, Spec determines the value as 92.66%, Prec is estimated as 60.94%, the F1.S is evaluated as 69.26%, and the value of the G.M is assessed 86.21% that shows the superiority of the proposed work. The results of the CGSVM are illustrated as the Acc is found 90.28%, the value of the Sens is determined as 78.01%, Spec is measured as 92.03%, Prec is calculated as 58.31%, F1.S is measured as 66.72%, and G.M is founded as 84.72%. In the same way, the results of the QSVM are illustrated as the Acc is determined as 90.61%, the value of Sens is as 79.61%, Spec is specified as 92.17%, the score of the Prec is as 59.23%, F1.S is as 67.92%, and G.M is evaluated as 85.66%. The results of the LSVM are illustrated as the Acc is measured as 90.55%, Sens value is as 77.41%, Spec determines as 92.43%, Prec number is as 59.36%, F1.S is delivered as 67.19%, and G.M informs about the value is 84.58%.

4.4.2 Experiment 2: Results with 500 features on KVASIR dataset. In test experiment 2, classifiers are trained and tested by using a 500-optimized feature set. The QSVM classifier provides high accuracy in experiment 2 as compared with other classifiers as shown in Fig 13. The results of the CSVM are illustrated as the Acc is 97.32%, Sens shows the value as 96.41%, Spec is measured as 97.46%, Prec is calculated as 84.41%, F1.S is achieved as 90.01, and G.M is attained as 96.93% that shows the progress of the proposed model over the other classifiers in experiment 2. The results of the CGSVM are illustrated as the Acc is determined as 93.03%, Sens is founded as 85.81%, Spec is measured as 94.06%, Prec is obtained as 67.35%, F1.S is measured as 75.46%, and G.M is acquired as 89.83%. The results of the QSVM are illustrated as the Acc is attained as 98.01%, Sens is taught as 98.21%, Spec is determined as 97.97%, Prec is achieved as 87.37%, F1.S is calculated as 92.47%, and G.M is obtained as 98.09%. The results of the LSVM are illustrated as the Acc is measured as 97.15%, Sens number is shown as 97.21%, Spec is calculated as 97.14%, Prec is taught as 82.94%, F1.S is quantified as 89.51%, and G.M achieved as 97.17%.

4.4.3 Experiment 3: Results with 1000 features on KVASIR dataset

In test experiment 3, classifiers are trained and tested by using a 1000-optimized feature set. The QSVM classifier provides high accuracy in experiment 3 as compared with other

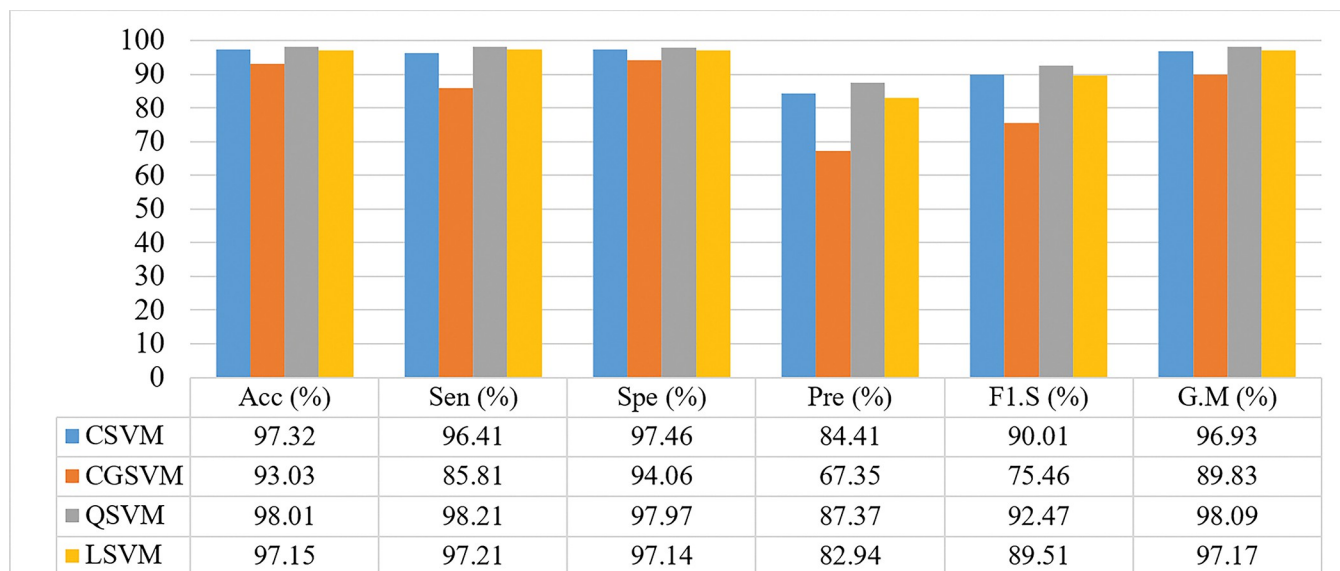


Fig 13. Evaluation of the model using 500 features.

