Computers, Materials & Continua
DOI: 10.32604/cmc.2022.030432
Article

Tech Science Press

A Two Stream Fusion Assisted Deep Learning Framework for Stomach Diseases Classification

Muhammad Shahid Amin[1], Jamal Hussain Shah[1], Mussarat Yasmin[1], Ghulam Jillani Ansari[2], Muhamamd Attique Khan[3], Usman Tariq[4], Ye Jin Kim[5] and Byoungchol Chang[6,*]

[1]Department of Computer Science, COMSATS University Islamabad, Wah Campus, Pakistan
[2]Department of Information Sciences University of Education, Lahore (Multan Campus), Pakistan
[3]Department of Computer Science, HITEC University Taxila, Pakistan
[4]College of Computer Engineering and Sciences, Prince Sattam bin Abdulaziz University, Al-Kharj, Saudi Arabia
[5]Department of Computer Science, Hanyang University, Seoul, 04763, Korea
[6]Center for Computational Social Science, Hanyang University, Seoul, 04763, Korea
*Corresponding Author: Byoungchol Chang. Email: bcchang@hanyang.ac.kr
Received: 25 March 2022; Accepted: 19 May 2022

Abstract: Due to rapid development in Artificial Intelligence (AI) and Deep Learning (DL), it is difficult to maintain the security and robustness of these techniques and algorithms due to emergence of novel term adversary sampling. Such technique is sensitive to these models. Thus, fake samples cause AI and DL model to produce diverse results. Adversarial attacks that successfully implemented in real world scenarios highlight their applicability even further. In this regard, minor modifications of input images cause "Adversarial Attacks" that altered the performance of competing attacks dramatically. Recently, such attacks and defensive strategies are gaining lot of attention by the machine learning and security researchers. Doctors use different kinds of technologies to examine the patient abnormalities including Wireless Capsule Endoscopy (WCE). However, using WCE it is very difficult for doctors to detect an abnormality within images since it takes enough time while inspection and deciding abnormality. As a result, it took weeks to generate patients test report, which is tiring and strenuous for them. Therefore, researchers come out with the solution to adopt computerized technologies, which are more suitable for the classification and detection of such abnormalities. As far as the classification is concern, the adversarial attacks generate problems in classified images. Now days, to handle this issue machine learning is mainstream defensive approach against adversarial attacks. Hence, this research exposes the attacks by altering the datasets with noise including salt and pepper and Fast Gradient Sign Method (FGSM) and then reflects that how machine learning algorithms work fine to handle these noises in order to avoid attacks. Results obtained on the WCE images which are vulnerable to adversarial attack are 96.30% accurate and prove that the proposed defensive model is robust when compared to competitive existing methods.

Keywords: WCE images; adversarial attacks; FGSM noise; salt and pepper noise; feature fusion; deep learning

1 Introduction

Over the last 20 years, advancements in the field of medical imaging have shown significant progress in computerized diagnosis of diseases in different parts of human body [1,2]. Similar to brain, stomach infections can also be detected early by using computer aided diagnosis (CAD) systems [3]. Gastrointestinal (GI) stomach infections comprise of bleeding, ulcer and polyps [4]. In United States alone, 7,65,000 people died due to stomach diseases in 2017 [5]. Since 2019, 27,570 new instances of GI stomach infections are analyzed in the United States that include 17,240 men and 10,330 women cases. Though approximated deaths are 11,130 including 6,795 men and 4,335 women [6,7]. The utilization of push gastroscopy apparatuses for identification and examination of GI diseases isn't appropriate for small bowls because of its compound architecture [8]. This issue was fixed in the year 2000 by presenting another innovation named as WCE [9]. As indicated by a yearly report of 2018, around 10,000,00 cases are effectively cured utilizing WCE.

In WCE, doctor inspects inside of GI tract to identify the infection. In this procedure, the patient gulps a capsule shape device consisting of a camera. Capsule naturally moves in GI tract and in the wake of transmitting constant video, the device expels over the anus. The doctor looks at the video received and tells about the infection [10,11]. Capsule endoscopy is utilized for detecting infection like polyps and ulcer in GI tract. This method is easy to use. During Capsule endoscopy therapies including Enteroscopy and CT-Enteroclysis, a patient experiences most of inconvenience and complexities. Detecting GI bleedings and tumor, particularly in small intestine, improves diagnostic accuracy [12,13]. The entire procedure takes more than 120 min on average. A camera captures video frames with a resolution of 255×255 pixels at a rate of 2 frames per second. On a normal, the entire method takes over 2 h. All the frames are compressed utilizing jpeg format [14].

It is very difficult for doctor to examine all the video frames. On an average, around 60,000 images of one person are analyzed manually, which almost seems impossible even for an experienced doctor. Although most of the frames are not needed to analyze but in the manual analysis, doctor checks all the video frames for results [15]. To tackle this issue, analysts have been attempting to utilize computerized techniques including steps of classification, segmentation and feature visualization [16]. Based on this, few techniques from previous work used attributes of texture and color [17]. All computed attributes are not relevant and should be omitted to achieve a good classification accuracy. As a result, feature selection methods are needed [18]. For computerized diagnosis of medical disorders, a deep neural networks-based approach was used [19]. It was first discovered that high performing deep networks are vulnerable to adversarial attacks; such attacks will change the network's classifier outputs, causing it to predict incorrect results with high confidence. The multiple networks classifier may be influenced by these image perturbations, causing it to forecast incorrect outcomes [20]. Deep networks' protection is becoming increasingly important as they become more mainstream and integrated into people's everyday lives. Deep networks also show incredible accuracy, yet these remain vulnerable to adversarial assaults [21]. Adversaries may effectively target the network by making minor changes to the input picture that are almost undetectable to the human eye but have catastrophic effects [22].

Adversarial example is relatively interesting and surprising, so it begins with the fact of training Convolutional Neural Networks (CNN) to do a very good job at recognizing what is in the image [23]. For example, Fig. 1 shows image of polyps can get correctly classified as a polyp image by Resnet101. But if the image is altered by adding some noise, that image gets classified wrongly with 99% confidence

because neural networks seem to be capturing on features and making decisions in a different way than human.

