Multimedia Tools and Applications (2020) 79:21941–21966
https://doi.org/10.1007/s11042-020-08905-7

Deep learning based image classification for intestinal hemorrhage

Husanbir Singh Pannu[1] · Sahil Ahuja[1] · Nitin Dang[1] · Sahil Soni[1] · Avleen Kaur Malhi[1]

Received: 30 May 2019 / Revised: 10 March 2020 / Accepted: 27 March 2020 /
Published online: 14 May 2020

Abstract

Convolutional neural networks (CNN) have become a popular choice for image segmentation and classification. Internal body images are obscure in nature with involvement of noise, luminance variation, rotation and blur. Thus optimal choice of features for machine learning model to classify bleeding is still an open problem. CNN is efficient for attribute selection and ensemble learning makes a generalized robust system. Capsule endoscopy is a new technology which enables a gastroenterologist to visualize the entire digestive tract including small bowel to diagnose bleeding, ulcer and polyp. This paper presents a supervised learning ensemble to detect the bleeding in the images of Wireless Capsule Endoscopy. It accurately finds out the best possible combination of attributes required to classify bleeding symptoms in endoscopy images. A careful setting for CNN layer options and optimizer for back propagation after reducing the color palette using minimum variance quantization has shown promising results. Results of testing on public and real dataset has been analyzed. Proposed ensemble is able to achieve 0.95 on the public endoscopy dataset and 0.93 accuracy on the real video dataset. A detailed data analysis has also been incorporated in the study including RGB pixel intensities, distributions of binary classes and various class ratios for training.

Keywords Convolutional neural networks · Ensemble · Image processing · Capsule endoscopy

1 Introduction

Convolutional Neural Networks (CNN) has become popular for wide range of image processing applications including Wireless Capsule Endoscopy (WCE) due to its powerful feature extraction and classification faculties. CNN is one of the most popular and robust

✉ Avleen Kaur Malhi
avleen@thapar.edu

Husanbir Singh Pannu
hspannu@thapar.edu

[1] Thapar Institute of Engineering and Technology, Patiala, India

class of deep neural networks [20]. It is particularly useful for finding patterns in the images to recognize objects, faces, and scenes. They have the ability to learn directly from the dataset of the images and thus detecting patterns to classify images and thereby eliminating the need for manual feature extraction. It is a synthetic neural network to perform image analysis through pattern recognition with the help of non linear activation functions and neuron weights, calculated during training process. This property of pattern detection is incredibly helpful to make CNN of the utmost importance. It has hidden layers known as convolutional layers as its basis which makes it completely different from normal Multi-Layer Perceptron (MLP) [8]. These convolutional layers perform scanning operations on the image to look for various features. They receive the input, rework in a way to concise it and forward it to selected subsequent layers. Nowadays CNN has become a really powerful tool for solving image classification tasks as they are able to learn highly discriminating features from raw pixel intensities. However, their applicability to medical image analysis is restricted by the non-availability of enormous sets of annotated knowledge needed for training the CNN model. Although image analysis has been the foremost widespread use of CNN, they can even be used for different data type analyses including classification and regression [21].

Wireless Capsule Endoscopy (WCE) is a new technique to visualize entire digestive trace including small bowel using non-invasive means. It is primarily used to examine areas in small intestine where other endoscopy methods fail. The paper proposes a new CNN augmented architecture for detecting the obscure bleeding cases from wireless capsule endoscopy images. Nevertheless, the proposed method is general and can be utilised for other image classification applications as well. Below are the challenges and contributions of the underlying study.

Challenges

1. In-vivo internal body organ images obtained from capsule endoscopy suffer from various issues such as (a) movement of capsule camera for site capturing (b) organ movements (c) non-ideal luminance and focal conditions to capture (d) no control over the movement of both camera and the organs [4].
2. Images suffer from compression resulting in information loss. Moreover noise and blur are involved during capture phase [15].
3. Data collected for anomaly detection is often imbalanced with greater instances of normal case images relatively. Sometimes the ratio is 1000:1 for healthy:sick while making the classification problem difficult to formulate [1].
4. Optimal CNN architecture also depends upon the underlying data distribution along with layer+parameter tuning. Therefore it is a meta-heuristic problem without a clear cut well defined solution [38].

Contribution

1. This paper proposes a CNN augmented ensemble architecture for training, validation and testing the obscure bleeding symptoms from Capsule Endoscopy (CE) images.
2. For ensemble, copies of the image data have been generated using synthetic augmentation by introducing rotation, blur, illuminance and noise.
3. Color palette reduction has been used using minimum variance quantization, studied in [32] to get 24 colors in the image for optimal balance of accuracy and time-space complexity.
4. Data distribution analysis of the RGB pixel intensities has been studied in detail to demonstrate the complexity of the classification problem.

5. The experimental analysis of the proposed technique has been performed on the public and real dataset and compared with other state-of-art methods.

This paper is structured as follow: Section 2 is about the background of CNN and WCE, Section 3 is about literature survey to trigger the research motivation, Section 4 is about proposed method, Section 5 is about results and experiments and finally Section 6 is conclusion and future possibilities.

2 Background - CNN and WCE

A brief discussion about CNN and WCE has been presened in this section with relevant illustrations.

2.1 Convolutional neural networks

According Professor Geoffrey Hinton (father of convolutional networks) in [22], factors responsible for increased use of CNN are as follows:

- Eliminate the need of manual feature extraction thereby, saving time and labour. CNN mimics human intelligence to learn from input image examples automatically.
- CNN produce state-of-the-art recognition results with improved quality and performance.
- It can be retrained for new recognition tasks, enabling you to build on pre-existing networks.

CNN is composed of an input layer, output layer, and many hidden convolutional layers as intermediate layers (Fig. 1). These layers perform operations that alter the data with the intent of learning features specific to the data. Fundamental inherent components of CNN are: convolution, activation function such as Rectified Linear Unit (ReLU), pooling and fully connected layer. CNN has varied filters to facilitate the pattern recognition ranging from simple to obscure for automatic image feature extraction.

Images features may be colors, texture, boundaries or shapes which are detected by using various geometric filters. Classical conception was that deeper the network, more refined are the filters to detect mature features at the cost of computation time. But latest architectures are able to perform better with fewer layers studied in [11]. In convolutional layer, the block of pixel set (convolution matrix) moves over the whole image while filtering across every block of pixels of the input image.

Output of last convolutional block will yield the feature volume. As an example, Fig. 12 shows the features from three convolution layers and fully connected layers for example

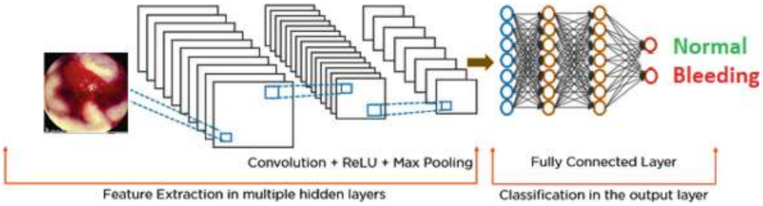

Normal
Bleeding
Convolution + ReLU + Max Pooling
Fully Connected Layer
Feature Extraction in multiple hidden layers
Classification in the output layer

Fig. 1 Basic structure of CNN involved in binary classification

gastral images case study. The features are present in the output of the convolutional channels. The fully connected layer produces representation vectors that already digested the features.

It explains about the automatic feature learning and classification phases of CNN framework which is similar to identify bleeding against normal images in endoscopy images for the underlying case.

CNN terminology Convolution means to blend two functions simultaneously and can be described as follows in case of two variables for digital images:

$$(f * g)(x, y) = \sum_{i=0}^{N} \sum_{j=0}^{N} f(i, j).g(x - i, y - j) \qquad (1)$$

where f, g are two functions with x, y variables, i, j are indices that run for a $N \times N$ pixel image for example. The dot '.' on the right hand side is the element by element product. Thus the convolution is a sum of pointwise products of function values, which is conditioned to traversal. CNN is mostly used to learn from three channel coloured image with input size of $256 \times 256 \times 3$ i.e. the intensity matrices (0-255 values) for R,G and B channels. Input images are known as input volume. Kernels are small squared matrices which are convolved with input volume. This process generates activation maps which indicate the highlighted regions detected in the image. During the training process, the kernel matrix values change to extract required features and highlight the important regions for learning. Details can be found in [31]. In CNN, the receptive field (a two dimensional region say 5×5) specify the required neuron connections instead of all possible neuron connections. So, if the input volume is $28 \times 28 \times 3$ and receptive field is 5×5, then in convolutional layer, every neuron is connected to $5 \times 5 \times 3$ in the input volume for all 3 color channels. Therefore, number of neuron weights will be 75 which gets updated during successive iterations of Back-propagation similar to a feed-forward network.

CNN architecture Architecture of Convolutional Neural Network consists of following layers (1) Convolutional layer (2) Pooling layer (3) Rectified Liner Unit layer (4) Fully connected layer. The CNN structure anatomy for the proposed network is shown in Table 1

1. Convolutional layer performs the operations of convolution on the input volume and responsible for neutron firing. It has three dimensional structure due to RGB channels. Neuron connections are defined by the receptive field. This layer calculates the low level image features such as lines, edges and corners. By combining multiple maps, each output map feature is calculated as follows:

$$y_k^t = f\left(\sum_{i \in M_k} x_i^{t-1} * K_{ik}^t + b_k^t\right) \qquad (2)$$

   where t is t^{th} layer, M is a set of input maps , K_{ik} is convolutional kernel, b_k is bias. Please refer [31] for details.
2. Pooling layer reduces the dimensions (down-sampling) of input volume for successive convolutional operations by keeping the depth same. Down-sampling avoids overfitting and reduces the overhead. Similar to convolutional layer, it uses sliding window using max or average operations.