<https://doi.org/10.1371/journal.pone.0292601.g013>

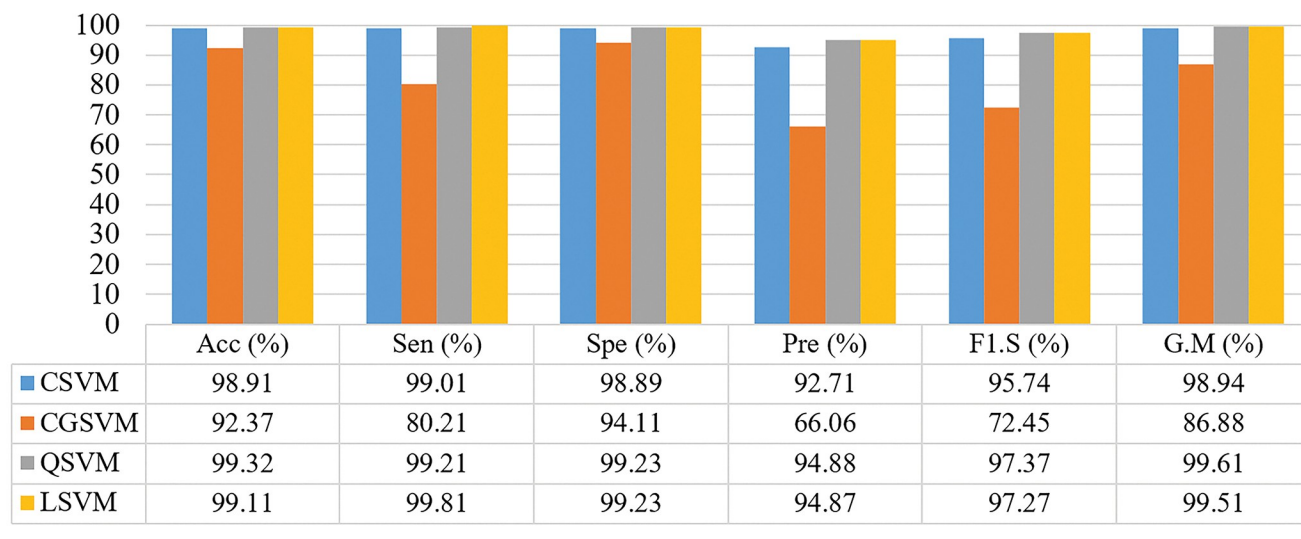


Fig 14. Evaluation of the model using 1000 features.

<https://doi.org/10.1371/journal.pone.0292601.g014>

classifiers as shown in Fig 14. The results of the CSVM are illustrated as the Acc is founded as 98.91%, Sens is measured as 99.01%, the figures of the Spec are as 98.89%, Prec is determined as 92.71%, F1.S calculated as 95.74%, and G.M is assessed as 98.94% that shows the better results of the classifier over the other classifiers. The results of the CGSVM are illustrated as the Acc is determined as 92.37%, Sens value is as 80.21%, Spec improved as 94.11%, Prec shows the value as 66.06%, F1.S is measured as 72.45%, and G.M reflects the value 86.88%. The results of the QSVM are illustrated as the Acc shows the number as 99.32%, Sens is calculated as 99.21%, the assessment of the Spec is determined as 99.23%, Prec is figured-out as 94.88%, F1.S is as 97.37%, and G.M values comes as 99.61%. The results of the LSVM are illustrated as the value of the Acc is acquired as 99.11%, Sens is achieved as (99.81%), Spec is determined as (99.23%), Prec is attained as 94.87%, F1.S is as 97.27%, and G.M is obtained as 99.51%.

Table 4 presents the confusion matrix obtained by QSVM with the best outcomes at 1000 features. The classes of diseases are illustrated in short form. Dyed-Lifted-Polyps (DLP), Dyed-Resection-Margins (DRM), Esophagitis (Eso), Normal-Cecum (NC), Normal-Pylorus (NP), Normal-Z-Line (NZL), Polyps (Pol), and Ulcerative-Colitis (UC). The AUC of the three experiments is illustrated in Fig 15.

Table 4. Confusion matrix of QSVM with best outcomes with 1000 features.

Actual Class	Predicted Class								
	A/P	DLP	DRM	Eso	NC	NP	NZL	Pol	UC
DLP		500	0	0	0	0	0	0	0
DRM		0	499	0	1	0	0	0	0
Eso		0	0	497	0	2	1	0	0
NC		0	0	0	498	0	0	2	0
NP		0	0	0	0	498	2	0	0
NZL		0	0	0	0	3	496	1	0
Pol		0	0	0	1	1	0	493	5
UC		0	0	0	0	0	0	8	492