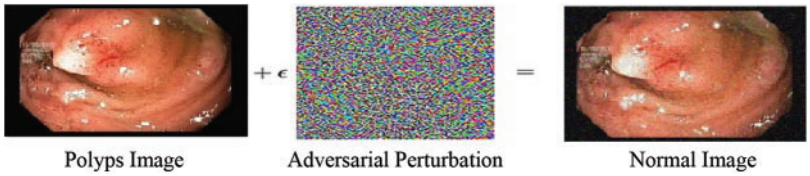

Figure 1: Polyp image before and after addition of noise

The key contributions of proposed work include:

- Assessment and analysis of adversarial attacks and defenses on WCE images expressed as advanced work.
- A model is proposed that contains FGSM and salt and pepper attacked images and a defensive technique against these attacks.
- Three different types of adversarial training are done to get best accuracy on attacked images. Notable results of WCE images are obtained by doing feature fusion of Squeeze Net, Resnet101 and handcrafted LBP features.

The remaining paper is organized as follows. Section 2 describes literature analysis of existing techniques including classification results accuracies, popular adversarial attack algorithms and defenses against these attacks. Section 3 represents the proposed work. Section 4 describes the experiments and their results while conclusion is presented in Section 5.

2 Related Work

In medical image classification, several Deep Neural Networks (DNN) have been suggested [24] and they have produced high accuracies to be used in diagnosis [25] but adversarial attacks are causing huge concern in deploying these models for clinical diagnosis [26]. Thus, critical inspection is required to assess the effectiveness of DNN models in case of adversaries, as it involves decision with high stakes dependent on the diagnosis. These adversarial attacks were first generated by Pan et al. [27] where minor perturbation has caused the network model to misclassify images. Goodfellow et al. [28] claimed that DNN models are at risk of adversarial attacks because of their linear nature, so non-linear models must be developed to make them robust against such attacks. Kurakin et al. [29] discovered that adversaries can also damage DNNs even in physical world set ups. A significant percentage of adversarial inputs are graded incorrectly although the pictures were captured from mobile phone camera. Carlini et al. [30] have developed three more powerful adversaries and encouraged researchers to consider these attacks in checking out the robustness of their models against adversaries. Papernot et al. [31] have suggested adverse saliency maps giving most impactful features to have notable effect on the performance of DNNs. Ma et al. [32] have consider adversarial attacks on medical images and yielded that adversaries are more damaging to these images because of their nature but assessment of adversaries is far easier. They achieved 98% classification accuracy on three different benchmark datasets of Diabetic Retinopathy (DR), chest and melanoma, respectively. Ren et al. [33] suggested in their study that robust model against these adversaries is yet to be made since the most effective defense adversarial training is too costly to be deployed in clinical assessments. Hirano et al. [34] observed in their findings that no focus was given on more realistic adversaries named

by them as Universal Adversarial Perturbation (UAP). They considered three different diseases of DR, pneumonia and skin cancer to check their robustness against these UAP attacks. Li et al. [35] have proposed unsupervised learning based model to catch adversaries on chest X-ray dataset. The authors claimed that their model is capable to recognize extensive variety of adversarial inputs and have also preserved the classification accuracy on both white and black box attacks. Joel et al. [36] have assessed robustness of DNN models on oncology images of (Computed Tomography) CT, mammography and Magnetic Resonance Imaging (MRI). Although adversarial attacks have reduced the classification accuracy of their model but the use of adversarial training has increased it. DL in medical imaging works better but accuracy decreased through Small Perturbation (SP) and wrong label predicated. Tabs. 1 and 2 show most popular and effective adversarial attacks and defenses respectively.

Table 1: Popular adversarial attack algorithms

Year	References	Adversarial attack
2013	[27]	L-BFGS
2015	[28]	Fast gradient sign method
2016	[37]	DeepFool
2016	[31]	Jacobian-based saliency map
2017	[30]	Carlini and Wagner
2017	[38]	One pixel

Table 2: Defenses against adversarial attacks

Year	References	Defense techniques
2014	[28]	Adversarial training
2014	[39]	Deep Contractive Networks (DCN)
2016 – 2017	[40]	Image data compression
	[41]	JPG compression
2017	[42]	Input gradient regularization

Souaidi et al. [43] proposed curvelet-based Local Binary Features (LBP) features to locate ulcer regions. Support Vector Machine (SVM) and Multilayer perceptron both are used giving an accuracy of 88% and 93.28% respectively. Also, LBP variance and discrete wavelet are utilized. Li et al. [44] found an anomaly in capsule endoscopy image automatically by computerized detection of ulcer. Szczypiński et al. [45] introduced a methodology dependent on LBP and Laplacian pyramid. In RGB and YCbCr color spaces, they found accuracies in CR and green components as 95.11% and 93.8% using SVM. Georgakopoulos et al. [46] calculated five features as LBP, color coherence vectors, curvelet transforms, color descriptors and HSV color histograms to find ulcer in capsule endoscopy images, they got an accuracy of 96% by using SVM classifier. Fan et al. [47] applied deep learning to evaluate WCE videos frame for the diagnosis of both ulcer and erosions in images. They used AlexNet CNN and attained high accuracy of 95.16% and sensitivity of 96.8% exhibiting the productivity of DL approach.

Xu et al. [48] Proposed diverse strategy where the authors used SqueezeNet and DenseNet201. Pei et al. [49] utilized Fully Convolutional Network (FCNs) concatenated with Long Short Term Memory (LSTM). Seguí et al. [50] presented a computerized decision system based on a feature learning approach in which 110,500 images were used to train CNN to achieve 96% accuracy. Wimmer et al. [51] trained a CNN by using different layers and different filters to find celiac disease. They concatenated SVM and CNN and predicted a high accuracy of 97%. In Jia et al. [52] work, they applied a CNN architecture for feature extraction and utilized these extracted features to train SVM and find GI infection in WCE videos. This work produced 91% accuracy. In Ronneberger et al. [53] research, they proposed DL with image manifold (SSAEIM) which was used to detect polyp infection in photographs with an accuracy of 99.5%and 99%. SSAEIM found bubbles, polyps and turbid images with 99%, 98%, and 99.50% accuracy. Sharif et al. [54] did work on the classification of GI tract diseases by using concatenation of deep CNN and geometric-based feature methods. In the first step, they separated infected part through a color-based approach and secondly, they used two CNN including VGG16 and VGG19 for feature extraction. Summary of literature review depicts existing techniques and their classification results accuracies as shown in Tab. 3. Similarly, same techniques and technology were used for classification Seguí et al. [50] and used intrusion detection system for robust classifier [51]. To conclude this section few more techniques including [55,56] are prominent to observe in this regard.