Table 1 CNN structure anatomy for the proposed network

Sr	Variable
1	Image Input 100x100x3
	8, 3x3 Convolution - 1
2	Batch Normalization - 1
	ReLU-1
	2x2 Max Pooling - 1
	16, 3x3 Convolution-2
3	Batch Normalization-2
	ReLU-2
	2x2 Max Pooling-2
	32, 3x3 Convolution-3
	Batch Normalization-3
4	ReLU-3
	Fully Connected
5	Softmax
6	Classification Output

3. ReLu layer uses the function $max(0, x)$ but it is not differentiable at $(0, 0)$. Therefore a smoother function is used. For example Softplus function which is integral of the Sigmoid function as studied in [47].

$$f(x) = ln(1 + e^x) \tag{3}$$

4. Fully connected layer is completely connected with its preceding layer and follows the output layer. It is the last stage of CNN.

After convolution, the size of the output image can be calculated using the following formula [31]:

$$W_2 = \frac{W_1 - F + 2P}{S} + 1 \tag{4}$$

where W is input volume, F is the receptive field, P is the padding value, S is the stride. For example, let the image has size $256 \times 256 \times 3$, $(W_1 = 256)$, $S = 2$, $F = 4$, $P = 0$. Then $W_2 = \frac{256-4+2.0}{2} + 1 = 128$.

The convolution, pooling, ReLu layer operations are repeated over tens or hundreds of layers, with each layer learning to identify different features. After CNN layers are trained, the CNN architecture shifts to classification phase which is the final phase of its life cycle. The last but one layer is a fully connected layer that outputs a vector equivalent to length of number of classes (binary in our case) that the network will be able to predict. This vector contains the probabilities for each class of any image being classified. The final layer of the CNN architecture uses a classification layer such as softmax to provide the output of the classification of the model.

2.2 Wireless capsule endoscopy

It involves a pill camera similar to the the size of a large vitamin tablet. Two frames per second are recorded by the camera. The patient swallows the capsule and the camera will capture the pictures of entire gastrointestinal tract. Capsule records for about 8 hours and delivers about 55,000 images to the receiver which is put on the waist of the person being examined. When this data is transferred into the computer, the gastroenterologist can see the entire GI tract. The example images of small bowel bleeding symptoms against normal images have been shown in the Fig. 2. The internal body images are blurry red brown in color but bleeding images are more saturated in redness. Digestive tract may involve bleeding, ulcers, tumors and polyp to damage on the intestinal wall or mucous membrane. Capsule endoscopy is mainly used to find the cause of unexplained bleeding in the digestive tract or in case of inflammation or tumors in the small intestine.

Capsule endoscopy is a safe procedure involving little risks only such as the capsule might remain lodged in the body rather than leaving the body through bowel movement. The purpose is not only to identify bleeding but also exact position of bleeding instances [5, 28]. Therefore, automation of this manual procedure would save a lot of valuable time of the doctors and patients waiting for the results of the diagnosis of the endoscopy recording. The current technologies are efficient to traverse WCE images swiftly but still lack automatic diagnosis capability with 100% accuracy. The capsule involved in WCE can be placed in a child with weight as low as 10kg making it a better option for people of all ages. In spite of its benefits, CE is neither a substitute for upper endoscopy nor for colonoscopy. The reason is that CE is only a diagnostic technique and also, it is non-invasive so it cannot help in biopsy.

The developing predominance of intestinal bleeding overall expands the quantity of cases that must be surveyed by doctors. Furthermore, the cost involving the regular check ups and examination are very high and absence of expert specialists keep many people away from accepting a sufficient and effective medication. Algorithmic image classification of WCE is an attempt to assist the analysis of huge image datasets with promising quality and lesser expenses. Below is the literature survey which motivated this research.

3 Literature survey

Authors of this current research paper have also recently worked on explainable machine learning model (based upon CNN) to classify in-vivo gastral images in [33]. It has the

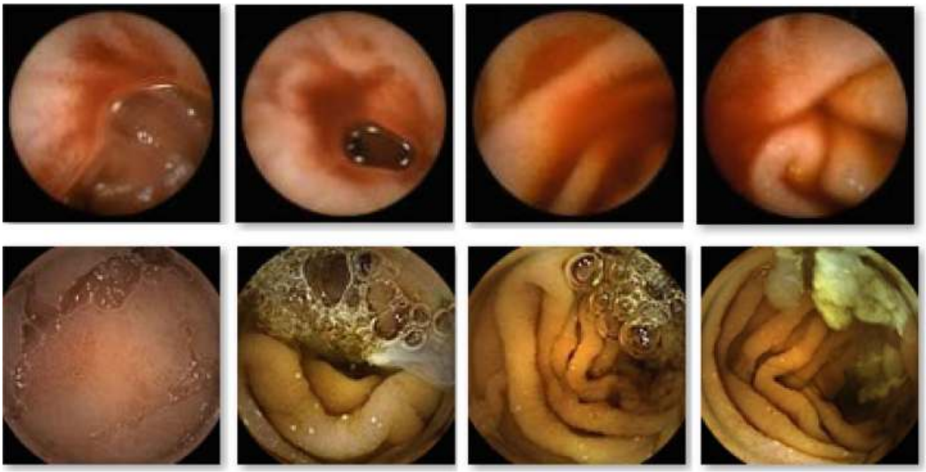


Fig. 2 Row 1 shows bleeding samples and row 2 shows normal images of GI tract from public data in [5]

capability to justify the test results in contrast to the features involved while highlighting the segmented region of interest in the medical images. Explanable Artificial Intelligence (XAI) provides the insight to a black box machine learning model to reason for the test results derivation [2].

Review on the Applications of Deep Learning for Gastral images has been studied in [6]. Automatic bleeding zone detection using low complexity CNN Structure has been studied in [12] and using multi-stage attention-Unet in [25]. Xing et al.[44] have generated a bleeding detection algorithm of three stages. Pre-processing along with key frame extraction and edge removal is computed in the first stage. The second stage is to differentiate the bleeding frames using a novel superpixel-color histogram (SPCH) feature build on the principle color spectrum and later the decision is made by a subspace KNN classifier. Lastly, segmentation of the bleeding regions is done by extracting a 9-D color feature vector from the multiple color spaces at the superpixel level. Jia et al.[16] present a technique for automatic GI bleeding detection based on a deep CNN. Then performance is calculated by constructing a big WCE dataset that contains nearly 10,000 annotated images for bleeding detection. An eight-layer convolutional neural network that contain three convolutional layers, two fully-connected layers, and three pooling layers are made. After that model is trained using stochastic gradient descent having a batch size of 100 examples, learning rate, momentum, and weight decay of 0.001, 0.9 and 0.004 respectively. The proposed method is lastly modified by replacing the second fully-connected layer with an SVM classifier and complete the bleeding frame detection.

In [34], Probabilistic Neural Network (PNN) is used to extract features of the bleeding region in WCE images distinguishing from the non-bleeding region. Bleeding detection algorithm along with programming are implemented, and the results correctly recognize the bleeding regions in WCE images and clearly mark them out with sensitivity and specificity as $93.1\%\ and\ 85.6\%$ respectively. Sekuboyina et al. [42] propose a technique in which splitting the image into several patches and extracting features pertaining to each block using CNN to automate the abnormality detection in WCE images. It increases their generality while overcoming the drawbacks of manually crafted features. This paper results in sensitivity, specificity as $71\pm 19\%$ and $72\pm 3\%$ respectively. In [13] Hajabdollahi et al. propose a simple and efficient method for segmentation of the bleeding regions in WCE captured images. Suitable colour channels are used and classified by MLP structure. They proposed that quantized neural network without any multiplication could be considered as automatic diagnostic approach inside the capsule. Ghosh et al. [9] trained CNN using SegNet layers with three classes. Then endoscopy image is segmented using the training network and the detected bleeding zones are marked. The best performance is achieved using the hue saturation and value (HSV) color space among different color planes. Performance is evaluated on a publicly available clinical dataset, and framework achieves 94.42% global accuracy.

A deep learning U-Net architecture was used to detect and segment red lesions in the small bowel by Coelho et al. in [5]. This U-Net was evaluated in an annotated sequence by using the Suspected Blood Indicator (SBI) tool and state of art techniques. U-net outperformed its peers in both detection and lesion segmentation. Its lesion detection accuracy outperformed the state-of-art methods by 1.78% in accuracy. Li et al [23] proposed a SVM based approach that combines chrominance moments and uniform LBP as color texture feature for bleeding detection on data collected from 10 patients. Further, Giritharan et al [10] also proposed a SVM based approach for rebalancing the samples by oversampling the minority class and undersampling the majority class and obtaied the accuracy of 90%. In [26], transfer learning has been studied for GI bleeding diagnosis using pre-trained

V3 model on ImageNet dataset. Data re-sampling have been used to increase the positive sample rate of the training sets for CNN.

Table 2 summarizes the comparison of recent techniques while listing their year, basic approach, dataset, highlights and results obtained. The state-of-the-art techniques only focus on the classification algorithms on static data without introducing any variations.