<https://doi.org/10.1371/journal.pone.0292601.t004>

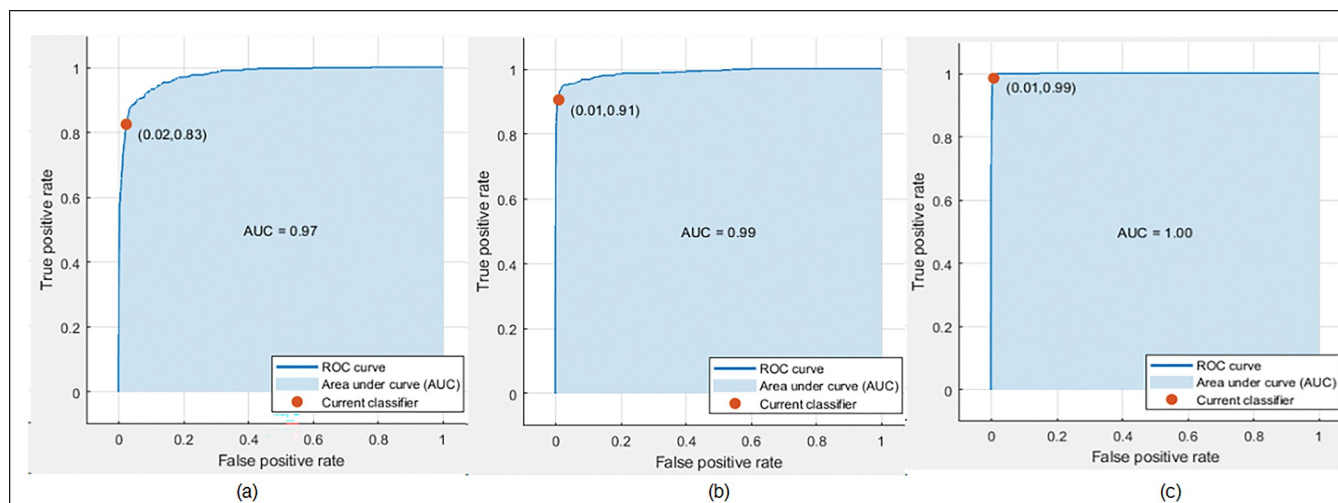


Fig 15. AUC of three experiments (a) CSVM in experiment 1 (b) QSVM in experiment 2 (c) QSVM in experiment 3.

<https://doi.org/10.1371/journal.pone.0292601.g015>

The graph in Fig 16 shows the training time of the best classifiers of the proposed model.

4.4.4 Results of Experiment1 (200), Experiment2 (500) and Experiment3 (1000) features using NERTHUS dataset. The intended model is assessed on the NERTHUS dataset which contains four classes. The results are taken using different features including 200, 500, and 1000 that are fused features of deep CNN models including InceptionNetV3 and GITNet. The outcomes are depicted in Table 5 in terms of Acc%, Sen%, Spe%, Pre%, F1.S%, and G.M%. In experiment 1, four classifiers are trained where the CSVM classifier performs well over other classifiers in terms of Acc(99.84%). Similarly, in experiment2 and 3, CSVM computes better over other classifiers in terms of Acc (99.82%) and Acc(99.89%) respectively. The results on the NERTHUS dataset show the SOTA performance of the proposed approach.

The performance of the classifiers is analyzed using a boxplot as shown in Fig 17. The analysis tests are conducted using 100, 200, 500, and 1000 features for each group of the classifiers. The variance of analysis is conducted over the two datasets. The tests of different classifiers using features of the KVASIR dataset consisted of adjacent values as CSVM lower (90.55), upper(98.91) and mean(94.17), CGSVM lower(90.28), upper(93.03) and mean(91.49), QSVM

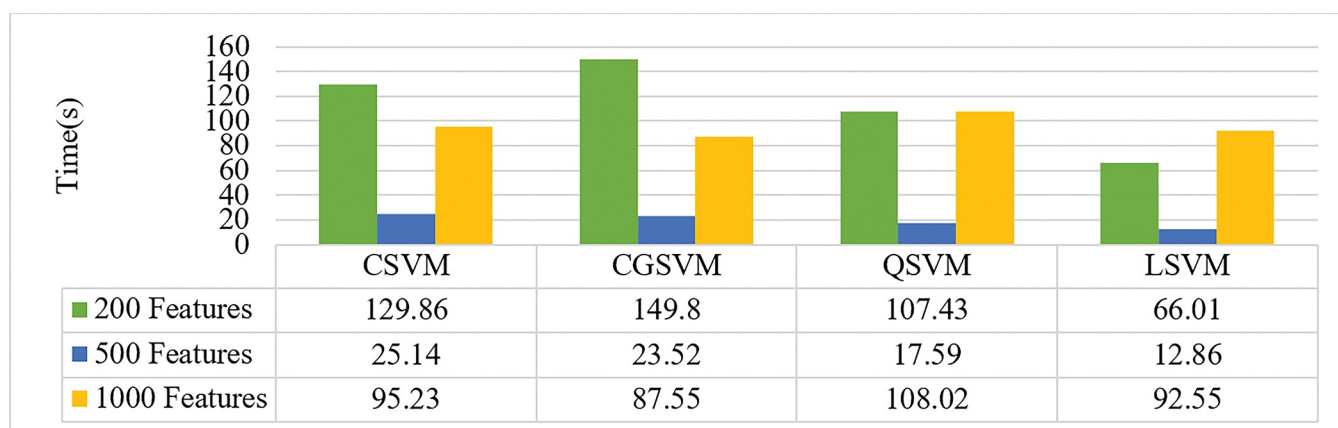


Fig 16. Classifier training time.

<https://doi.org/10.1371/journal.pone.0292601.g016>

Table 5. Evaluation of the model with 200, 500, and 1000 features using the NERTHUS dataset.

