Received January 24, 2021, accepted February 13, 2021, date of publication February 23, 2021, date of current version March 4, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3061592

Wireless Capsule Endoscopy Bleeding Images Classification Using CNN Based Model

FURQAN RUSTAM[1], MUHAMMAD ABUBAKAR SIDDIQUE[1], HAFEEZ UR REHMAN SIDDIQUI[1], SALEEM ULLAH[1], ARIF MEHMOOD[2], IMRAN ASHRAF[3], AND GYU SANG CHOI[3]

[1]Department of Computer Science, Khwaja Fareed University of Engineering and Information Technology, Rahim Yar Khan 64200, Pakistan
[2]Department of Computer Science & IT, The Islamia University of Bahawalpur, Punjab 63100, Pakistan
[3]Department of Information and Communication Engineering, Yeungnam University, Gyeongbuk 38541, South Korea

Corresponding authors: Imran Ashraf (ashrafimran@live.com) and Gyu Sang Choi (castchoi@ynu.ac.kr)

This work was supported in part by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education under Grant NRF-2019R1A2C1006159, and in part by the Ministry of Science and ICT (MSIT), South Korea, through the Information Technology Research Center (ITRC) Support Program, supervised by the Institute for Information & Communications Technology Promotion (IITP) under Grant IITP-2020-2016-0-00313.

ABSTRACT Wireless capsule endoscopy (WCE) is an efficient tool to investigate gastrointestinal tract disorders and perform painless imaging of the intestine. Despite that, several concerns make its wide applicability and adaptation challenging like efficacy, tolerance, safety, and performance. Besides, automatic analysis of the WCE provided dataset is of great importance for detecting abnormalities. Imaging of the patient's digestive tract through WCE produces a large dataset that requires a substantial amount of time and a special skill set from a medical practitioner for analysis. Several computer-aided and vision-based solutions have been proposed to resolve these issues, yet, they do not provide the desired level of accuracy and further improvements are still needed. The current study aims to devise a system that can perform the task of automatic analysis of WCE images to identify abnormalities and assist practitioners for robust diagnosis. This study adopts a deep neural network approach and proposes a model name BIR (bleedy image recognizer) that combines the MobileNet with a custom-built convolutional neural network (CNN) model to classify WCE bleedy images. BIR uses the MobileNet model for initial-level computation for its lower computation power requirement and subsequently the output is fed to the CNN for further processing. A dataset of 1650 WCE images is used to train and test the model using the measures of accuracy, precision, recall, F1 score, and Cohen's kappa to evaluate the performance of the BIR. Results indicate the promising outcomes with achieved accuracy, precision, recall, F1 score, and Cohen's kappa of 0.993, 1.000, 0.994, 0.997, and 0.995 respectively. The accuracy of the BIR model is 0.978 with the Google collected WCE image dataset which is better than the state-of-art approaches.

INDEX TERMS Wireless capsule endoscopy, deep learning, computer vision, gastrointestinal tract infection, classification, convolutional neural networks.

I. INTRODUCTION

Presently gastrointestinal tract infections such as ulcer polyps, bleeding, cancer, and Crohn's have become fairly frequent whereas ulcers and bleeding are commonplace diseases. A survey by the world health organization (WHO) indicates the leading causes of death as lung cancer (1.1 million deaths), stomach cancer (765 000 deaths), colon and rectum cancer (525 000 deaths) liver cancer, (505 000 deaths), and breast cancer (385 000 deaths) [1]. According to a study [2] in America, 135,430 new gastrointestinal tract infection diseases occurred since 2017, and approximately 200,000 new cases are being registered every year since 2011 worldwide. This gastrointestinal tract infection can be rectified if detected and diagnosed at an early stage [3]. For the detection of gastrointestinal tract infection, wireless capsule endoscopy is the preferred technology used by expert physicians.