4 Proposed method

This sections explains the basic architecture of CNN model used followed by the proposed CNN ensemble based on data augmentation and finally details the proposed model training.

4.1 CNN architecture layers

There are a variety to deep learning models available for CNN architecture. The architecture of each CNN in the proposed ensemble has been inspired from [17, 41, 46] and modified to tune the learning accuracy of the model on the underlying dataset. Similar to the CNN model proposed by Jia et al. [16], each of the proposed CNN involved in the ensemble has been comprised of eight-layers that involve three convolutional layers (C1-C3), two fully-connected layers (FC1, FC2) and three pooling layers (MP1-MP3). Flow diagram for CNN starting from input image, role of various layers and the final binary outcome has been shown in Fig. 3. Table 3 contains the variable values of the layers used in CNN architecture.

4.2 Data augmentation

The proposed method uses dynamic data for CNN training using augmentation during input phase. This augmentation involves generating copies of one image into multiple possibilities of blur, luminance, noise and rotation variances of the image, for efficient and generalized training. The reason of data augmentation is to increase the size of dataset for real time application and avoid overfitting in case of deep learning according to [35, 40]. By utilising blur, luminance, noise and rotation variances of the images, the labelled dataset size can be increased. Labelled data is difficult and expensive to obtain from expert gastroenterologists and also time consuming. It also involves patient's private information and thus difficult to acquire from hospitals.

For the proposed CNN ensemble, the data is augmented with *four* transformations and five copies of each data image ($4 \times 5 = 20$ copies per image). Afterwards, each of these five copies are fed into each of the CNN and the aggregate result calculates the binary classification of bleeding and normal image. Figure 4 shows original image, rotation, blur, illuminance alteration, and noise into the original image. Details of each step is sequentially represented in the following subsections. In order to reduce the variance error (in the bias variance graph) for the CNN model, it is useful to consider various deformations in the training data using data augmentations and use ensemble of CNN learning. Various data versions would be infused into CNN ensemble for the aggregate decision. It will make the model capable of dealing with such spectrum of degradation. The CNN ensemble would help finding the generalized decision function thereafter.

While the wireless endoscopy camera captures the images in the small intestine, it might take pictures with random arbitrary angles and distances. The images captured might be blurry due to poor focusing of the camera or it might contain bubbles as shown in Fig. 2. Due

Table 2 Caomparative analysis of literature review

Sr	Ref	Year	Approach	Dataset	Highlights	Results
1	Coelho et al.[5]	2018	U-Net	Training on their own dataset. of set 1: 3,295 images and set2: 600 images 512×512×3	Deep learning U-Net used, outperformed SBI and studied algorithms	acc=0.9588
2	Xing et al.[44]	2018	SVM, KNN and SPCH	500 -normal; 500 -bleeding WCE img.; 30 Patients; Training set : 680 img; Testing set : 320 img. Size of img. =256 × 256	Principal Color Spectrum Extraction using superpixeled-color Histogram; 9D feature using RGB and HSV	acc=0.99; sen=0.98; spc=0.99
3	Jia et al.[16]	2016	CNN	7150 normal; 2850 bleeding WCE img; Training set : 2050-bleeding; 6150-normal; Testing set :800-bleeding and 1000-normal; Size of img =240 × 240 × 3	Large WCE dataset; CNN used	F1 = 0.99
4	Pan et al.[34]	2010	PNN	20 img. with 339850-bleeding pixels; 590725-normal pixels; 150 full WCE video clips with 3172-bleeding img. and 11458-normal img.	RGB and HSI space; Probabilistic Classification with ANN; 6D feature vector.	sen.=0.93; spec.=0.85
5	Li et al.[23]	2009	SVM and MLP	From 10 Patients; Training:1800-bleeding and 1800-normal img; Testing:1800-bleeding and 1800-normal img	Combines Chrominance moments and uniform LBP as color textural feature	acc=0.90
6	Sekubo- yina et al.[42]	2017	CNN	Images rate of 3fps with Size= 320×320 px; 137 img were used, 77 images with abnormalities	Determine optimal color space components for feature extraction/classifier design	acc=0.79; sens=0.71, spec=0.72

Table 2 (continued)

Sr	Ref	Year	Approach	Dataset	Highlights	Results
7	Hajabdollahi et al.[13]	2018	MLP	50 images used for training and testing, Images show that only 0.2% of the pixels belong to the bleeding class	Efficient method for segmentation of bleeding regions in WCE	Min. MLP:0.68;Max. MLP:0.98;Avg. MLP:0.85
8	Girith- aran et al. [10]	2008	SVM	Approx 55000 images: minority:majority= 1:145. The training set: 550 bleeding img Size= 256×256 in size with the color depth=8-bit.	Re-balancing the training samples by over-sampling the minority class and under-sampling the majority class.	spec: 0.90; sen: 0.60-0.70;
9	Jia et al. [17]	2017	CNN and hand-crafted (HC) features	Training set: Randomly sampled 200-bleeding; 800-normal WCE img. Testing set: 100-bleeding; 400-normal; WCE img. Size= 240 × 240 × 3.	GI bleeding detection via CNN and HC; Performing Fewer training samples and low computing cost.	R=0.91; P=0.94;F1=0.92
10	Ghosh et al .[9]	2018	SegNet	Public data [15], train:test = 201:134 using semantic segmentation layers; Size=256 × 256 × 3.	Trained using SegNet layers with 3 classes. Best perform. using HSV	global acc = 0.94
11	Li et al.[26]	2017	CNN	2890 total images: 185-bleeding img; Size=480 × 480 × 3.	Inception V3 model trained on ImageNet dataset for mid-level img. representation.	acc=0.95
12	Sainju et al.[39]	2014	ONN	Training with 1500 images. and testing with 100 total images: 50-bleeding and 50-non bleeding	Statistical features and region growing, integration of a partially auto mated annotation	acc=0.93

Table 2 (continued)

Sr	Ref	Year	Approach	Dataset	Highlights	Results
13	Figueiredo et al.[7]	2013	Anisotropic diffusion, segmentation	Three data sets D1:700 with 28 bleeding images, D2:1200 with 62 bleeding images, D3:2087 with 367 bleeding images and fourth data set D4, combination of all the three data-sets together of 4000 images.	second component of CIE Lab colour space used, several medical data-sets	acc=0.92
14	Usman et al.[43]	2016	SVM	8500 images from 30 WCE videos; bleeding: 3000	Threshold analysis in HSV color space, image enhancement and edge removal	acc=0.92; sens=0.94; spec=0.91
15	Xiong et al.[45]	2015	MST and SVM	3596 WCE images from 5 patients; bleeding: 599 Size=256 × 240 × 8 pixels with 3-channel color depth	MST and K mean clustering algorithms used, enhancement operations to improve contrast	acc=0.94; sens=0.92; spec=0.94

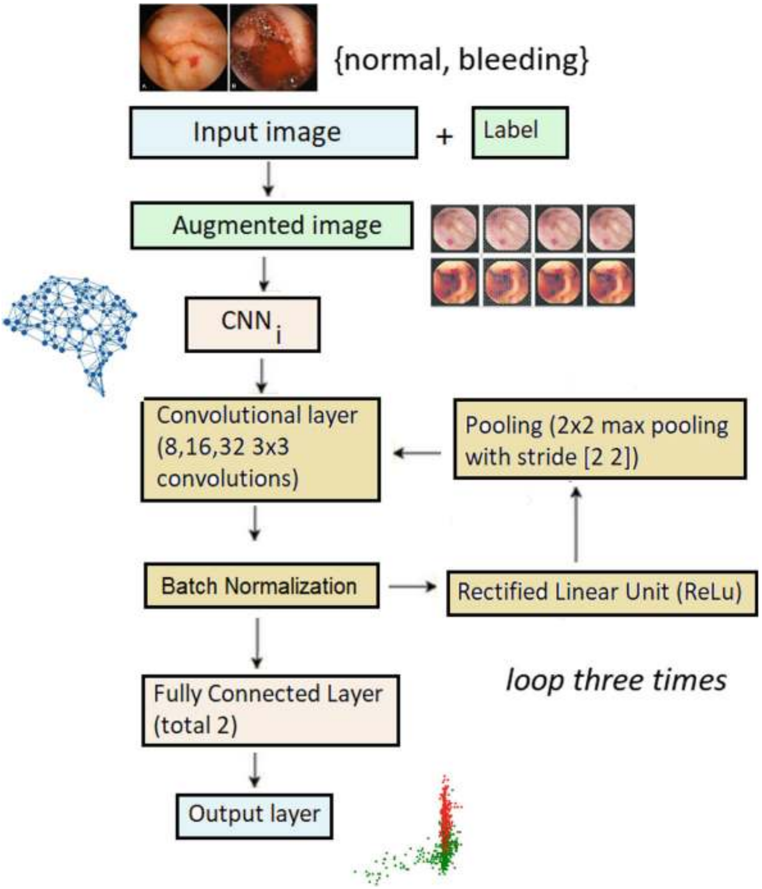

{normal, bleeding}
Input image
+
Label
Augmented image
CNN$_i$
Convolutional layer (8,16,32 3x3 convolutions)
Pooling (2x2 max pooling with stride [2 2])
Batch Normalization
Rectified Linear Unit (ReLu)
loop three times
Fully Connected Layer (total 2)
Output layer

Fig. 3 Flow diagram for individual CNNs involved in the proposed ensemble. The output layer goes into aggregation function which is the majority voting for binary classification

to capsule movement, variations in the angles of image rotation (R), illuminance (L) conditions, blurring (B), and noise (N) [24] are simulated. Again, the results might be affected by illuminance conditions. The proposed algorithm should be able to deal with these types of degraded images and the following transformations are considered for studying these degradations which are also shown in Fig. 4. Figure 5 presents the block diagram of our proposed ensemble for image classification. Each of the augmentation is randomly input into CNN models. The ensemble learning helps to reduces the bias caused by individual classifiers for the given set of parameters. Thus, ensemble is more generalised as compared to individual classifiers. In [48], the review of various ensemble learning models and motivation has been studied in detail. A similar ensemble setup has been studied in [30] for traffic surveillance to classify vehicle types.