Table 3: Summary of existing techniques

Year	References	Techniques	Classifier	Accuracy (%)
2019	[43]	Curvelet-based LBP	SVM	88
2019	[54]	CNN and geometric based features	KNN	99.5
2016	[46]	LBP and color coherencevector	SVM	96
2016	[51]	Deep CNN	F-measure	99
2016	[52]	Deep CNN	SVM	91
2015	[53]	Deep learning with image manifold		99
2014	[45]	LBP and laplacian pyramid	SVM	93.8

3 Proposed Methodology

The proposed methodology performs WCE classification using original and perturbed images, and the accuracy is preserved by including different adversarial attacks, such as FGSM and S&P, in which a Salt & Pepper (S&P) attack is a novel attack. The proposed method consists of two parts: First part is related to adversarial training and have three folded. In the first type, a dataset is prepared that includes half of the images taken from original dataset and other half is FGSM attacked images of three classes including Normal, Polyps and Ulcer images. This combinatorial dataset is trained using deep CNN Alexnet model through transfer learning. In the second type of adversarial training, another dataset is made that has half of the images taken from original dataset and other half of images consists of salt and pepper (S&P) attacked images. Likewise in the third type, a complete dataset is prepared that is equally divided into three parts in which original images, FGSM attacked images and salt and pepper noise images are included. Second part of proposed method involves adversarial attacks on original dataset to make adversary images, pre-processing to convert all images to same size of 227×227 and

extraction of LBP features. Then hand-crafted features LBP are fused with deep features extracted from SqueezeNet and ResNet101 for getting results. Finally, performance evaluations are obtained including Accuracy, Sensitivity, Specificity, FPR, and Precision. The proposed model is shown in Fig. 2. Further details of model are given in below section.

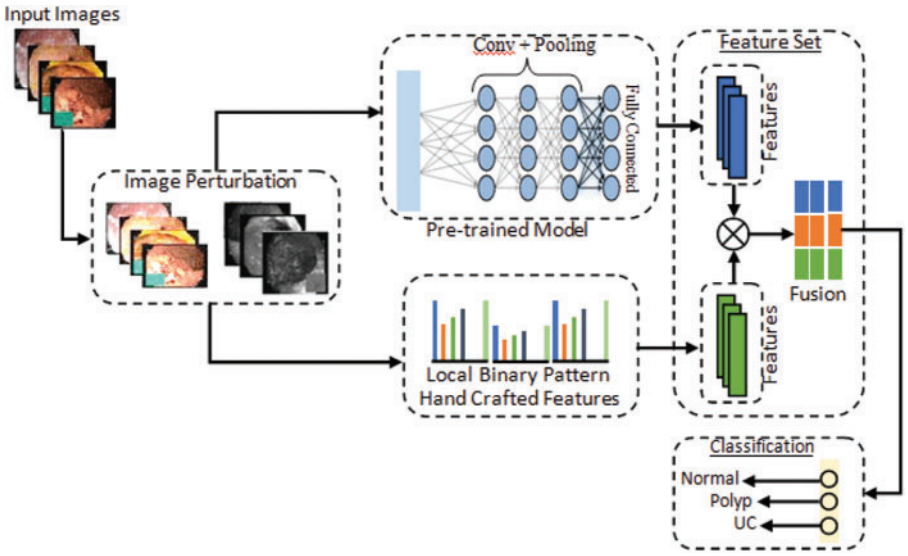

Figure 2: Proposed computerized method for stomach diseases classification

3.1 Fast Gradient Sign Method (FGSM) Attack

The quick gradient sign method uses a neural grid gradient to create an example of adversary image. For the input image, this method uses loss gradient to create a new image that maximizes losses and is named as racing image. One of the first and most popular adversarial attacks to date is called FGSM Attack developed by Goodfellow et al. [28]. The attacks are incredibly powerful, yet intuitive and specifically designed to bout neural networks. The idea is simple in that instead of minimizing losses by adjusting weights based on the multiplication of inverse gradient, the attack adjusts the input to maximize losses based on the same inverse multiplied gradient. FGSM attack model is shown in Fig. 3. The following expression can be used to summarize it:

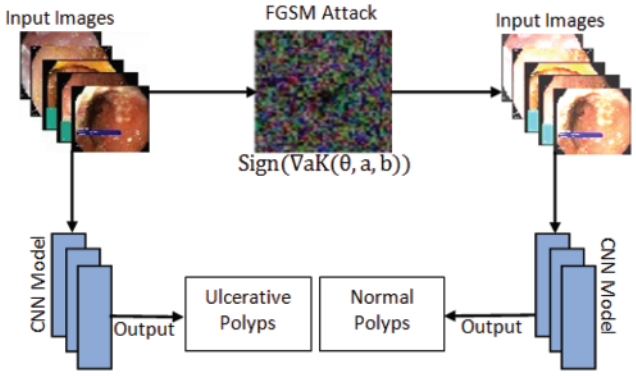

Figure 3: FGSM attack model

$$\text{Adversarial}_a = a + \in * \text{Sign}(\nabla a K\,(\theta, a, b) \quad (1)$$

where Adversarial is adversary perturbed image, a and b is original image and label respectively while € is multiplier to ensure the perturbation are small when K = Loss.

3.2 Salt and Pepper Attack

The self-generated attack known as salt & pepper is the second attack created on dataset in this work. S&P exposure is a sort of noise that occasionally appears on pictures. Impulsive noise is a term used to describe this type of noise, which could cause by abrupt and quick changes in the visual signal. It appears like a scattering of black and white pixels. Fig. 4 shows salt and pepper attack.

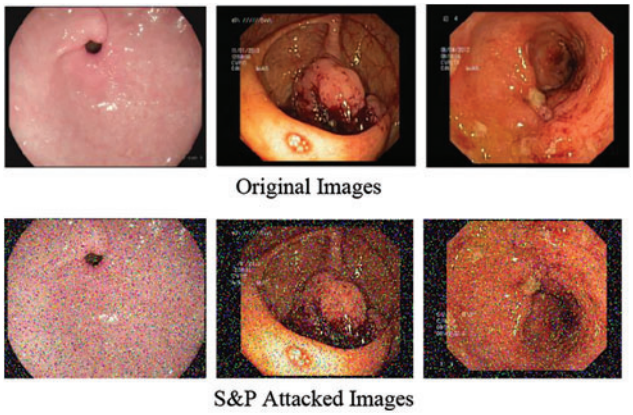

Figure 4: Salt and pepper attack

3.3 Preprocessing

In this study, Kvasir dataset includes three classes Normal pylorus, Polyps and Ulcerative Colitis. Since the dataset contains images of different sizes such as few have 1134×629 dimensions, whereas others have 814×605 dimensions. Therefore, images are resized to $227 \times 227 \times 3$ because high dimension sizes images require greater training time and hence enhance model inference time. The outcome of our preprocessing technique is shown in Fig. 5 that converts all images into grayscale.