No of Features (Experiment1)	Classifiers	Acc (%)	Sen (%)	Spe (%)	Pre (%)	F1.S (%)	G.M (%)
200	CSVM	99.84	99.20	99.90	99.00	99.10	99.55
	CGSVM	94.39	83.60	95.46	64.71	72.95	89.33
	QSVM	99.76	99.20	99.82	98.22	98.71	99.51
	LSVM	98.50	97.00	98.65	87.70	92.12	97.82
No of Features (Experiment2)	Classifiers	Acc (%)	Sen (%)	Spe (%)	Pre (%)	F1.S (%)	G.M (%)
500	CSVM	99.82	98.60	99.94	99.40	99.00	99.27
	CGSVM	95.15	88.60	95.80	67.74	76.78	92.13
	QSVM	99.73	98.80	99.82	98.21	98.50	99.31
	LSVM	98.84	98.00	98.93	90.07	93.87	98.46
No of Features (Experiment3)	Classifiers	Acc (%)	Sen (%)	Spe (%)	Pre (%)	F1.S (%)	G.M (%)
1000	CSVM	99.89	99.40	99.94	99.40	99.40	99.67
	CGSVM	95.55	89.00	96.20	69.97	78.35	92.53
	QSVM	99.80	99.80	99.80	98.04	98.91	99.80
	LSVM	98.93	98.20	99.00	90.76	94.33	98.60

<https://doi.org/10.1371/journal.pone.0292601.t005>

lower(90.24), upper(99.32) and mean(94.31), LSVM values of lower(90.21), upper(99.11), and mean(93.85). The adjacent minimum and maximum values are 98.99 and 99.89, respectively, and the mean value is 99.83 for the CSVM classifier analysis. In the CGSVM analysis test, lower and upper adjacent values are 93.94 and 95.55 respectively with a mean value of 94.77. The classifier QSVM is analyzed by quantifying low and upper adjacent points 98.96 and 99.80 respectively having Acc 99.75. The LSVM test contains lower and upper adjacent values of 97.89 and 98.93 with a mean value of 98.67. The Variance analysis shows that QSVM performs well on the NERTHUS dataset.

4.4.5 Performance comparison with existing methods. The comparison of the results of existing SOTA approaches with the proposed work is performed, as shown in Table 6. The

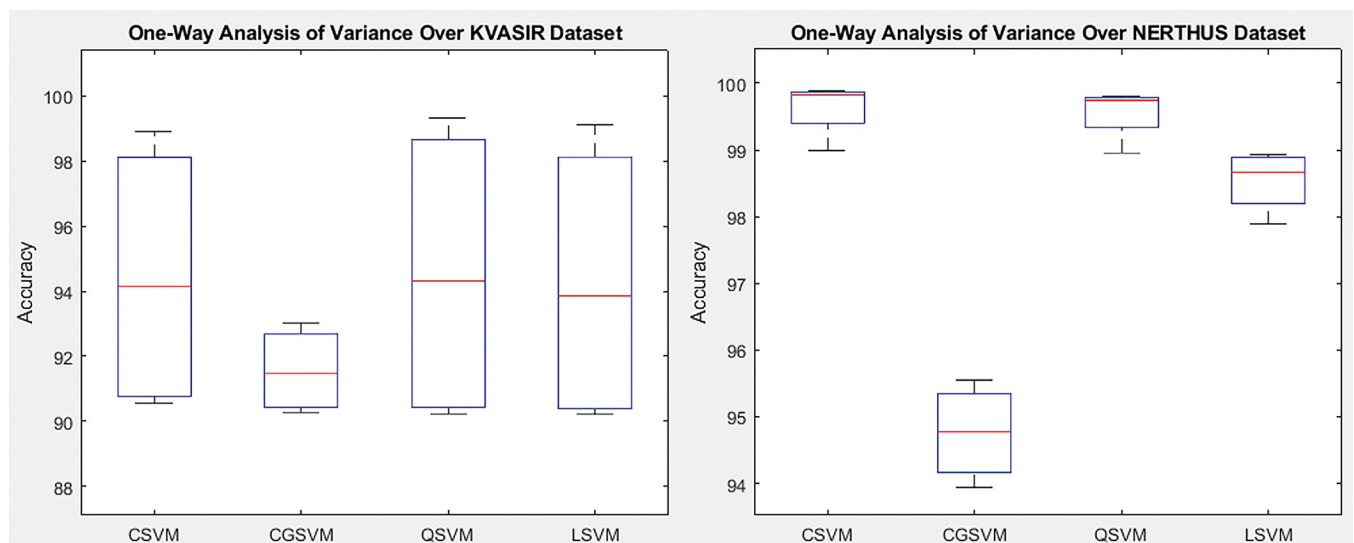


Fig 17. Analysis of Variance for evaluating classifiers' performance.

<https://doi.org/10.1371/journal.pone.0292601.g017>

Table 6. Comparison with existing approaches.

Refs/Years	Methods	Acc%
[61] 2017	GF Random Forrest	95.90
[64] 2017	Ensemble of Inception, fine-tuning, data augmentation	91.50
[65] 2019	Inception-ResNet-v2	98.48
[66] 2021	VGG+ResNet+Inception	95.02
[67] 2022	VGG16, ResNet, MobileNet, Inception-v3, and Xception	98.30
[68] 2022	VGG, InceptionNetV3	98.00
[69] 2022	ResNet-50	95.10
Proposed	InceptionNetV3, GITNet	99.32

<https://doi.org/10.1371/journal.pone.0292601.t006>

global features (GF) random forest approach achieved 95.9% Acc using the KVASIR dataset. The features were integrated into InceptionNetV3 by applying the data augmentation technique and the result was 91.5% Acc. The technique of feature fusion was accomplished in 2021 with achieving 95.02% Acc. The five deep learning models were used for feature extraction and the technique of feature fusion was applied to achieve 98.3% Acc. The features of two deep learning models were combined with achieving 98% Acc. The resNet-50 was used for feature extraction that attains 95.1% Acc.