1. **Rotation:** Images have been rotated with {60°, 120°, 180°, 240°, 300°} as 5 different values of degrees by using nearest-neighbor interpolation.
2. **Luminance change:** To tune the luminance, the image is first converted from RGB to YCbCr. A scalar L is chosen in a random and uniform manner in the interval of [0,

Table 3 MATLAB value of variables and their names used in the CNN layers for training

Sr	Variable value	Variable name	Description
1	‘imageinput’	Image Input	100x100x3 images with ‘zerocenter’ normalization
2	‘conv_1’	Convolution	8 3x3 convolutions with stride [1 1] and padding ‘same’
3	‘batchnorm_1’	Batch Normalization	Batch normalization
4	‘relu_1’	ReLU	Rectified Linear Unit
5	‘maxpool_1’	Max Pooling	2x2 max pooling with stride [2 2] and padding [0 0 0 0]
6	‘conv_2’	Convolution	16 3x3 convolutions with stride [1 1] and padding ‘same’
7	‘batchnorm_2’	Batch Normalization	Batch normalization
8	‘relu_2’	ReLU	Rectified Linear Unit
9	‘maxpool_2’	Max Pooling	2x2 max pooling with stride [2 2] and padding [0 0 0 0]
10	‘conv_3’	Convolution	32 3x3 convolutions with stride [1 1] and padding ‘same’
11	‘batchnorm_3’	Batch Normalization	Batch normalization
12	‘relu_3’	ReLU	Rectified Linear Unit
13	‘fc’	Fully Connected	2 fully connected layer
14	‘softmax’	Softmax	Softmax
15	‘classoutput’	Classification Output	crossentropyex - hyperparameter for loss function
16	‘he’	He method	Weights Initializer using He’s method [14]

1] and multiplying L with the Y channel. Thus, the modified image is converted back to RGB. Five different variations of YCbCr versions of the image were obtained using random variations in luminance (Y values). The random factors help to spread out the luminance across whole value space possible.

3. **Blurring:** A circular-symmetric 2-D Gaussian kernel of $w \sigma w$ pixels and Standard deviation σ is used for RGB components for image filtering. RGB components were blurred using this kernel generating 5 versions of each frame with random w and σ values as follows: $(w, \sigma) \in \{(11, 9), (9, 7), (7, 5), (5, 3), (3, 1)\}$. The original and blurred version is shown in Fig. 4 which illustrates that the details in the image are weakened by using blurred operation.
4. **Poisson Noise:** The photon counter endoscopic camera introduces noise following Poisson distribution [37]. The WCE images are being added a Poisson noise. Thus, the output is generated for each pixel using Poisson distribution by computing mean of pixel value. Five versions of noise variants have been generated using MATLAB function `imnoise` as illustrated in Fig. 4.

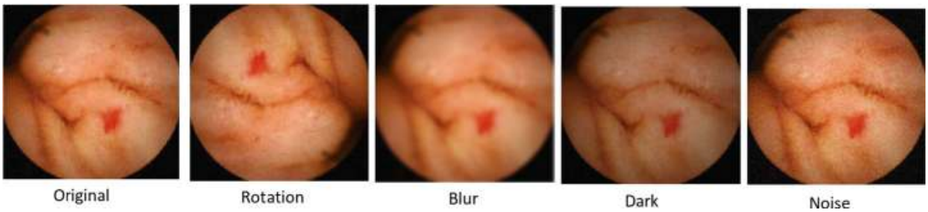


Fig. 4 Various image augmentations of the original image for ensemble setup shown in Figure 5

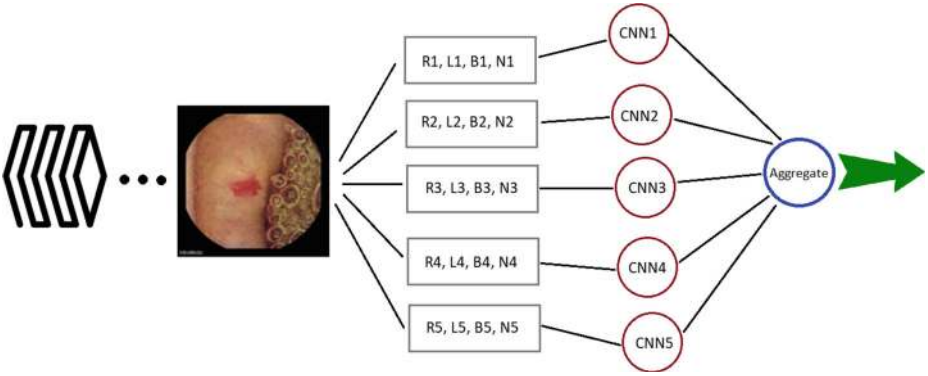


Fig. 5 Proposed flow of CNN ensemble (left to right). One image generates $4 \times 5 = 20$ copies by augmenting rotation, illuminance, blur and noise assigned randomly to CNN models. Aggregate function is majority voting for binary classification

Thus in total, $4 \times 5 = 20$ images have been generated for each image input and introduced into 5 different CNN models in the ensemble framework. There are almost 60,000 images for a real capsule endoscopy video, but our public dataset [5] contains only 3,895 images. Thus 20 fold augmentation helps to achieve the real data simulation $(3,895 \times 20 = 77,900$. Since only a small proportion of intestinal images contain , so augmentation of the images is beneficial for robust model training and data enhancement [35]. In Fig. 5, $\{R_i, L_i, B_i, N_i\}_{i=1}^{5}$ signify various versions of rotation, luminance, blur and noise as explain in above bullet points.

4.3 Training of proposed method

Fine tuning of the model is vital for achieving a premium quality performance in especially in case of medical image analysis. It is because health related decision are critical and internal body images obtained are not often the best quality for automation. Parameter tuning helps to mitigate the problems of over-fitting and speeding up the training process as well [36]. CNN model have been trained and Table 3 highlights the CNN network layer configuration used in training of proposed model. He's method has been used for weights initialization [14]. For back propagation in CNN, there are various optimizing algorithms options available such as Stochastic Gradient Descent with Momentum (SGDM) solver [3], Root Mean Square Propagation (RMSPROP) and, Adam Stochastic Gradient Descent (ASGD). SGDM has been found to perform relatively better in the experiments. SGDM optimizer specifications have been summarized in Table 4. The epochs were initially set to 20 but the similar accuracy was achieved after 4 epochs as depicted in Figs. 6 and 7. Hence, the number of epochs were reduced to 4 gradually to increase the computation time.

5 Imagenet vs training from scratch

This section discusses about pre-training using ImageNet versus training from scratch for the underlying medical dataset specifically related to endoscopy. Medical dataset has two primitive attributes while working with deep learning: (a) data content is different from pre-trained (Imagenet) model (b) smaller sample dataset. Capsule endoscopy dataset is quite

Table 4 Auxiliary table for MATLAB CNN training options for SGDM optimizer

Sr	Training Options of SGDM	Properties
1	Momentum:	0.9
2	InitialLearnRate:	0.01
3	LearnRateScheduleSettings:	[1 × 1 struct]
4	L2Regularization:	1.00E-04
5	GradientThresholdMethod:	‘l2norm’
6	GradientThreshold:	Inf
7	MaxEpochs:	4*
8	MiniBatchSize:	128
9	Verbose:	0
10	VerboseFrequency:	50
11	ValidationData:	[1 × 1 matlab. io. datastore. ImageDatastore]
12	ValidationFrequency:	3
13	ValidationPatience:	Inf
14	Shuffle:	‘every-epoch’
15	CheckpointPath:	‘ ’
16	ExecutionEnvironment:	‘auto’
17	WorkerLoad:	[ ]
18	OutputFcn:	[ ]
18	Plots:	‘training-progress’
20	SequenceLength:	‘longest’
21	SequencePaddingValue:	0

*(Epochs were initially set to 20 but decremented to 4 gradually for the similar accuracy but faster computation as shown in Figs. 6 and 7). Validation Frequency means frequency of network validation in number of iterations