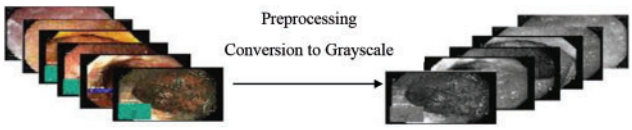

Figure 5: Preprocessing technique (Conversion to grayscale)

3.4 Deep Feature Extraction

Features are numerical values of the object representing its local and global characteristics. In our work, single feature is not useful, so we need to combine different types of features for obtaining satisfactory performance of a proposed model. The presented method uses handcrafted features (LBP) with a combination of deep features (SqueezeNet and ResNet101) for robust feature vector. The model also represents the single use of these feature extraction approaches for classification. The selection of best informative features and descriptors are problem specific and are generally founded on the

understanding of experts. Fig. 6 illustrates the general structure of deep feature extraction method and fusion.

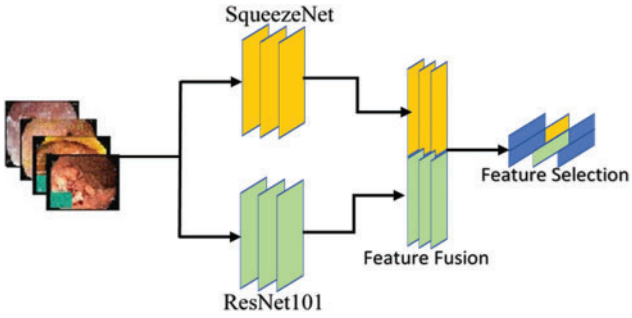

Figure 6: Deep feature extraction method

3.5 Hand-Crafted Feature Extraction

The first phase of feature extraction using handcrafted methods extracts local features of input images. In the proposed model, LBP traditional feature extraction method is applied for the extraction of robust features based on texture and shape which are important factors in describing properties of images. Fig. 7 gives illustration of hand-crafted feature extraction procedure.

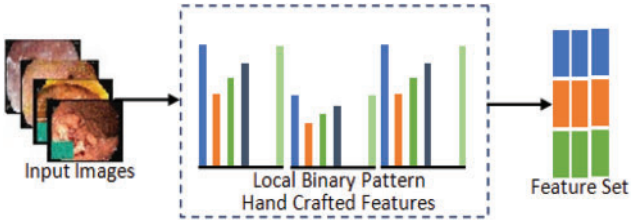

Figure 7: Hand-crafted feature extraction method

Local Binary Pattern: The LBP descriptor describes an input image through its texture spectrum. LBP is very efficient for capturing local areas of images such as boundaries, spots and smooth areas. For feature extraction, Normal pylorus, Polyps and Ulcerative-colitis images are passed to LBP which use sign part of the image to generate an 8-bit binary number using Eq. (2) that is converted into decimal followed by the computation of histogram of processed image so that the values are used as a feature vector of images.

$$\mathrm{B} \approx \left(\mathrm{TV}\left(\mathrm{C}\left(\mathrm{y}_0 - \mathrm{y}_1\right), \mathrm{C}\left(\mathrm{y}_2 - \mathrm{y}_1\right), \ldots \mathrm{C}\left(\mathrm{y}_{\mathrm{p}-2} - \mathrm{y}_1\right)\right)\right. \tag{2}$$

where B is resultant binary pixel value y_p and y_1 is the intensity of current and neighbor pixels which in this case is $p = 8$. TV is the threshold value used for generating the binary number. Normally the value of $\mathrm{TV} = 0$ if not defined specifically.

$$LBP_r = \sum_{p=0}^{A-1} s\left(\mathrm{y}_\mathrm{p} - \mathrm{y}_\mathrm{c}\right) 2^\mathrm{p} \tag{3}$$

where A is the number of neighbor point, c represents the center pixel, y_p represents the *pth* neighboring pixel and 2^p shows the histogram features that are extracted from LBP code. The features that are extracted using LBP are invariant to change in scalar and illumination because of the magnitude intensities and their relation.

3.6 Feature Fusion

In the proposed method, let there are feature vectors $F_{lbp}(C_v)$ such as LBP and deep features *SqueezeNet* (S_v) and $F_{res}(T_v)$ such as SqueezeNet and ResNet101 respectively. These features are fused using Eq. (4).

$$F_{x\times y} = \begin{pmatrix} A_v, B_{v,} C_v \\ S_v, R_v, T_v \end{pmatrix} \tag{4}$$

where fused feature vector is denoted by $F_{x\times y}$ of size 1×12850 and $y = (A + B + C + S + R + T)$ represents n gallery sample. Proposed model uses hand-crafted features with a combination of deep features to get the robust and discriminative features that increase the accuracy of classification task.

3.7 Classification

The classification in medical image analysis is one of the most demanding issues with the goal of classifying medical images into different groups to aid clinicians in disease diagnosis and study. The classification of medical images can be divided into two parts. In the first step, useful features of the images are extracted and then these characteristics are used to create models that classify image dataset in the second stage. In the presented model, classification is the main step in which dataset is classified into three different classes including Normal-pylorus, Polyps and Ulcerative-colitis. The extracted features from LBP and pre-trained CNN are concatenated through feature fusion. All the extracted and selected information is passed to different classifiers including M-SVM, weighted KNN, L-SVM, Bagged Tree, Fine KNN, Course SVM and Subspace KNN. The experiments are performed and all the classifiers are tested on the given input.

3.8 Proposed Defense Against Adversarial Attack

In this defense model against adversarial attacks, three different types of adversarial training are done on dataset. In the first adversarial training, a dataset is prepared such that half images are taken from original dataset and rest half are FGSM attacked images of three classes including Normal, Polyps and Ulcer images. Then this prepared dataset is trained using deep CNN Alexnet model through transfer learning. Model extracts the features to make prediction. After the completion of training, trained dataset is tested through perturbed and original images. Accuracy is improved by further moving to other adversarial training. In the second type of training, dataset is prepared including half of the image from original dataset and half of the images are S&P attacked images of three classes Normal, Polyps and Ulcer images. Likewise, in third training, a complete dataset is prepared that is equally divided into three parts in which original images, FGSM attacked images and S&P noise images are included. Adversarial training model is shown in Fig. 8. Adversarial training is listed as follows.