The analysis of prior work indicates that existing work employed already crafted, pre-trained transfer learning with feature fusion methods that provide less Acc as compared with our proposed work. In our suggested work, pre-trained model inceptions are used with the proposed GITNet model for feature extraction that achieved 99.32% Acc which is better than existing methods.

4.5. Discussion

In this manuscript, GIT disease classification is performed. The preprocessing technique, AGCWD performs the transformation of the pixel energy level of the endoscopic frames. Features are extracted by the deep CNN models, including GITNet and InceptionNetV3. After that, the features of both deep models are optimized by the ACO method, which lowers the computation cost of the classifiers. Additionally, features obtained by the ACO are fused serially. The four classifiers are used for the GIT disease classification including CSVM, CGSVM, QSVM, and LSVM. Correspondingly, in the proposed work, multiple experiments are carried out but only the best three are illustrated after visual analysis using KVASIR and NERTHUS datasets. The experiments are carried out by tuning feature values, but the best results are obtained using total features such as 200, 500, and 1000 features. In experiment 1 using the KVASIR dataset, with 200 features, the CSVM provides higher accuracy than other classifiers. Similarly, the QSVM classifier provides improved accuracy in experiments 2 and 3, which are 98.00% and 99.32%, respectively. The training time of conducted experiment over the KVASIR dataset using four classifiers is depicted in Fig 16. Grad-CAM is implemented, which represents the learning pattern of the deep learning model. The proposed model is also evaluated using NERTHUS dataset where it performs well. The results of conducted three experiments over the NERTHUS dataset are depicted in Table 5. The results' comparison over the existing SOTA approaches is done which shows that the performance of the suggested work is better than the existing methodologies.

5. Conclusion and future work

This research work presents an approach for the disease prediction of GIT using endoscopic images. The work is evaluated over two publically datasets including the KVASIR and

NERTHUS datasets. The technique AGCWD is employed as a preprocessing phase for enhancing the pixel intensity level, which helps to get better features for disease prediction. The proposed GITNet is a CNN-based deep learning model which is trained and tested by using a third-party dataset i.e. CIFAR [21]. The GITNet and an existing pre-trained model named InceptionNetV3 are employed for feature extraction from the fully connected layers. A bio-inspired approach name ACO is applied for feature subset selection. The selected features are fused serially and given to the SVM-based classifiers for predictions. The selected classifiers are trained on the different feature sets of 200, 500, and 1000 and higher accuracy is achieved on 1000 features. The four classifiers considered in this research include LSVM, CSVM, QSVM, and CGSVM which are assessed during experimentation. This SOTA approach provides a classification accuracy of 99.32% on the KVASIR dataset by using QSVM. The proposed approach performed well in terms of Acc (99.89%) when evaluated over the NERTHUS dataset. Moreover, the comparison of the result is done with existing SOTA approaches. From the specified results, it can be concluded that the CNN-based deep learning models including the proposed GITNet and pre-trained InceptionNetV3 with the feature fusion method give outstanding performance for the classification of GIT images.

In future work, a different combination of hand-crafted and deep feature extraction methods can be evaluated. Different metaheuristic feature selection methods may yield better results. The improved techniques for image preprocessing and enhancement may be integrated into the existing model to improve its performance.

Author Contributions

Conceptualization: Muhammad Ramzan.

Data curation: Muhammad Irfan Sharif, Faisal Azam.

Investigation: Jungeun Kim.

Methodology: Mudassar Raza, Seifedine Kadry.

References

1. Robert M. E., Crowe S. E., Burgart L., Yantiss R. K., Lebowitz B., Greenson J. K., et al., "Statement on best practices in the use of pathology as a diagnostic tool for celiac disease," *The American journal of surgical pathology*, vol. 42, pp. e44–e58, 2018.
2. Arnold M., Abnet C. C., Neale R. E., Vignat J., Giovannucci E. L., McGlynn K. A., et al., "Global burden of 5 major types of gastrointestinal cancer," *Gastroenterology*, vol. 159, pp. 335–349. e15, 2020. <https://doi.org/10.1053/j.gastro.2020.02.068> PMID: 32247694
3. Li C., Lin L., Zhang L., Xu R., Chen X., Ji J., et al., "Long noncoding RNA p21 enhances autophagy to alleviate endothelial progenitor cells damage and promote endothelial repair in hypertension through SESN2/AMPK/TSC2 pathway," *Pharmacological Research*, vol. 173, p. 105920, 2021. <https://doi.org/10.1016/j.phrs.2021.105920> PMID: 34601081
4. Zhu Y., Huang R., Wu Z., Song S., Cheng L., and Zhu R., "Deep learning-based predictive identification of neural stem cell differentiation," *Nature communications*, vol. 12, p. 2614, 2021. <https://doi.org/10.1038/s41467-021-22758-0> PMID: 33972525
5. Lu L., Dong J., Liu Y., Qian Y., Zhang G., Zhou W., et al., "New insights into natural products that target the gut microbiota: Effects on the prevention and treatment of colorectal cancer," *Frontiers in Pharmacology*, vol. 13, p. 964793, 2022. <https://doi.org/10.3389/fphar.2022.964793> PMID: 36046819
6. WHO, "Cancer Diseases " 3 March 2021 2021.
7. W. H. Organization, "Cancer cases over the world," 2022.
8. Qin K., Li J., Fang Y., Xu Y., Wu J., Zhang H., et al., "Convolution neural network for the diagnosis of wireless capsule endoscopy: a systematic review and meta-analysis," *Surgical Endoscopy*, pp. 1–16, 2021. <https://doi.org/10.1007/s00464-021-08689-3> PMID: 34426876