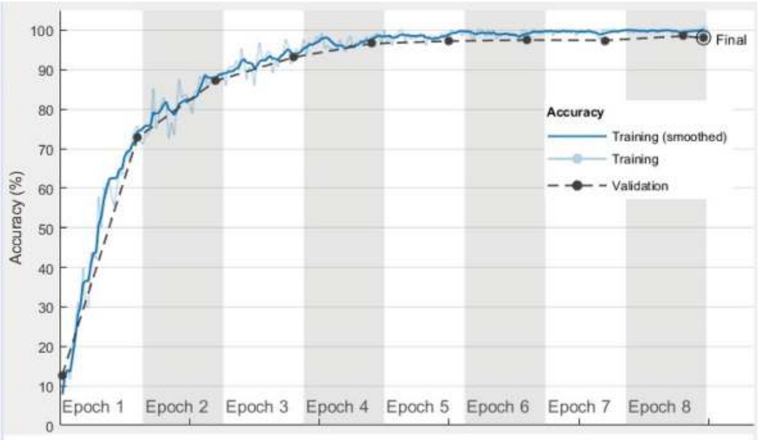


Fig. 6 CNN accuracy curve for the public dataset

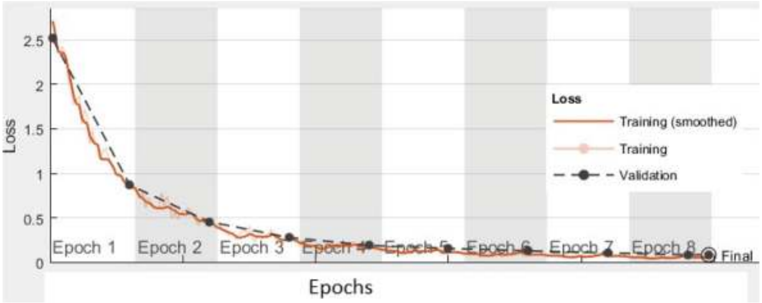
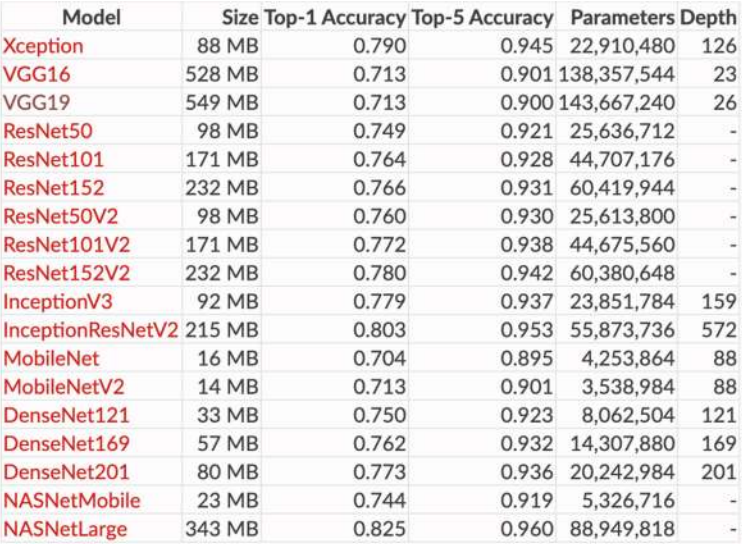


Fig. 7 CNN loss curve for the public datset

different and obscure as compared to the available natural images in Imagenet. Therefore primitive features contained in the first layer of pre-trained Imagenet model may not be useful. Hence model training from scratch has found to work better in the underlying case study. Moreover Fig. 8 shows the size of ImageNet files in MB which needs to be downloaded for the pre-trained model (obtained from https://keras.io/applications/). Even by using the initial parameters obtained through Imagenet model followed by updating the network parameter results in same classification accuracy. To summarize below are the major aspects to be considered before using transfer learning (such as Imagenet):

- If the trained dataset has enough labelled examples and nature of the dataset is quite different from ImageNet then it is better to train CNN from scratch. ImageNet does not have endoscopy medical dataset [29].
- The data distribution of underlying data should be similar to the pre-trained ImageNet dataset. Otherwise training and testing will not be consistent.



Model	Size	Top-1 Accuracy	Top-5 Accuracy	Parameters	Depth
Xception	88 MB	0.790	0.945	22,910,480	126
VGG16	528 MB	0.713	0.901	138,357,544	23
VGG19	549 MB	0.713	0.900	143,667,240	26
ResNet50	98 MB	0.749	0.921	25,636,712	-
ResNet101	171 MB	0.764	0.928	44,707,176	-
ResNet152	232 MB	0.766	0.931	60,419,944	-
ResNet50V2	98 MB	0.760	0.930	25,613,800	-
ResNet101V2	171 MB	0.772	0.938	44,675,560	-
ResNet152V2	232 MB	0.780	0.942	60,380,648	-
InceptionV3	92 MB	0.779	0.937	23,851,784	159
InceptionResNetV2	215 MB	0.803	0.953	55,873,736	572
MobileNet	16 MB	0.704	0.895	4,253,864	88
MobileNetV2	14 MB	0.713	0.901	3,538,984	88
DenseNet121	33 MB	0.750	0.923	8,062,504	121
DenseNet169	57 MB	0.762	0.932	14,307,880	169
DenseNet201	80 MB	0.773	0.936	20,242,984	201
NASNetMobile	23 MB	0.744	0.919	5,326,716	-
NASNetLarge	343 MB	0.825	0.960	88,949,818	-

Fig. 8 Top-1 and top-5 accuracy shows the validation performance of model on ImageNet data. Depth inclues activation layers and batch normalization layers (obtained from https://keras.io/applications/)

- If the labelled data is less than pre-trainined model (such as ResNet) which comprises million parameters, it could result in overfitting [29]. Number of layers is one of the hyper-parameter which is not easy to reduce in order to deal with overfitting. Moreover, it is computationally expensive to figure out the layers and neurons which needs to be eliminated for dealing with overfitting problem.

Thus, pre-trained networks can be used but one has to fine-tune the parameters, number of layer/neurons making the results not better as compared to training from scratch. The coverage time could be better for transfer learning but performance may not, since the data distribution is different. In [19], a special transfer learning based upon modality-bridge has been proposed which uses a special bridge database for the underlying medical imaging to reduce the domain difference between nature images and medical images.

6 Experiments

This section contains discussion about public dataset, system configuration, minimum variance quantization for preprocessing, performance metrics, and results analysis on public and real dataset.

6.1 Public dataset

Public dataset for small bowel lesion has been obtained from [5] containing two sets of $3,295 + 600 = 3,895$ as shown in Table 5. The dataset also contains target labels of bleeding/normal along with bleeding regions segmented manually which played a crucial role in the performance analysis. These image samples are good representative of small bowel situation including normal and bleeding cases. All the images have been re-sized to 256×256.

6.2 System configuration

The HP Pavilion X360 on the 64-bit Windows 10 operating system, having the Intel i5 processor has been used to run the experiments. HP Pavilion X360 graphics were produced courtesy of the integrated Intel UHD 620 Graphics along with the NVDIA MX130. Latest version of MATLAB R2018b has been used for programming the proposed method.

6.3 Minimum variance quantization

Quantization helps to reduce the color pallete in the underlying images to help focus on the region of interest and faster computation. This algorithm begins with under sampling the lattice of image pixels, use low-pass filter to blur the image to get the average colors present in uniform sections and finally find the color clusters by using fast quantization methods

Table 5 Dataset description of normal and bleeding examples obtained from [5]

Data sets	Normal	Bleeding	Total
Set 1	2,164	1,131	3,295
Set 2	161	439	600
Total	2,325	1,570	3,895

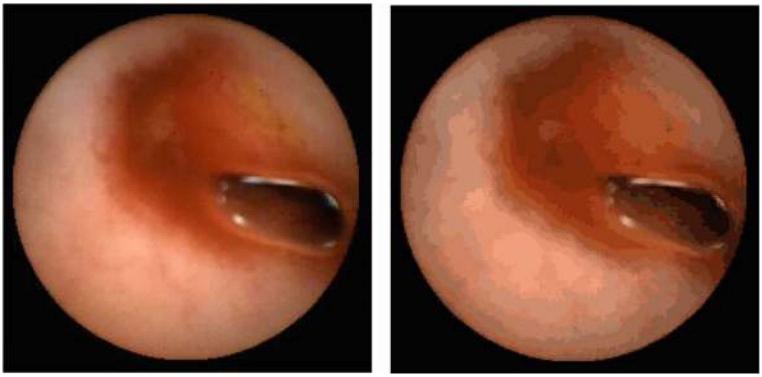


Fig. 9 Color palette reduced to 24 by using minimum variance quantization for accuracy and complexity balance

studied in [32]. The input layer of the CNN architecture fetches the images in RGB color space after reduction in the color palette using *minimum variance quantization*. Number of colors chosen to be 24 has resulted in good balance between the accuracy and computational complexity, also studied in [39]. Figure 9 illustrates the increased contrast of the transformed crisp image. Color reduction helps in better image segmentation by simplifying the obscure nature of internal body image which is originally red in color. It also enhances the computational efficiency of the proposed model through this pre-processing.

6.4 Performance metrics

The prominent metrics for comparative analysis have been used which include, accuracy, sensitivity, specificity, precision, recall and F1-score. Their formulae have been given in the Table 6.

6.5 Data analysis

Machine learning model selection along with its performance depends on the underlying data distribution [18]. Therefore the data analysis is imperative before model design and selection. Two experiments have been performed to study the data distribution of normal and sick (bleeding) images (1) average red color (R) intensity comparison in sick versus healthy images (2) RGB values analysis in healthy and sick images to see the dispersion and separability of two classes in 2-d plots. For sake of clarity in the plots, less than 10% images

Table 6 Metric definitions

Metric	Formula
Accuracy	(tp+tn)/(tp+fp+fn+tn)
Sensitivity	tp/(tp+fn)
Specificity	tn/(tn+fp)
Precision (P)	tp/(tp+fp)
Recall (R)	tp/(tp+fn)
F1	2*R*P/(R+P)

have been chosen randomly out of 3,895 images in the underlying public dataset obtained from [5]).