Wireless capsule endoscopy (WCE) is a noninvasive technique considered essentially to provide diagnostic imaging of the small intestine [4]. In the WCE technique, the patient gulps a capsule containing a camera-embedded device that passes through the gastrointestinal tract and captures images and transmits these captured images to an

The associate editor coordinating the review of this manuscript and approving it for publication was Yuan Zhuang.

external receiver. Incepted in 1989 by a research group, the WCE technology went under many improvements over time by various companies to become more accurate and efficient, as we find it today [2]. This technology replaced the conventional endoscopy method and made the diagnosis easier for both patient and examiner because traditional endoscopy is very painful and distressful for patients. During the WCE process, a camera captures images continuously for 8 hours inside the patient and transmits these images to the receiver, approximately 60000 images [5]. A physician examines these images to identify which frames/images contain the infection/disease. Because of the large number of images, it requires a substantial amount of time for a medical expert to examine these images, which may result in an extra burden for the practitioner and may lead to the poor diagnosis of the infected region of the intestine. So, the identification of images that contain infected regions using various statistical and machine learning algorithms has been an attractive research area during the last decade or so. An automatic identification system performs analysis on the large volume of images and finds the infected frames so it is easy for the physician to examine only those frames that contain visual contents of the affected area and start appropriate remedial actions, in good time.

Machine learning techniques have shown significant competence to perform such automatic tasks and hold the potential to make the medical domain advanced and more accurate. Many researchers work on automatic systems for gastrointestinal tract infection detection [6] and gastrointestinal tract infection classification [7] using artificial intelligence. Predominantly, the researchers consider a specific disease in automatic detection systems, such as ulcer [8], polyps [9] and bleeding [10] in their approaches because it makes efficient and specialized for a specific task. Along the same direction, this study aims at classifying gastrointestinal tract infection (bleeding) images using the machine learning paradigm.

In this study, an artificial intelligence model, named BIR (bleedy image recognition), is developed to classify infected (bleeding) images. First, an image dataset is obtained by performing WCE of 33 patients. Afterward, clean and infected images are extracted separately, after the examination is done by a medical expert on a large image dataset of 33 patients, to make an infected image dataset and normal image dataset.

Although the number of images in a normal image dataset is large enough to train the model, the same is not true for its counterpart, i.e., infected image dataset (See section III). To overcome the imbalance of datasets, data augmentation techniques are used to increase the number of images for the infected class by flipping each infected image. After flipping each image from an infected set we get twice the size of infected image data. BIR model is then trained on these infected and normal images and subsequently, test data is passed to the trained model to evaluate the performance of models. We evaluate the BIR performance in terms of accuracy precision, recall, F1 score, and Cohen's kappa. This study makes the following contributions.

- A large dataset of WCE images for 33 patients is gathered using $PillCam^{TM}\,SB_3$. In addition, a validation dataset is collected using the Google search engine that contains 22 normal and 25 infected images.
- A novel model called BIR is proposed that can classify WCE images into normal and infected images to help medical experts in a timely and efficient diagnosis of gastrointestinal tract infection. BIR combines MobileNet and convolutional neural networks to enhance the efficiency of image classification.
- The impact of a balanced and imbalanced dataset is investigated as well with the proposed approach. Data augmentation is done to make the dataset balanced for this purpose.
- The performance of the proposed model is analyzed against five machine learning algorithms like support vector machine (SVM), logistic regression (LR), random forest (RF), extra tree classifier (ETC), and decision tree classifier (DTC), as well as, MobileNet, and convolutional neural networks. Furthermore, the performance is compared with four state-of-the-art classification approaches for bleedy image recognition.

The rest of the paper is organized as follows. Section II discusses research works related to the current study. Section III contains the dataset description while the proposed methodology is presented in Section IV. Section VII analyzes experiment results, whereas, the discussion is performed in section VIII. In the end, the conclusions are given in Section IX.

II. RELATED WORK

This study focuses on the automatic prediction of infected images obtained using wireless capsule endoscopy. The detected infected images make it easy for the medical expert to perform a timely diagnosis and start the treatment readily. Significant research in this domain makes automatic detection of infection from a large set of image data which saves both time and effort of medical experts, as well as, increases diagnosis efficiency. Automatic detection of infected images has been an active area of research recently and a large variety of work can be found in the literature. Such studies make use of three kinds of techniques in essence: classic machine learning algorithms like SVM, k-NN, etc., proposed/modified models like modified PCA, and deep learning techniques like convolutional neural networks (CNN), multilayered perceptron, etc.