1. Adversarial Training One (ATO): Original Images + FGSM attacked Images
2. Adversarial Training Two (ATT): Original Images + S&P attacked Images
3. Mixed Adversarial Training (MAT): Original Images + FGSM + S&P

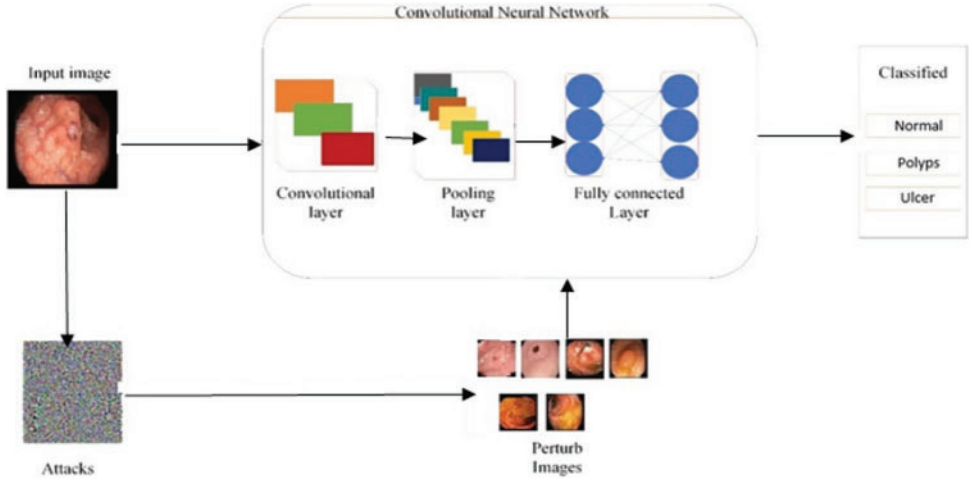

Figure 8: Adversarial training process model

4 Results and Analysis

This section discusses experiments and their results. Tab. 4 displays experimental findings for two different kinds of images attacked by FGSM and S&P noise. During this stage of experiment, unexpected changes occurred in accuracy outcomes.

Table 4: Original training network testing of attacked images

Class labels	Applied attack	After attack predicted label	Normal (%)	Polyps (%)	UC (%)
Normal	FGSM	Normal	75	5.13	19.87
Polyps	FGSM	UC	19.45	35.79	44.76
UC	FGSM	Polyps	5	51	44
Normal	S&P	UC	6.511	35.45	55
Polyps	S&P	UC	1.45	36.1	62.45
UC	S&P	Polyps	1.08	75.5	23.42

When original trained network was tested with attacked images, they incorrectly predicted their classes hence showed incorrect labels for each class. When FGSM attacked images of class Polyps were tested, they were correctly categorized as belonging to class UC with 44.76% accuracy despite the fact that they are from class Polyps, just like other WCE classes. When S&P images of class Polyps were tested, they were classified as class UC with an accuracy of 62.45%. The images for Normal and UC classes were likewise mislabeled.

Tab. 5 shows the results of first adversarial training in which a dataset is prepared including half images from original dataset and half images from FGSM attacked images of three classes as Normal, Polyps and Ulcer images. Then this dataset is trained using deep CNN Alexnet model through transfer learning. The model took the features and used them to make predictions, which were then tested

through adversary and original images. The accuracy is improved when compared to previous testing on the trained network.

Table 5: Testing of ATO network with attacked images

Class labels	Applied attack	After attack predicted label	Normal (%)	Polyps (%)	UC (%)
Normal	FGSM	Normal	86.37	2.94	10.69
Polyps	FGSM	Polyps	10.54	54.48	34.98
Ulcer	FGSM	UC	4.02	26.47	69.51
Normal	S&P	Normal	57	4	39
Polyps	S&P	UC	1	42	57
UC	S&P	Polyps	1.7	55.3	43

Tab. 6 shows the results of second adversarial training in which the prepared dataset has half images from original dataset whereas half images are S&P attacked images of three classes including Normal, Polyps and Ulcer images. Then this dataset is trained using deep CNN Alexnet model through transfer learning. The model took the features and used them to make predictions, which were then tested through original and adversary images. The detected anomaly resolved all S&P attacked images classified with high 99.5%, 70.56%, and 89.25% accuracy.

Table 6: Testing of ATT network with attacked images

Class labels	Applied attack	After attack predicted label	Normal (%)	Polyps (%)	UC (%)
Normal	FGSM	Polyps	0	64.51	35.32
Polyps	FGSM	Normal	74	18.6	7.4
UC	FGSM	Normal	36	30	34
Normal	S&P	Normal	99.5	0.4	0.1
Polyps	S&P	Polyps	4.675	70.56	24.765
UC	S&P	UC	1.94	8.81	89.25

Tab. 7 shows the results of mixed adversarial training such that a complete dataset is prepared equally divided into original images, FGSM attacked images and S&P noise images. This defense is more robust than the first since more data is provided in this training, the classifier learns to work best, and the odds of model fooling are reduced in comparison to others.

The results of MAT depicted that the majority of attacked images were properly classified with high accuracy and precision, and the trained model became most resilient. Furthermore, for adversarial training, original, FGSM, and S&P datasets were used which performed well and accurately labeled the majority of labels. Summary of proposed defense model is shown in Tab. 8.

Table 7: Testing of MAT network with attacked images

Class labels	Applied attack	After attack predicted label	Normal (%)	Polyps (%)	UC (%)
Normal	FGSM	Normal	99.5	0.5	0
Polyps	FGSM	Polyps	0.4	90	9.6
UC	FGSM	UC	0	23	77
Normal	S&P	Normal	96	3.7	0.3
Polyps	S&P	Polyps	2.13	89	8.87
UC	S&P	UC	0.29	6.71	93

Table 8: Summary of proposed defense model

Training dataset	Testing dataset	Correct prediction (%)
Original dataset	Original dataset	99.6
Original dataset	FGSM attacked dataset	51.5
Original dataset	S&P attacked dataset	22
Adversarial training 1	Original + FGSM	70.12
Adversarial training 2	Original + S&P	86.43
Mixed adversarial training	Original + FGSM + S&P	96.30

4.1 Feature Fusion Defense

For the results of feature fusion defense, three classes were used including Normal Pylorus, Polyps and Ulcerative Colitis. Two different experiments were performed by generating adversarial samples of S&P and FGSM on dataset and applying two pre trained CNN including Resnet 101 &SqueezeNet.