In Fig. 10, the average red color in the sick and healthy classes has been sorted and plotted for 120 images from each category. Although all gastral images are reddish brown in color, but the sick images are more saturated with redness. In Fig. 11, various intensity plots of 300 images have been illustrated. These four figures show the following: (a) RGB intensity distributions of 120 bleeding (sick) and 180 normal (healthy) images for comparison of dispersion in red color. Red is more dispersed on the left half (sick class) which means higher range of red values. Other three plots (b-d) show for all images, the R versus G; R versus B and B versus G plots for these 120 sick + 180 healthy images. In (b-d) images show that there is tremendous amount of overlap and a simple linear regression may not be sufficient for the bleeding classification. Thus, there is a need for powerful non-linear learning system (such as CNN) to discern among sick and healthy classes.

6.6 CNN parameters and optimizers

The output layer of the proposed model is a binary output, which is either bleeding or normal. Hidden layers act as intermediate layers with the freedom to tune the parameters to achieve best possible decision function for classification. The training cycle included 4 epochs to yield the required accuracy. Epochs were initially set to 20 but decremented to 4 gradually for the similar accuracy but faster computation during empirical analysis (Figs. 6 and 7). The validation frequency was set to 3 iterations with Infinite Patience and hardware included single CPU with constant learning rate schedule and 0.01 learning rate. There are three 2-D convolutional layers in the proposed network. The initial layers are relatively basic to learn low level features while maturing up towards the end of the network. Layers enlarge in the receptive field size capability to learn subtle features from left to right in CNN. Figure 12 shows the outputs for convolution layers (1-3) and fully connected layer. Figure 12 has been drawn using MATLAB function *deepDreamImage* to visualise network

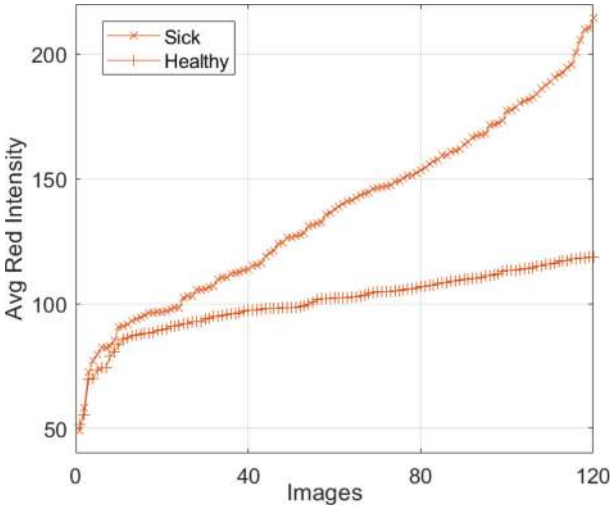


Fig. 10 Sorted average red pixel intensities for 120 normal and sick images. The upper line is redness in sick images which has relatively higher values compared to normal images

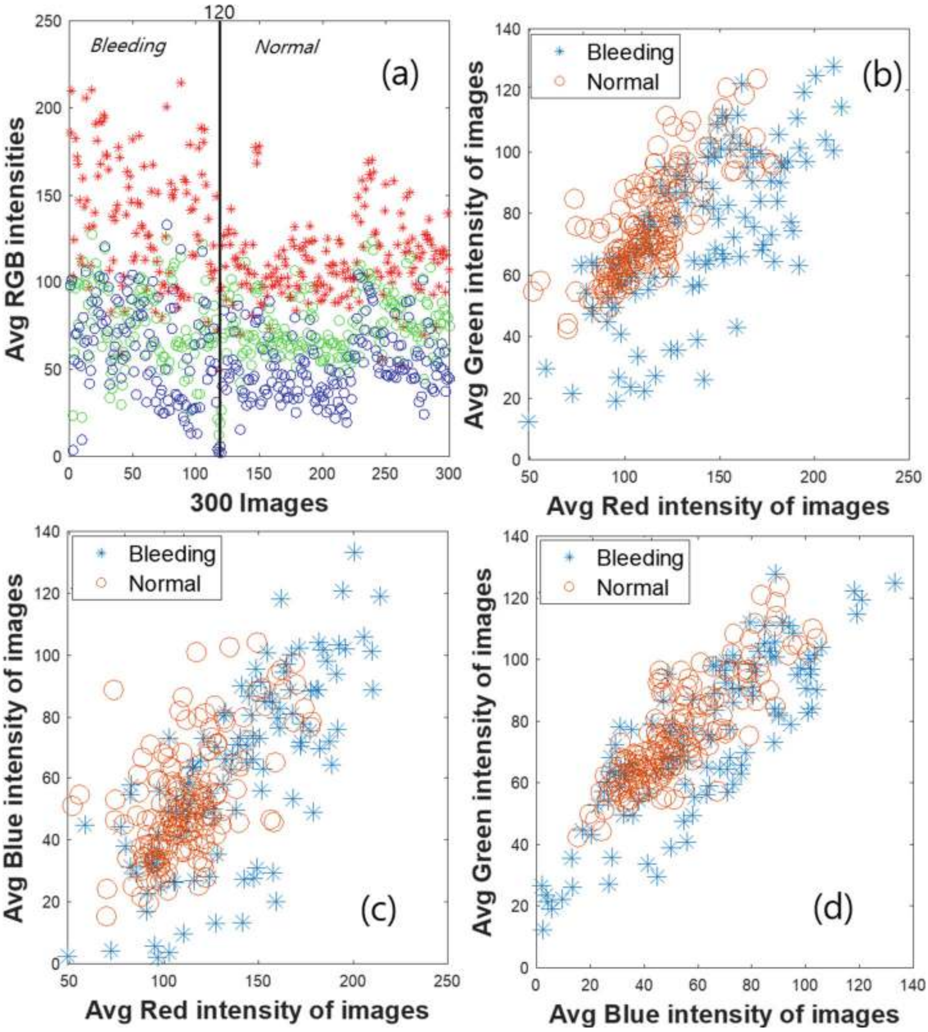
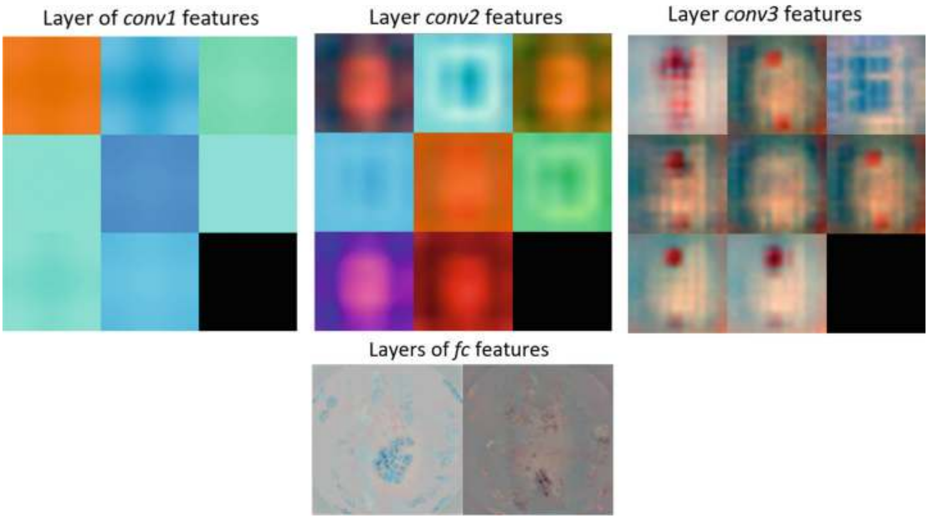


Fig. 11 Only 300 images out of 3,895 (from public dataset [5]) has been chosen for clarity of illustration (a) RGB intensity of 300 images (180:120 for healthy:sick cases) (b) Red vs Green average intensities (c) Red vs Blue avg. intensities (d) Blue vs. Green intensities plotted of all 300 images. It is clear from (b-d) that none of the colour intensities are easily separable for healthy and sick cases

features. It shows semantic learning prowess in subsequent CNN layers as the data goes through. Refer Table (3) for variables for layers corresponding to entries {2, 6, 10, 13}. Thus the network learns progressively about colors and edges at different angles for complex features. It is clear from the Fig. (12) that by the end, the model is able to classify bleeding versus normal frames due to their striking differences in the contrast after the transformation through CNN layers.

Three optimizer options were tried for CNN ensemble (sgdm, adam and rmsprop) to compare the performance of proposed model as shown in Table 7. It can be observed that for better overall performance, *sgdm* (Stochastic Gradient Descent with Momentum) solver


Layer of *conv1* features

Layer *conv2* features

Layer *conv3* features

Layers of *fc* features

Fig. 12 CNN features from 3 convolution layers and fully connected layer

should be used for CNN, which is related to the works [3]. Thus *sgdm* optimizer is good choice over *rmsprop* (Root Mean Square Propagation) and *adam* (Adam Stochastic Gradient Descent) for this dataset to classify bleeding. Now, keeping the solver fixed to *sgdm*, the dataset ratio in training phase has been varied by decreasing the normal images randomly to make the data relatively balanced for binary classes (60:35, 60:40 and 50:50).