The study [11] used the support vector machine (SVM) with a radial basis function for the classification of bleedy and normal images and achieves a 0.978 accuracy. The study [12] used the Contourlet transform and local binary pattern (LBP) to classify bleedy and non-bleedy frames. They achieve an accuracy of 0.963 through the k-Nearest Neighbor (k-NN) classifier. The study [13] designed an algorithm using a joint diagonalization principal component analysis (PCA)

combined with the color coherence vector (CCV) where no iterations, approximations, and inverting procedures are required. This method overcomes the problem of PCA and the 'curse of dimensionality' of the original asymptotic PCA.

The study [11] proposed an approach for automatic detection of gastrointestinal tract infection of various lesions and abnormalities. The color features and texture features based approach is used for bleeding detection in the study. Color information is used to extract the lesion captured in the image. Color information is an important feature for the initial detection of bleeding images while texture can be used later to extract lesions in the detected images. The proposed approach can distinguish finely between borderline cases. Another study [14] proposed an algorithm for segmentation of bleedy WCE frame using HIS intensity channel for texture features extraction and use support vector machine for classification of frames. They calculated a histogram for the uniform LBP with 8×8 regions in terms of mean, variance, entropy, kurtosis, Skewness, and energy. The study [15] proposed an approach for the classification of abnormalities such as bleeding and ulcer and introduced an ANOVA using the F-statics measure and the sequential floating forward search to classify frames using color and texture features.

Recently, several deep learning approaches have been proposed and employed for the automatic detection of infection lesions using WCE images. For example, the study [16] used chrominance-moments-based texture to detect the bleedy region in WCE video. This study deployed a multilayered perceptron neural network to classify normal regions and bleeding regions in WCE images. Similarly, [17] used the WCE images to detect Ulcer using the deep convolutional neural networks. They proposed the HAnet architecture using ResNet-34 as the base model. A dataset containing 1,416 independent videos is collected during the WCE procedure and frames from these videos are later used in the study. HAnet achieved a 92.05% accuracy score and its sensitivity and specificity are 91.64% and 92.42%, respectively. In the same vein, the author makes use of a CNN model in [18] to perform the diagnosis of middle ear diseases. They use the CNN model to diagnosis the tympanic membrane. A CNN with six-layers is developed and a class activation map is used as the feature extraction method. CNN provides a significant performance in achieving 97.9% accuracy. These approaches indicate the potential of deep learning approaches to perform classification in medical imaging.

Similarly the CNN model, MobileNet model has also been used for various medical domain-related tasks, although not with WCE images. For example, the study [19] used MobileNet to detect skin cancer disease. MobileNet is applied to a large dataset containing 1,280,000 images for detecting skin cancer. MobileNet achieved 83.10% accuracy on the dataset with seven classes. Another study [20] performed image classification using the MobileNet on the COVID-19 image dataset to prove the efficiency of the MobileNet in effectively distinguishing the Covid-19 cases from viral and bacterial pneumonia cases. Another study [21] used MobileNet for the lung nodules malignancy classification with other machine learning and deep learning models and showed the significant performance of the MobileNet model on the images dataset.

The above-cited research works are limited by several factors. Firstly, many among them do not specifically utilize WCE images for detecting infected images to diagnose gastrointestinal tract infections. For example, MobileNet is used for skin cancer, COVID-19, and lung ailment detection but never used with WCE images for intestine infection. To the best of our knowledge, this study is the first to make use of MobileNet on WCE images to detect bleedy images. Secondly, despite that many machine learning, as well as, modified machine learning models show promising efficiency, their performance can be improved further, as the results of deep learning approaches suggest. CNN has shown significant improvement over other deep learning models with WCE images and this study intends to improve its performance further by devising a combination of pooling, normalization, and dropout layers.