9. Lu S., Yang B., Xiao Y., Liu S., Liu M., Yin L., et al., "Iterative reconstruction of low-dose CT based on differential sparse," *Biomedical Signal Processing and Control*, vol. 79, p. 104204, 2023.
10. Thakur N., Yoon H., and Chong Y., "Current trends of artificial intelligence for colorectal cancer pathology image analysis: a systematic review," *Cancers*, vol. 12, p. 1884, 2020. <https://doi.org/10.3390/cancers12071884> PMID: 32668721
11. Ali H., Sharif M., Yasmin M., Rehmani M. H., and Riaz F., "A survey of feature extraction and fusion of deep learning for detection of abnormalities in video endoscopy of gastrointestinal-tract," *Artificial Intelligence Review*, vol. 53, pp. 2635–2707, 2020.
12. Ali S., Ghatwary N., Braden B., Lamarque D., Bailey A., Realdon S., et al., "Endoscopy disease detection challenge 2020," *arXiv preprint arXiv:2003.03376*, 2020.
13. Kumar A., Singh S. B., Satapathy S. C., and Rout M., "MOSQUITO-NET: A deep learning based CADx system for malaria diagnosis along with model interpretation using GradCam and class activation maps," *Expert Systems*, p. e12695, 2021.
14. Zhang Z., Wang L., Zheng W., Yin L., Hu R., and Yang B., "Endoscope image mosaic based on pyramid ORB," *Biomedical Signal Processing and Control*, vol. 71, p. 103261, 2022.
15. Tanaka S., Kashida H., Saito Y., Yahagi N., Yamano H., Saito S., et al., "JGES guidelines for colorectal endoscopic submucosal dissection/endoscopic mucosal resection," *Digestive Endoscopy*, vol. 27, pp. 417–434, 2015. <https://doi.org/10.1111/den.12456> PMID: 25652022
16. Sharif M., Attique Khan M., Rashid M., Yasmin M., Afza F., and Tanik U. J., "Deep CNN and geometric features-based gastrointestinal tract diseases detection and classification from wireless capsule endoscopy images," *Journal of Experimental & Theoretical Artificial Intelligence*, pp. 1–23, 2019.
17. Zhuang Y., Chen S., Jiang N., and Hu H., "An Effective WSSENet-Based Similarity Retrieval Method of Large Lung CT Image Databases," *KSII Transactions on Internet & Information Systems*, vol. 16, 2022.
18. Zhuang Y., Jiang N., and Xu Y., "Progressive distributed and parallel similarity retrieval of large CT image sequences in mobile telemedicine networks," *Wireless Communications and Mobile Computing*, vol. 2022, pp. 1–13, 2022.
19. Khan M. A., Sarfraz M. S., Alhaisoni M., Albeshier A. A., Wang S., and Ashraf I., "StomachNet: Optimal deep learning features fusion for stomach abnormalities classification," *IEEE Access*, vol. 8, pp. 197969–197981, 2020.
20. Khan M. A., Zhang Y.-D., Sharif M., and Akram T., "Pixels to classes: intelligent learning framework for multiclass skin lesion localization and classification," *Computers & Electrical Engineering*, vol. 90, p. 106956, 2021.
21. Recht B., Roelofs R., Schmidt L., and Shankar V., "Do cifar-10 classifiers generalize to cifar-10?," *arXiv preprint arXiv:1806.00451*, 2018.
22. Wu Y., Zheng Y., Wang X., Tang P., Guo W., Ma H., et al., "Ginseng-Containing Sijunzi Decoction Ameliorates Ulcerative Colitis by Orchestrating Gut Homeostasis in Microbial Modulation and Intestinal Barrier Integrity," *The American Journal of Chinese Medicine*, vol. 51, pp. 677–699, 2023. <https://doi.org/10.1142/S0192415X23500325> PMID: 36883990
23. Zheng Y., Zhang Z., Tang P., Wu Y., Zhang A., Li D., et al., "Probiotics fortify intestinal barrier function: a systematic review and meta-analysis of randomized trials," *Frontiers in Immunology*, vol. 14, p. 1143548, 2023. <https://doi.org/10.3389/fimmu.2023.1143548> PMID: 37168869
24. Vachutka J., Trneckova M., Salzman R., Kolarova H., and Belakova P., "Optimal Light Source Intensity Setting in Endoscopic Ear Surgery," *Otology & Neurotology*, vol. 43, pp. e205–e211, 2022. <https://doi.org/10.1097/MAO.0000000000003410> PMID: 34855680
25. Sharma P., Hans P., and Gupta S. C., "Classification Of Plant Leaf Diseases Using Machine Learning And Image Preprocessing Techniques," in *2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence)*, 2020, pp. 480–484.
26. Maghsoudi O. H., "Superpixel based segmentation and classification of polyps in wireless capsule endoscopy," in *2017 IEEE Signal Processing in Medicine and Biology Symposium (SPMB)*, 2017, pp. 1–4.
27. Thangaraj R., Anandamurugan S., Pandiyan P., and Kaliappan V. K., "Artificial intelligence in tomato leaf disease detection: a comprehensive review and discussion," *Journal of Plant Diseases and Protection*, pp. 1–20, 2021.
28. Muruganantham P. and Balakrishnan S. M., "A survey on deep learning models for wireless capsule endoscopy image analysis," *International Journal of Cognitive Computing in Engineering*, vol. 2, pp. 83–92, 2021.
29. Naz J., Sharif M., Yasmin M., Raza M., and Khan M. A., "Detection and classification of gastrointestinal diseases using machine learning," *Current Medical Imaging*, vol. 17, pp. 479–490, 2021. <https://doi.org/10.2174/1573405616666200928144626> PMID: 32988355