4.2 Experiment 1 (Salt & Pepper Attack Results Using LBP + SqueezeNet + ResNet101)

In experiment 1, images are noisy with S&P noise, and pre trained CNN Resnet 101 with pool5 layer and SqueezeNet with pool10 layer are used. Image is resized into 224×224 dimensions along with fused features of ResNet101, SqueezeNet and hand-crafted LBP to calculate evaluation metrics. Tab. 9 shows that best accuracy achieved is 94.7% from Quadratic SVM classifier. Graphical comparison of classification methods is shown in Fig. 9.

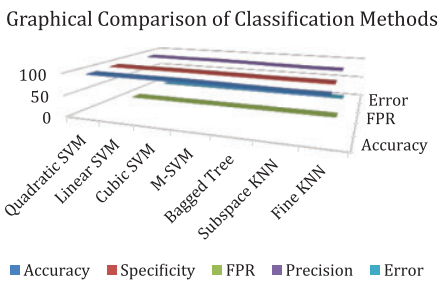

Figure 9: Graphical comparison of experiment 1 results

Table 9: Classification results on S&P attack using LBP + Resnet101 + SqueezeNet

Methods	Accuracy (%)	Sen (%)	Spe (%)	FPR	Pre (%)	Error
Quadratic SVM	94.7	94.6	97..3	0.026	94.68	0.053
Linear SVM	93.8	93.8	96.9	0.030	93.8	0.062
Cubic SVM	93.8	93.8	96.9	0.030	93.8	0.062
M-SVM	92.8	92.8	96.4	0.035	92.8	0.072
Weighted KNN	90.2	90.1	95.1	0.049	90.09	0.098
Medium KNN	90	90	95	0.05	89.9	0.1
Fine KNN	87.3	87.3	93.6	0.063	87.2	0.127

4.3 Experiment 2 (FGSM Attack Results Using ResNet101 + SqueezeNet + LBP)

In experiment 2, images are noisy with FGSM noise, and pre trained CNN Resnet 101 with pool5 layer and SqueezeNet with pool10 layer are used. Image is resized into 224×224 dimensions along with fused features of ResNet101, SqueezeNet and hand-crafted LBP to calculate evaluation metrics. Tab. 10 shows that best accuracy achieved is 99.6% from Cubic SVM classifier. Graphical comparison of classification methods is shown in Fig. 10.

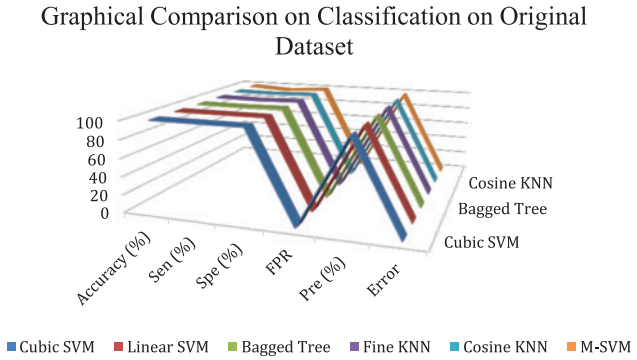

Figure 10: Graphical comparison of experiment 2 results

Table 10: Classification results on FGSM attack using LBP + Resnet101 + SqueezeNet

Method	Accuracy (%)	Sen (%)	Spe (%)	FPR	Pre (%)	Error
Cubic SVM	99.6	99.58	99.82	0.0017	99.20	0.004
Linear SVM	99.5	99.53	99.77	0.0022	98.72	0.005
Bagged tree	98.9	98.56	99.45	0.0054	97.94	0.011

(Continued)

Table 10: Continued

Method	Accuracy (%)	Sen (%)	Spe (%)	FPR	Pre (%)	Error
Fine KNN	98.9	98.09	99.43	0.0056	98.23	0.011
Cosine KNN	98.8	98.14	99.38	0.0061	98.32	0.012
M-SVM	98.8	95.62	98.79	0.014	97.65	0.012

5 Conclusion

Attacks on artificial intelligence models have the potential to degrade model performance. To address this issue, an adversarial training-based model against unfavorable interruptions is presented. In addition to WCE images process, method enhancement requires not only sophisticated grasp of image processing methods, but also critical medical input, such as professional knowledge of its screening procedure. To decrease the impact of adversarial attacks, several types of adversarial training are used, which lowered the unfavorable effect and ensured that the model could not be deceived when compared to current models. The results are 96.30% correct, demonstrating the robustness of the suggested defensive model. For adversarial attacks, another defense model based on feature fusion was suggested in which deep and handcrafted features i.e., SqueezeNet and ResNet101 deep features and LBP features were fused, and the accuracy was enhanced by 99.6%. In future, the same issue will be discussed with other set of noises. The purpose of doing all this is to give more control to different DL and AI models to perform accurately on preturbed images, efficiently and robustly against different adversarial attacks.

Funding Statement: This work was supported by "Human Resources Program in Energy Technology" of the Korea Institute of Energy Technology Evaluation and Planning (KETEP), granted financial resources from the Ministry of Trade, Industry & Energy, Republic of Korea. (No. 20204010600090).

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] K. Muhammad, M. Sharif, T. Akram and S. Kadry, "Intelligent fusion-assisted skin lesion localization and classification for smart healthcare," *Neural Computing and Applications*, vol. 21, no. 1, pp. 1–16, 2021.

[2] K. Jabeen, M. Alhaisoni, U. Tariq, Y. D. Zhang and A. Hamza, "Breast cancer classification from ultrasound images using probability-based optimal deep learning feature fusion," *Sensors*, vol. 22, no. 3, pp. 807, 2022.

[3] J. Naz, M. Alhaisoni, O. Y. Song, U. Tariq and S. Kadry, "Segmentation and classification of stomach abnormalities using deep learning," *Computers, Materials & Continua*, vol. 69, no. 3, pp. 607–625, 2021.