III. DATASET

WCE is one of the promising and forthcoming technology today for examining gastrointestinal tract infections because the WCE process is easy and less distressing than those of traditional endoscopy. In WCE patient gulps a camera-embedded capsule that is connected with an external receiver. This capsule passes through the throat, esophagus, stomach, small intestine, large intestine, rectum, and anus respectively. This process requires approximately 8 hours and captures as much as 60,000 images. The captured images can be later used for infection detection and diagnosis by medical experts. The dataset containing the images proves very helpful for the medical community to study various diseases. This study uses a dataset that contains WCE images of 33 patients.

According to gastroenterologist observations 14 patients have moderate bleeding, 11 have mild bleeding while 9 patients have severe bleeding. The dataset contains a total of 5,000 images from 33 patients. Based on the presence of bleeding, WCE images are classified into normal and bleedy images. A sample of the normal and bleedy(infected) image is shown in Figure 1.

Normal images are those that do not contain the bleeding. Out of the total of 5000 images, there are 450 bleedy images and 4550 are normal as shown in Table 1. Some of these bleedy images are also from the early stage of infection where the bleeding is not very clear.

The dataset is collected by gastroenterologists at Sheikh Zayed Hospital Rahim Yar Khan, Pakistan. The *PillCam*^{TM} *SB*3 capsule has been used in this study to collect the image dataset [22], [23]. This capsule captures images at a frame rate of 2 to 6 frames per second (fps) based on capsule speed. Figure 2 shows the interface used to collect the WCE image dataset from 33 patients.

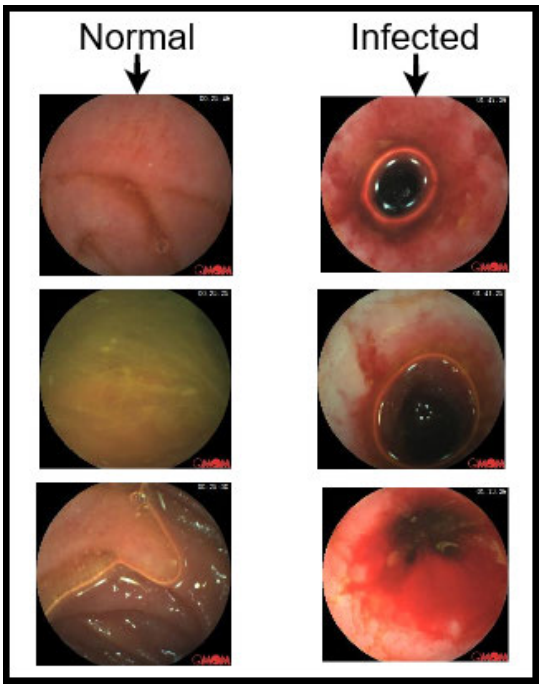


FIGURE 1. Sample of data for both normal and infected class.

TABLE 1. Dataset count according to patients severity.

Category	No. of Patients	Bleeding Images	Normal Images
Total	33	450	4550
Mild Bleeding	11	150	1550
Moderate Bleeding	14	150	1550
Severe Bleeding	9	150	1450

FIGURE 2. Interface for data accusation.

IV. PROPOSED MODEL

This study uses a deep learning approach to develop the proposed model. Deep learning is a sub-class of artificial intelligence that uses artificial neural networks (ANNs) to train learning models. Several types of deep learning networks have been developed and deployed, but the two most commonly used neural networks are recurrent neural networks (RNN) [24] and convolutional neural networks (CNN) [25]. Predominantly, the CNN model has been used in a variety of computer vision tasks like object detection, image classification, etc. and proven superior to other deep learning models [4], [26], [27]. Figure 3 shows the approach of this study for bleedy image classification.