6.7 Results analysis for class ratios

Various data ratios of healthy and sick must be explored to analyse the performance due to data distribution disparity (in healthy and sick classes) as studied in Section 6.5. All 3,895 images have been considered for training and testing, with random selection of 85:15 for training:testing ratio. Afterwards, 10-trails average has been used to report the performance

Table 7 Effects in performance for variations in **a** training data ratios (healthy:sick) and **b** various optimizer methods for proposed method

	Optimizer	Healthy:Sick	Sen	Spc	P	R	F1	Acc
1	SGDM	65:35	0.92	0.96	0.91	0.94	0.92	**0.95**
2		60:40	0.94	0.94	0.92	0.92	0.92	**0.94**
3		50:50	0.95	0.92	0.90	0.95	0.92	**0.93**
4	ADAM	65:35	0.88	0.94	0.90	0.89	0.89	0.92
5		60:40	0.90	0.92	0.91	0.91	0.91	0.91
6		50:50	0.91	0.88	0.91	0.91	0.91	0.89
7	RMSPROP	65:35	0.87	0.92	0.89	0.87	0.88	0.90
8		60:40	0.90	0.91	0.88	0.89	0.88	0.91
9		50:50	0.91	0.88	0.92	0.89	0.90	0.89

Training:Testing ratio has been set to 85:15. Average of 10 trials have been reported

Bold entries try to emphasize on the proposed model results

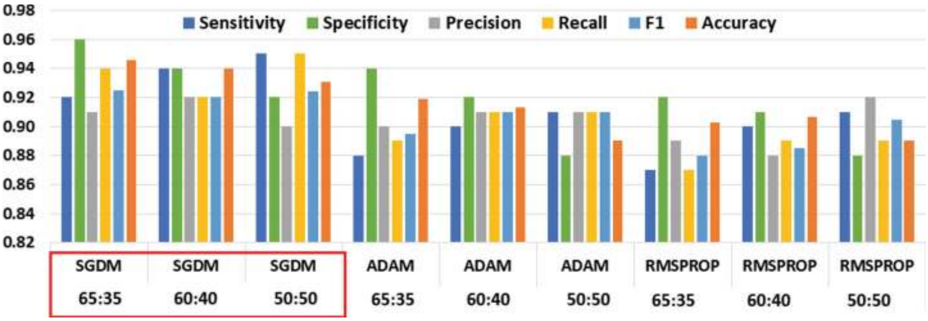


Fig. 13 SGDM optimizer in CNN ensemble gives better results with accuracy over various ratios of normal:bleeding instances. Best results were found with SGDM and 65:35 data ratios of healthy:sick images during training

results. In the training phase, normal:bleeding class ratios has been altered (65:35, 60:40 and then 50:50) for each optimizer and performance has been recorded for analysis. The sick examples were kept intact and normal image examples were decreased gradually. It has been observed that by reducing the data for normal examples, the decision boundary gets biased towards sick images resulting in higher sensitivity. On the other hand, specificity and overall accuracy decreases as the data balances in class ratio by successive reductions. Thus in the underlying case study, more instances with normal examples are beneficial for global predictions for both sick and healthy.

Visual representation of the Table 7 has been plotted as Fig. 13. It clearly demonstrates the victory of *sgdm* optimizer in CNN over various data ratios of normal and bleeding image sets. Although the best *sgdm* results were obtained by using 65:35 ratio (for normal and bleeding images). The results from proposed method have been compared with other state-of-art algorithms as shown in Table 8. For consistent comparison, same dataset (of 3,895 images from [5]) has been used with 85:15 for train:test ratio and average over 10 trials. It can be observed that proposed CNN model can achieve accuracy as high as 0.95 outperforming other modern methods. Also, an optimal split ratio for training the model depends on underlying data distribution i.e. ratio of healthy to sick images. With consideration of the system used for experimentation, data analysis results might be useful in the future for

Table 8 Performance comparison with state-of-art techniques (using public dataset [5] of 3,895 images) with 85:15 ratio of training:testing

Sr	Method	Accuracy
1	Statistical features and region growing [39]	0.87
2	Computerized method[7]	0.91
3	SVM [43]	0.93
4	MST and SVM[45]	0.93
5	Suspected Blood Indicator tool (SBI) [27]	0.73
6	U-Net [5]	0.94
7	SegNet [9]	0.93
8	**Proposed CNN framework**	**0.95**

Average of 10 trials have been recorded

Bold entries try to emphasize on the proposed model results researchers to improve their model. Thus CNN ensemble strives to find out the best possible combination of attributes required to classify bleeding symptoms in CE images after setting up the layer options, SGDM optimizer and color palette reduction using minimum variance quantization technique.

6.8 Real dataset

Real endoscopy dataset have been obtained from a known gastroenterologist with manual annotation of the binary labels (normal or bleeding). We considered a subset of this endoscopic video for testing, stretching up to 42 minutes which included significant bleeding frames. The video to image conversion has been done using MATLAB using function *readFrame* with the rate of 2 frames per second. It yielded 5,000 images (normal:bleeding ratio is 4000:1000). Sample frames from the real dataset video has been shown in the Fig. 14. The previous trained model on the public dataset [5] with augmentations has been used to test this real world video dataset. Training rules were defined in Tables 3, 4, 5. Prior to the testing real dataset on the proposed algorithm, basic pre-processing has been performed such as (a) histogram equalization for contrast enhancement (b) image resizing to match the training images (c) color palette reduction using minimum variance quantization. Proposed ensemble was able to achieve 0.93 and 0.91 accuracy with and without preprocessing on the real video dataset while outperforming state-of-art algorithms.

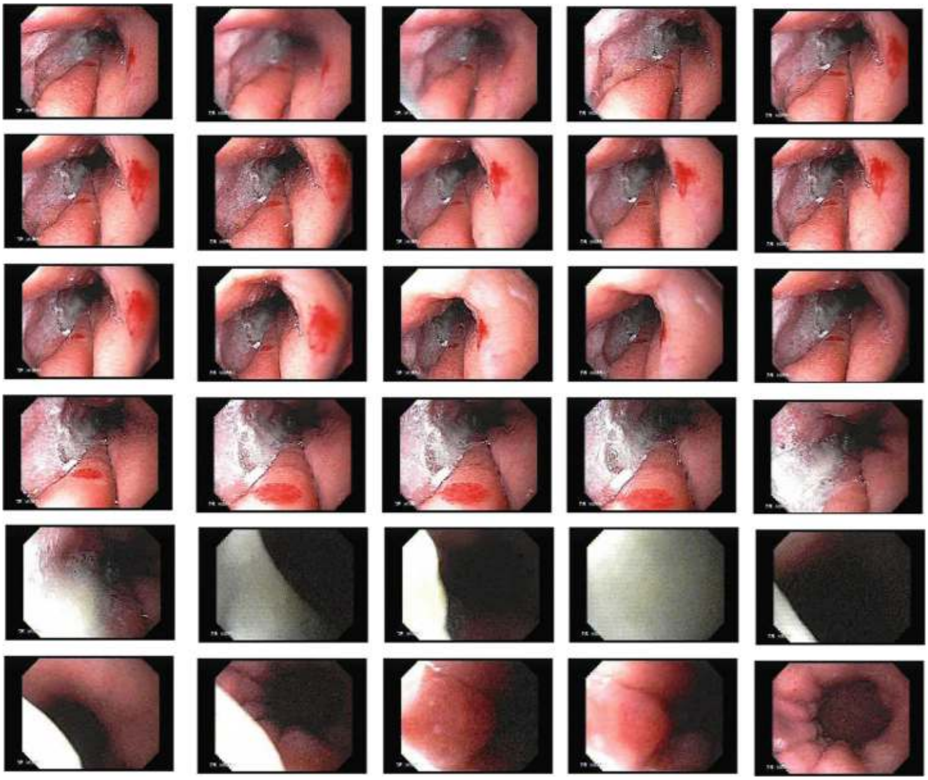


Fig. 14 Bleeding classification on the real dataset endoscopy video

Thus convolutional neural networks ensemble made from data augmentation with minimum variance quantization with a specific setting of layers and optimizer makes it well suitable for building highly dependable image classification system for internal organ images.

7 Conclusion

This paper has defined a CNN augmented ensemble architecture to diagnose bleeding symptoms from Capsule Endoscopy (CE) images. Data distribution and RGB pixel intensities has been studied in detail to demonstrate the complexity of the classification problem. Copies of the image data have been generated using synthetic augmentation by introducing rotation, blur, illuminance and noise. Afterwards, Color palette reduction has been used using minimum variance quantization, studied in [32] to get 24 colors in the image for optimal balance of accuracy and time-space complexity. The experimental analysis of the proposed technique has been performed on both the public and real dataset and compared with other state-of-art methods.

The limitation of the proposed technique is that if images quality is poor, then the detection accuracy might degrade. So the assumption is a good quality image dataset for training and testing. Image size should also be optimized to balance the trade off with meta-heuristic performance accuracy and computational time on the underlying machine. For high volume video data processing, GPU workstation is also required which might be expensive. Nevertheless, the proposed CNN ensemble is well suitable for building highly dependable image classification system for any similar type of image classification application. Future plan is to incorporate longer and obscure videos for model training and design automatic video annotation and segmentation system by utilizing cross-modal features of image and collateral text obtained from expert gastroenterologists.

Acknowledgments This research work has been sponsored by the Seed Money project grant at Thapar Institute of Engineering and Technology Patiala India under Grant TU/DORSP. Authors are thankful to Dr. Sunil Arya Gastroenterologist, Patiala India for providing the dataset and technical recommendations.

Funding Information Open access funding provided by Aalto University.