30. Goel N., Kaur S., Gunjan D., and Mahapatra S., "Investigating the significance of color space for abnormality detection in wireless capsule endoscopy images," *Biomedical Signal Processing and Control*, vol. 75, p. 103624, 2022.
31. Bouyaya D., Benierbah S., and Khamadja M., "An intelligent compression system for wireless capsule endoscopy images," *Biomedical Signal Processing and Control*, vol. 70, p. 102929, 2021.
32. Eze P., Udaya P., Evans R., and Liu D., "Comparing Yiq and Ycbr Colour Image Transforms for Semi-Fragile Medical Image Steganography," 2019.
33. Tuba E., Tuba M., and Jovanovic R., "An algorithm for automated segmentation for bleeding detection in endoscopic images," in *2017 International Joint Conference on Neural Networks (IJCNN)*, 2017, pp. 4579–4586.
34. Xu Y., Zhang F., Zhai W., Cheng S., Li J., and Wang Y., "Unraveling of advances in 3D-printed polymer-based bone scaffolds," *Polymers*, vol. 14, p. 566, 2022. <https://doi.org/10.3390/polym14030566> PMID: 35160556
35. Liu M., Zhang X., Yang B., Yin Z., Liu S., Yin L., et al., "Three-dimensional modeling of heart soft tissue motion," *Applied Sciences*, vol. 13, p. 2493, 2023.
36. Li B. and Meng M. Q.-H., "Tumor recognition in wireless capsule endoscopy images using textural features and SVM-based feature selection," *IEEE Transactions on Information Technology in Biomedicine*, vol. 16, pp. 323–329, 2012. <https://doi.org/10.1109/TITB.2012.2185807> PMID: 22287246
37. Charfi S., El Ansari M., and Balasingham I., "Computer-aided diagnosis system for ulcer detection in wireless capsule endoscopy images," *IET Image Processing*, vol. 13, pp. 1023–1030, 2019.
38. Suman S., Hussin F. A. B., Malik A. S., Pogorelov K., Riegler M., Ho S. H., et al., "Detection and classification of bleeding region in WCE images using color feature," in *Proceedings of the 15th International Workshop on Content-Based Multimedia Indexing*, 2017, pp. 1–6.
39. Altini N., Marvulli T. M., Caputo M., Mattioli E., Prencipe B., Cascarano G. D., et al., "Multi-class Tissue Classification in Colorectal Cancer with Handcrafted and Deep Features," in *International Conference on Intelligent Computing*, 2021, pp. 512–525.
40. Majid A., Khan M. A., Yasmin M., Rehman A., Yousafzai A., and Tariq U., "Classification of stomach infections: A paradigm of convolutional neural network along with classical features fusion and selection," *Microscopy research and technique*, vol. 83, pp. 562–576, 2020. <https://doi.org/10.1002/jemt.23447> PMID: 31984630
41. Lu S., Liu S., Hou P., Yang B., Liu M., Yin L., et al., "Soft Tissue Feature Tracking Based on DeepMatching Network," *CMES-Computer Modeling in Engineering & Sciences*, vol. 136, 2023.
42. Krizhevsky A., Sutskever I., and Hinton G. E., "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, pp. 1097–1105, 2012.
43. Simonyan K. and Zisserman A., "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv: 1409.1556*, 2014.
44. Dung C. V., "Autonomous concrete crack detection using deep fully convolutional neural network," *Automation in Construction*, vol. 99, pp. 52–58, 2019.
45. Lee J. H., Kim Y. J., Kim Y. W., Park S., Choi Y.-i, Kim Y. J., et al., "Spotting malignancies from gastric endoscopic images using deep learning," *Surgical endoscopy*, vol. 33, pp. 3790–3797, 2019. <https://doi.org/10.1007/s00464-019-06677-2> PMID: 30719560
46. Ghosh T., Palash M. I. A., Yousuf M. A., Hamid M. A., Monowar M. M., and Alassafi M. O., "A Robust Distributed Deep Learning Approach to Detect Alzheimer's Disease from MRI Images," *Mathematics*, vol. 11, p. 2633, 2023.
47. Sharif M., Attique Khan M., Rashid M., Yasmin M., Afza F., and Tanik U. J., "Deep CNN and geometric features-based gastrointestinal tract diseases detection and classification from wireless capsule endoscopy images," *Journal of Experimental & Theoretical Artificial Intelligence*, vol. 33, pp. 577–599, 2021.
48. Ramzan M., Raza M., Sharif M. I., and Kadry S., "Gastrointestinal Tract Polyp Anomaly Segmentation on Colonoscopy Images Using Graft-U-Net," *Journal of Personalized Medicine*, vol. 12, p. 1459, 2022. <https://doi.org/10.3390/jpm12091459> PMID: 36143244
49. Suman S., Hussin F. A., Malik A. S., Ho S. H., Hilmi I., Leow A. H.-R., et al., "Feature selection and classification of ulcerated lesions using statistical analysis for WCE images," *Applied Sciences*, vol. 7, p. 1097, 2017.
50. Hossain M. M., Hasan M. M., Rahim M. A., Rahman M. M., Yousuf M. A., Al-Ashhab S., et al., "Particle swarm optimized fuzzy CNN with quantitative feature fusion for ultrasound image quality identification," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 10, pp. 1–12, 2022. <https://doi.org/10.1109/JTEHM.2022.3197923> PMID: 36226132
51. Jain S. and Salau A. O., "An image feature selection approach for dimensionality reduction based on kNN and SVM for AkT proteins," *Cogent Engineering*, vol. 6, p. 1599537, 2019.