Even though the use of the traditional machine learning models such as random forest, support vector machine, logistic regression, etc., in medical image classification, began long ago but the disadvantage of these traditional models is that their performance is far from the practical standard, and the improvement and development have been quite slow in recent years. The feature extraction and feature selection for traditional models are very time consuming and vary with respect to the desired objective [28]. On the other hand, CNN based model has shown significant performance in image classification tasks [29]. The achieved performances by CNN in medical research are human competitive. For example, the CNN based architecture CheXNet model with 121 layers perform as better as four average radiologists do on a 100000 chest X-ray image dataset [28]. Similarly, the CNN model used to classify 108,309 Optical coherence tomography images achieves the weighted average error equal to the average performance of 6 human experts [30]. As a consequence of such performances, this study considers the use of CNN based model for classifying the WCE images. For this purpose, the CNN model is combined with the MobileNet model to enhance the classification performance.

A. MobileNet

MobileNet is a lightweight deep neural network model with fewer parameters and higher classification accuracy [31]. It is a CNN architecture model for image classification and mobile vision applications. MobileNet uses depth-wise separable convolutions in each color channel rather than combining all three and flattening them. The depth-wise separable split into two layers, a separate layer for combining and a separate layer of filtering. This factorization of MobileNet architecture reduces the computation and model size. MobileNet is very suitable for embedded systems because it requires very low computation power to run and can be very effective in the medical domain. We can build an artificially intelligent medical device through the MobileNet model which will take very low computational power with optimal time and give very high accuracy. MobileNet's architecture is best for embedded vision applications. Another quality of MobileNet architecture is two global hyper-parameters that efficiently introduce the trade-off between accuracy and latency. For a complete description of its parameters and detailed discussion about the MobileNet architecture, readers are referred to [31].

B. CONVOLUTIONAL NEURAL NETWORK

CNN is a powerful network type of deep neural network to handle a great amount of complexity around preprocessing and computation of data. The common building blocks of the CNN model consist of a convolutional layer, pooling layer, flattening layer, nonlinear activation layer, and dropout layer. The convolutional layer extracts the feature maps out