[4] A. Majid, N. Hussain, M. Alhaisoni, Y. D. Zhang and S. Kadry, "Multiclass stomach diseases classification using deep learning features optimization," *Computers, Materials & Continua*, vol. 69, no. 2, pp. 1–15, 2021.

[5] M. Sharif, T. Akram, M. Yasmin and R. S. Nayak, "Stomach deformities recognition using rank-based deep features selection," *Journal of Medical Systems*, vol. 43, no. 4, pp. 1–15, 2019.

[6] H. Arshad, M. I. Sharif, M. Yasmin, J. M. R. Tavares and Y. D. Zhang, "A multilevel paradigm for deep convolutional neural network features selection with an application to human gait recognition," *Expert Systems*, vol. 4, no. 3, pp. e12541.

[7] A. F. Peery, S. D. Crockett, C. C. Murphy, J. L. Lund and E. S. Dellon, "Burden and cost of gastrointestinal, liver, and pancreatic diseases in the United States: Update 2018," *Gastroenterology*, vol. 156, no. 22, pp. 254–272. e11, 2019.
[8] L. Lan, C. Ye, C. Wang and S. Zhou, "Deep convolutional neural networks for WCE abnormality detection: CNN architecture, region proposal and transfer learning," *IEEE Access*, vol. 7, no. 2, pp. 30017–30032, 2019.
[9] T. Akram, M. Sharif, N. Muhammad, M. Y. Javed and S. R. Naqvi, "Improved strategy for human action recognition; experiencing a cascaded design," *IET Image Processing*, vol. 12, no. 2, pp. 1–21, 2019.
[10] Q. Wang, N. Pan, W. Xiong, H. Lu and N. Li, "Reduction of bubble-like frames using a RSS filter in wireless capsule endoscopy video," *Optics & Laser Technology*, vol. 110, no. 2, pp. 152–157, 2019.
[11] S. Pecere, C. Senore, C. Hassan, E. Riggi and N. Segnan, "Accuracy of colon capsule endoscopy for advanced neoplasia," *Gastrointestinal Endoscopy*, vol. 91, no. 5, pp. 406–414. e1, 2020.
[12] H. G. Lee, M. K. Choi, B. S. Shin and S. C. Lee, "Reducing redundancy in wireless capsule endoscopy videos," *Computers in Biology and Medicine*, vol. 43, no. 7, pp. 670–682, 2013.
[13] E. R. Kim, "Roles of capsule endoscopy and device-assisted enteroscopy in the diagnosis and treatment of small-bowel tumors," *Clinical Endoscopy*, vol. 53, no. 22, pp. 410, 2020.
[14] R. Sharma, R. Bhadu, S. K. Soni and N. Varma, "Reduction of redundant frames in active wireless capsule endoscopy," in *Presented at the Proc. of the Second Int. Conf. on Microelectronics, Computing & Communication Systems*, NY, USA, pp. 1–6, 2019.
[15] M. Pennazio, C. Spada, R. Eliakim and M. Keuchel, "Small-bowel capsule endoscopy and device-assisted enteroscopy for diagnosis and treatment of small-bowel disorders," *Endoscopy*, vol. 47, no. 12, pp. 352–376, 2015.
[16] M. Nawaz, T. Nazir, A. Javed, U. Tariq and H. S. Yong, "An efficient deep learning approach to automatic glaucoma detection using optic disc and optic cup localization," *Sensors*, vol. 22, no. 3, pp. 434, 2022.
[17] F. Afza, M. Sharif and A. Rehman, "Microscopic skin laceration segmentation and classification: A framework of statistical normal distribution and optimal feature selection," *Microscopy Research and Technique*, vol. 82, no. 20, pp. 1471–1488, 2019.
[18] J. K. Sethi and M. Mittal, "A new feature selection method based on machine learning technique for air quality dataset," *Journal of Statistics and Management Systems*, vol. 22, no. 2, pp. 697–705, 2019.
[19] M. A. Khan, T. Akram, M. Sharif, K. Javed and S. A. C. Bukhari, "An integrated framework of skin lesion detection and recognition through saliency method and optimal deep neural network features selection," *Neural Computing and Applications*, vol. 32, no. 25, pp. 15929–15948, 2020.
[20] A. Krizhevsky, I. Sutskever and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, vol. 21, no. 3, pp. 1097–1105, 2022.
[21] M. I. Sharif, M. Alhussein, K. Aurangzeb and M. Raza, "A decision support system for multimodal brain tumor classification using deep learning," *Complex & Intelligent Systems*, vol. 2, no. 5, pp. 1–14, 2021.
[22] M. Nasir, M. Sharif, M. Alhaisoni, S. Kadry and S. A. C. Bukhari, "A blockchain based framework for stomach abnormalities recognition," *Computers, Materials & Continua*, vol. 67, no. 5, pp. 141–158, 2021.
[23] H. H. Syed, U. Tariq, A. Armghan, F. Alenezi and J. A. Khan, "A rapid artificial intelligence-based computer-aided diagnosis system for COVID-19 classification from CT images," *Behavioural Neurology*, vol. 2021, no. 5, pp. 1–15, 2021.
[24] A. Aqeel, A. Hassan, S. Rehman, U. Tariq and S. Kadry, "A long short-term memory biomarker-based prediction framework for Alzheimer's disease," *Sensors*, vol. 22, no. 5, pp. 1475, 2022.
[25] F. Afza, M. Sharif, U. Tariq, H. S. Yong and J. Cha, "Multiclass skin lesion classification using hybrid deep features selection and extreme learning machine," *Sensors*, vol. 22, no. 4, pp. 799, 2022.
[26] M. Arshad, U. Tariq, A. Armghan, F. Alenezi and M. Younus Javed, "A computer-aided diagnosis system using deep learning for multiclass skin lesion classification," *Computational Intelligence and Neuroscience*, vol. 2021, no. 3, 2021.
[27] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 3, pp. 1345–1359, 2009.