52. Mohammad F. and Al-Razgan M., "Deep Feature Fusion and Optimization-Based Approach for Stomach Disease Classification," *Sensors*, vol. 22, p. 2801, 2022. <https://doi.org/10.3390/s22072801> PMID: 35408415
53. Naz J., Sharif M., Raza M., Shah J. H., Yasmin M., Kadry S., et al., "Recognizing gastrointestinal malignancies on WCE and CCE images by an ensemble of deep and handcrafted features with entropy and PCA based features optimization," *Neural Processing Letters*, pp. 1–26, 2021.
54. Huang W., Huang Y., Wu Z., Yin J., and Chen Q., "A Multi-Kernel Mode Using a Local Binary Pattern and Random Patch Convolution for Hyperspectral Image Classification," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 4607–4620, 2021.
55. Devanathan K., Ganesan K., and Swaminathan R., "An automated classification of HEp-2 cellular shapes using Bag-of-keypoint features and Ant Colony Optimization," *Biocybernetics and Biomedical Engineering*, vol. 41, pp. 376–390, 2021.
56. Liang M., Ren Z., Yang J., Feng W., and Li B., "Identification of colon cancer using multi-scale feature fusion convolutional neural network based on shearlet transform," *IEEE Access*, vol. 8, pp. 208969–208977, 2020.
57. Huang S.-C., Cheng F.-C., and Chiu Y.-S., "Efficient contrast enhancement using adaptive gamma correction with weighting distribution," *IEEE transactions on image processing*, vol. 22, pp. 1032–1041, 2012. <https://doi.org/10.1109/TIP.2012.2226047> PMID: 23144035
58. Aisu N., Miyake M., Takeshita K., Akiyama M., Kawasaki R., Kashiwagi K., et al., "Regulatory-approved deep learning/machine learning-based medical devices in Japan as of 2020: A systematic review," *PLOS Digital Health*, vol. 1, p. e0000001, 2022. <https://doi.org/10.1371/journal.pdig.0000001> PMID: 36812514
59. Szegedy C., Vanhoucke V., Ioffe S., Shlens J., and Wojna Z., "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2818–2826.
60. Dong N., Zhao L., Wu C.-H., and Chang J.-F., "Inception v3 based cervical cell classification combined with artificially extracted features," *Applied Soft Computing*, vol. 93, p. 106311, 2020.
61. Pogorelov K., Randel K. R., Griwodz C., Eskeland S. L., de Lange T., Johansen D., et al., "Kvasir: A multi-class image dataset for computer aided gastrointestinal disease detection," in *Proceedings of the 8th ACM on Multimedia Systems Conference*, 2017, pp. 164–169.
62. Gammulle H., Denman S., Sridharan S., and Fookes C., "Two-stream deep feature modelling for automated video endoscopy data analysis," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2020, pp. 742–751.
63. Pogorelov K., Randel K. R., de Lange T., Eskeland S. L., Griwodz C., Johansen D., et al., "Nerthus: A bowel preparation quality video dataset," in *Proceedings of the 8th ACM on Multimedia Systems Conference*, 2017, pp. 170–174.
64. Asperti A. and Mastronardo C., "The effectiveness of data augmentation for detection of gastrointestinal diseases from endoscopic images," *arXiv preprint arXiv: 1712.03689*, 2017.
65. Cogan T., Cogan M., and Tamil L., "MAPGI: Accurate identification of anatomical landmarks and diseased tissue in gastrointestinal tract using deep learning," *Computers in biology and medicine*, vol. 111, p. 103351, 2019. <https://doi.org/10.1016/j.compbiomed.2019.103351> PMID: 31325742
66. Ramzan M., Raza M., Sharif M., Khan M. A., and Nam Y., "Gastrointestinal Tract Infections Classification Using Deep Learning," *CMC-COMPUTERS MATERIALS & CONTINUA*, vol. 69, pp. 3239–3257, 2021.
67. Dheir I. M. and Abu-Naser S. S., "Classification of Anomalies in Gastrointestinal Tract Using Deep Learning," *International Journal of Academic Engineering Research (IJAER)*, vol. 6, 2022.
68. Haile M. B., Salau A. O., Enyew B., and Belay A. J., "Detection and classification of gastrointestinal disease using convolutional neural network and SVM," *Cogent Engineering*, vol. 9, p. 2084878, 2022.
69. Muruganantham P. and Balakrishnan S. M., "Attention aware deep learning model for wireless capsule endoscopy lesion classification and localization," *Journal of Medical and Biological Engineering*, vol. 42, pp. 157–168, 2022.