References

1. Abouelenien M, Yuan X, Giritharan B, Liu J, Tang S (2013) Cluster-based sampling and ensemble for bleeding detection in capsule endoscopy videos. Am J Sci Eng 2(1):24–32
2. Anjomshoae S, Främling K, Najjar A (2019) Explanations of black-box model predictions by contextual importance and utility. In: International Workshop on Explainable, Transparent Autonomous Agents and Multi-Agent Systems. Springer, New York, pp 95–109

3. Bottou L (2010) Large-scale machine learning with stochastic gradient descent. In: proceedings of COMPSTAT'2010. Springer, New York, pp 177–186
4. Carpi F, Shaheed H (2013) Grand challenges in magnetic capsule endoscopy. Expert Rev Med Devices 10(4):433–436
5. Coelho P, Pereira A, Salgado M, Cunha A (2018) A deep learning approach for red lesions detection in video capsule endoscopies. In: International Conference Image Analysis and Recognition. Springer, New York, pp 553–561
6. Du W, Rao N, Liu D, Jiang H, Luo C, Li Z, Gan T, Zeng B (2019) Review on the applications of deep learning in the analysis of gastrointestinal endoscopy images. IEEE Access 7:142053–142069
7. Figueiredo IN, Kumar S, Leal C, Figueiredo PN (2013) Computer-assisted bleeding detection in wireless capsule endoscopy images. Comput Methods Biomec Biomed Eng Imaging Visualization 1(4):198–210
8. Fuentes Álvarez JR Deep learning in hierarchical neural networks applied as pattern classifiers for massive information systems
9. Ghosh T, Li L, Chakareski J (2018) Effective deep learning for semantic segmentation based bleeding zone detection in capsule endoscopy images. In: 2018 25th IEEE International Conference on Image Processing (ICIP). IEEE, Los Alamitos, pp 3034–3038
10. Giritharan B, Yuan X, Liu J, Buckles B, Oh J, Tang SJ (2008) Bleeding detection from capsule endoscopy videos. In: Engineering in Medicine and Biology Society, 2008. EMBS 2008, 30th Annual International Conference of the IEEE. IEEE, Los Alamitos, pp 4780–4783
11. Gu J, Wang Z, Kuen J, Ma L, Shahroudy A, Shuai B, Liu T, Wang X, Wang G, Cai J et al (2018) Recent advances in convolutional neural networks. Pattern Recogn 77:354–377
12. Hajabdollahi M, Esfandiarpoor R, Najarian K, Karimi N, Samavi S, Soroushmehr SR (2019) Low complexity cnn structure for automatic bleeding zone detection in wireless capsule endoscopy imaging. In: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, Los Alamitos, pp 7227–7230
13. Hajabdollahi M, Esfandiarpoor R, Soroushmehr S, Karimi N, Samavi S, Najarian K (2018) Segmentation of bleeding regions in wireless capsule endoscopy images an approach for inside capsule video summarization,arXiv:1802.07788
14. He K, Zhang X, Ren S, Sun J (2015) Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In: proceedings of the IEEE international conference on computer vision, pp 1026–1034
15. Iakovidis DK, Koulaouzidis A (2015) Software for enhanced video capsule endoscopy: challenges for essential progress. Nat Rev Gastr Hepat 12(3):172
16. Jia X, Meng MQH (2016) A deep convolutional neural network for bleeding detection in wireless capsule endoscopy images. In: Engineering in Medicine and Biology Society (EMBC), 2016 IEEE 38th Annual International Conference of the. IEEE, Los Alamitos, pp 639–642
17. Jia X, Meng MQH (2017) Gastrointestinal bleeding detection in wireless capsule endoscopy images using handcrafted and cnn features. In: 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, Los Alamitos, pp 3154–3157
18. Kaur H, Pannu HS, Malhi AK (2019) A systematic review on imbalanced data challenges in machine learning: Applications and solutions. ACM Comput Surveys (CSUR) 52(4):79
19. Kim HG, Choi Y, Ro YM (2017) Modality-bridge transfer learning for medical image classification. In: 2017 10th International Congress on Image and Signal Processing BioMedical Engineering and Informatics (CISP-BMEI). IEEE, Los Alamitos, pp 1–5
20. Lawrence S, Giles CL, Tsoi AC, Back AD (1997) Face recognition: a convolutional neural-network approach. IEEE trans Neural Netw 8(1):98–113
21. LeCun Y, Bengio Y et al (1995) Convolutional networks for images, speech, and time series. Handbook Brain Theory Neural Netw 3361(10):1995
22. LeCun Y, bengio Y, Hinton G (2015) Deep learning. Nature 521(7553):436–444
23. Li B, Meng MQH (2009) Computer-aided detection of bleeding regions for capsule endoscopy images. IEEE Trans Biomed Eng 56(4):1032–1039
24. Li P, Li Z, Gao F, Wan L, Yu J (2017) Convolutional neural networks for intestinal hemorrhage detection in wireless capsule endoscopy images. In: 2017 IEEE International Conference on Multimedia and Expo (ICME). IEEE, Los Alamitos, pp 1518–1523
25. Li S, Zhang J, Ruan C, Zhang Y (2019) Multi-stage attention-unet for wireless capsule endoscopy image bleeding area segmentation. In: 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE, Los Alamitos, pp 818–825
26. Li X, Zhang H, Zhang X, Liu H, Xie G (2017) Exploring transfer learning for gastrointestinal bleeding detection on small-size imbalanced endoscopy images. In: Engineering in Medicine and Biology Society (EMBC), 2017 39th Annual International Conference of the IEEE. IEEE, Los Alamitos, pp 1994–1997

27. Liangpunsakul S, Mays L, Rex DK (2003) Performance of given suspected blood indicator. Amer J Gastro 98(12):2676–2678
28. Liu J, Yuan X (2009) Obscure bleeding detection in endoscopy images using support vector machines. Optim Eng 10(2):289–299
29. Liu S, Wang Y, Yang X, Lei B, Liu L, Li SX, Ni D, Wang T (2019) Deep learning in medical ultrasound analysis: a review. Engineering
30. Liu W, Zhang M, Luo Z, Cai Y (2017) An ensemble deep learning method for vehicle type classification on visual traffic surveillance sensors. IEEE Access 5:24417–24425
31. Lu Y, Yi S, Zeng N, Liu Y, Zhang Y (2017) Identification of rice diseases using deep convolutional neural networks. Neurocomputing 267:378–384
32. Lucchese L, Mitra S (1999) An algorithm for fast segmentation of color images. In: Multimedia Communications. Springer, New York, pp 110–119
33. Malhi A, Kampik T, Pannu HS, Madhikermi M, Framling K (2019) Explaining machine learning based classifications of in-vivo gastral images. In: 2019 IEEE International Conference on Digital Image Computing: Techniques and Applications (DICTA 2019). IEEE, Los Alamitos, pp 1–7
34. Pan G, Yan G, Qiu X, Cui J (2011) Bleeding detection in wireless capsule endoscopy based on probabilistic neural network. J Med Syst 35(6):1477–1484
35. Perez L, Wang J (2017) The effectiveness of data augmentation in image classification using deep learning, arXiv:1712.04621
36. Radenović F, Tolias G, Chum O (2018) Fine-tuning cnn image retrieval with no human annotation. IEEE transactions on pattern analysis and machine intelligence
37. Raginsky M, Jafarpour S, Harmany ZT, Marcia RF, Willett RM, Calderbank R (2011) Performance bounds for expander-based compressed sensing in poisson noise. IEEE Trans Signal Process 59(9):4139–4153
38. Razzak MI, Naz S, Zaib A (2018) Deep learning for medical image processing: Overview, challenges and the future. In: Classification in BioApps. Springer, New York, pp 323–350
39. Sainju S, Bui FM, Wahid KA (2014) Automated bleeding detection in capsule endoscopy videos using statistical features and region growing. J Med Syst 38(4):25
40. Salamon J, Bello JP (2017) Deep convolutional neural networks and data augmentation for environmental sound classification. IEEE Signal Process Letters 24(3):279–283
41. Seguí S, Drozdzal M, Pascual G, Radeva P, Malagelada C, Azpiroz F, Vitrià J (2016) Generic feature learning for wireless capsule endoscopy analysis. Comput Bio Med 79:163–172
42. Sekuboyina AK, Devarakonda ST, Seelamantula CS (2017) A convolutional neural network approach for abnormality detection in wireless capsule endoscopy. In: 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017). IEEE, Los Alamitos, pp 1057–1060
43. Usman MA, Satrya GB, Usman MR, Shin SY (2016) Detection of small colon bleeding in wireless capsule endoscopy videos. Comput Med Imaging Graph 54:16–26
44. Xing X, Jia X, Meng MH (2018) Bleeding detection in wireless capsule endoscopy image video using superpixel-color histogram and a subspace knn classifier. In: 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, Los Alamitos, pp 1–4
45. Xiong Y, Zhu Y, Pang Z, Ma Y, Chen D, Wang X (2015) Bleeding detection in wireless capsule endoscopy based on mst clustering and svm. In: 2015 IEEE Workshop on Signal Processing Systems (SiPS). IEEE, Los Alamitos, pp 1–4
46. Yuan Y, Meng MQH (2017) Deep learning for polyp recognition in wireless capsule endoscopy images. Med physics 44(4):1379–1389
47. Zheng H, Yang Z, Liu W, Liang J, Li Y (2015) Improving deep neural networks using softplus units. In: 2015 International Joint Conference on Neural Networks (IJCNN). IEEE, Los Alamitos, pp 1–4
48. Zhou Z-H (2015) Ensemble learning. Encyclopedia of biometrics, 411–416