[28] I. J. Goodfellow, J. Shlens and C. Szegedy, "Explaining and harnessing adversarial examples," *Surgical Endoscopy*, vol. 31, no. 5, pp. 1–18, 2014.
[29] A. Kurakin, I. Goodfellow and S. Bengio, "Adversarial examples in the physical world," *ACM Computing Surveys*, vol. 51, no. 5, pp. 1–42. 2016.
[30] N. Carlini and D. Wagner, "Towards evaluating the robustness of neural networks," in *Presented at the 2017 IEEE Symp. on Security and Privacy*, New Delhi, India, pp. 1–8, 2017.
[31] N. Papernot, P. McDaniel, S. Jha, M. Fredrikson and A. Swami, "The limitations of deep learning in adversarial settings," in *Presented at the 2017 IEEE Symp. on Security and Privacy*, New Delhi, India, pp. 1–8, 2016.
[32] X. Ma, Y. Niu, L. Gu, Y. Wang and Y. Zhao, "Understanding adversarial attacks on deep learning based medical image analysis systems," *Pattern Recognition*, vol. 110, no. 34, pp. 107332, 2021.
[33] K. Ren, T. Zheng, Z. Qin and X. Liu, "Adversarial attacks and defenses in deep learning," *Engineering*, vol. 6, no. 2, pp. 346–360, 2020.
[34] H. Hirano, A. Minagi and K. Takemoto, "Universal adversarial attacks on deep neural networks for medical image classification," *BMC Medical Imaging*, vol. 21, no. 5, pp. 1–13, 2021.
[35] X. Li and D. Zhu, "Robust detection of adversarial attacks on medical images," in *Presented at the 2020 IEEE 17th Int. Symp. on Biomedical Imaging (ISBI)*, NY, USA, pp. 1–6, 2020.
[36] M. Z. Joel, S. Umrao, E. Chang, R. Choi and D. X. Yang, "Adversarial attack vulnerability of deep learning models for oncologic images," *BMC Medical Imaging*, vol. 21, no. 5, 2021.
[37] M. A. Khan, T. Akram, M. Sharif, M. Y. Javed and N. Muhammad, "An implementation of optimized framework for action classification using multilayers neural network on selected fused features," *Pattern Analysis and Applications*, vol. 22, no. 4, pp. 1377–1397, 2019.
[38] J. Su, D. V. Vargas and K. Sakurai, "One pixel attack for fooling deep neural networks," *Sensors*, vol. 23, no. 4, pp. 828–841, 2019.
[39] S. Gu and L. Rigazio, "Towards deep neural network architectures robust to adversarial examples," *Applied Sciences*, vol. 5, no. 1, pp. 1–21, 2015.
[40] G. K. Dziugaite, Z. Ghahramani and D. M. Roy, "A study of the effect of JPG compression on adversarial images," *Applied Sciences*, vol. 6, no. 2, pp. 1–18, 2016.
[41] N. Das, M. Shanbhogue, S. T. Chen, F. Hohman and L. Chen, "Keeping the bad guys out: Protecting and vaccinating deep learning with JPEG compression," *Sensors*, vol. 21, no. 5, pp. 1–21, 2017.
[42] A. Ross and F. Doshi-Velez, "Improving the adversarial robustness and interpretability of deep neural networks by regularizing their input gradients," in *Proc. of the AAAI Conf. on Artificial Intelligence*, vol. 32, no. 1, 2018.
[43] M. Souaidi, A. A. Abdelouahed and M. El Ansari, "Multi-scale completed local binary patterns for ulcer detection in wireless capsule endoscopy images," *Multimedia Tools and Applications*, vol. 78, no. 5, pp. 13091–13108, 2019.
[44] B. Li and M. Q. H. Meng, "Ulcer recognition in capsule endoscopy images by texture features," in *2008 7th World Congress on Intelligent Control and Automation*, NY, USA, pp. 234–239, 2008.
[45] P. Szczypiński, A. Klepaczko, M. Pazurek and P. Daniel, "Texture and color based image segmentation and pathology detection in capsule endoscopy videos," *Computer Methods and Programs in Biomedicine*, vol. 113, pp. 396–411, 2014.
[46] S. V. Georgakopoulos, D. K. Iakovidis, M. Vasilakakis, V. P. Plagianakos and A. Koulaouzidis, "Weakly-supervised convolutional learning for detection of inflammatory gastrointestinal lesions," in *2016 IEEE Int. Conf. on Imaging Systems and Techniques (IST)*, NY, USA, pp. 510–514, 2016.
[47] S. Fan, L. Xu, Y. Fan, K. Wei and L. Li, "Computer-aided detection of small intestinal ulcer and erosion in wireless capsule endoscopy images," *Physics in Medicine & Biology*, vol. 63, no. 9, pp. 165001, 2018.
[48] Y. Xu, T. Mo, Q. Feng, P. Zhong and M. Lai, "Deep learning of feature representation with multiple instance learning for medical image analysis," in *2014 IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, NY, USA, pp. 1626–1630, 2014.

[49] M. Pei, X. Wu, Y. Guo and H. Fujita, "Small bowel motility assessment based on fully convolutional networks and long short-term memory," *Knowledge-Based Systems*, vol. 121, pp. 163–172, 2017.
[50] S. Seguí, M. Drozdzal, G. Pascual, P. Radeva and C. Malagelada, "Generic feature learning for wireless capsule endoscopy analysis," *Computers in Biology and Medicine*, vol. 79, no. 21, pp. 163–172, 2016.
[51] G. Wimmer, S. Hegenbart, A. Vécsei and A. Uhl, "Convolutional neural network architectures for the automated diagnosis of celiac disease," in *Int. Workshop on Computer-Assisted and Robotic Endoscopy*, Springer, Cham, pp. 104–113, 2016.
[52] X. Jia and M. Q. H. Meng, "A deep convolutional neural network for bleeding detection in wireless capsule endoscopy images," in *2016 38th Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society*, NY, USA, pp. 639–642, 2016.
[53] O. Ronneberger, P. Fischer and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Int. Conf. on Medical Image Computing and Computer-Assisted Intervention*, Cham, Springer, pp. 234–241, 2015.
[54] M. Sharif, M. Attique Khan, M. Rashid, M. Yasmin and F. Afza, "Deep CNN and geometric features-based gastrointestinal tract diseases detection and classification from wireless capsule endoscopy images," *Journal of Experimental & Theoretical Artificial Intelligence*, vol. 25, no. 2, pp. 1–23, 2019.
[55] X. Zhang, X. Sun, W. Sun, T. Xu and S. K. Jha, "Deformation expression of soft tissue based on BP neural network," *Intelligent Automation and Soft Computing*, vol. 32, no. 2, pp. 1041–1053, 2022.
[56] W. Sun, G. Zhang, X. Zhang and N. Ge, "Fine-grained vehicle type classification using lightweight convolutional neural network with feature optimization and joint learning strategy," *Multimedia Tools and Applications*, vol. 80, no. 7, pp. 30803–30816, 2021.