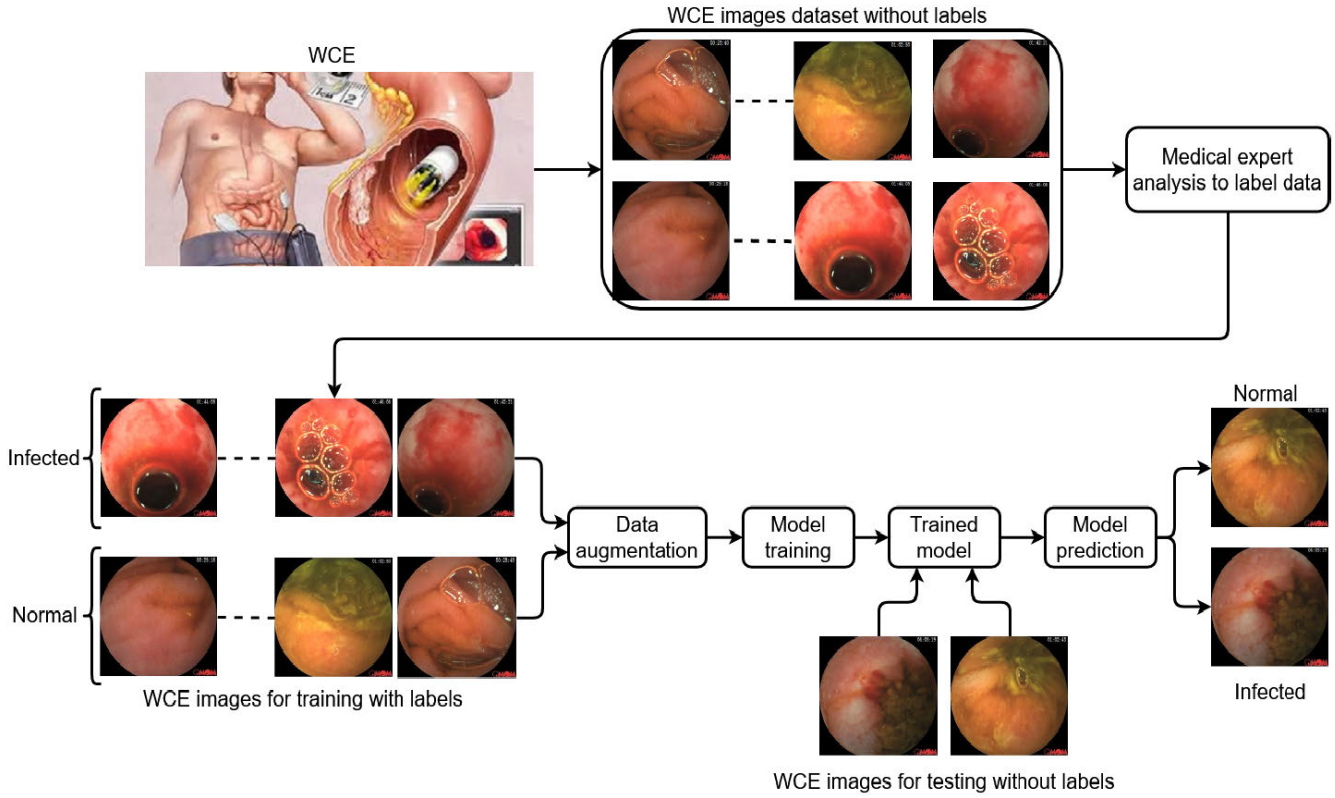


FIGURE 3. WCE images classification approach.

of input images which is also the main layer in the CNN model. The pooling layer is also known as the sub-sampling layer which is an important building block of CNN. The pooling layer works on feature maps extracted through the convolutional layer. It reduces the size of the features to extract the important features from the feature maps to avoid the over-fitting of the model. Pooling can be a max, average, or sum. The max-pooling is used in the proposed model because max-pooling can find sharp features as compared to that of average and sum pooling. The flattening layer converts (flatten) data into an array so the dense layer can perform computation on the data. This study uses the rectified linear unit (ReLU) activation function with the convolutional layer because natural images are non-linear so ReLU can be helpful to increase the non-linearity in the input image and CNN with ReLU is much easier and faster. ReLU can be defined as:

$$Y = max(0, i) \tag{1}$$

Here Y is output for the function which will be zero for all negative numbers and positive numbers will remain constant. So ReLU is zero for all negative numbers as illustrated in Figure 4. Over-fitting is a big issue while training a deep neural network and it happens when two or more neurons detect the same results repeatedly [28]. So, the dropout layer helps to drop the neurons in the neural network which helps to avoid over-fitting. We used the dropout layer with the 0.2 and 0.5 dropout rates.

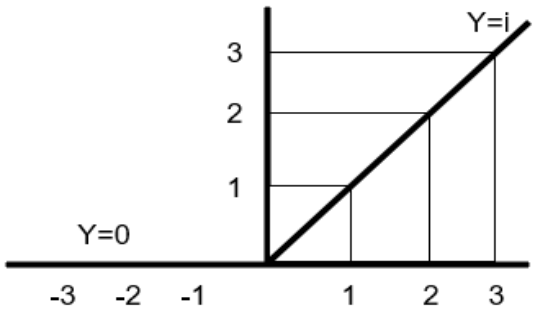


FIGURE 4. ReLU function.

C. ARCHITECTURE OF BIR

The proposed model consists of the MobileNet model and CNN. The output of MobileNet will be the input for the CNN. The model consists of a MobileNet model and other layers like one convolutional layer, one max-pooling layer, three dropout layers, one activation function layer, one batch normalization layer, one flatten layer, and three fully connected layers. The model input shape is $(150 \times 150 \times 3)$ indicating the image with 150 height and 150 widths with 3 channels red, green, and blue (RGB). MobileNet takes the input data and performs computation on it. The output of MobileNet serves as input for the Conv2D layer. In the Conv2D layer, 64 filters and 3*3 kernel size with the ReLU activation function are used. Max-pooling layer with 2×2 pooling size and batch normalization layer is used with the axis $= -1$ argument.

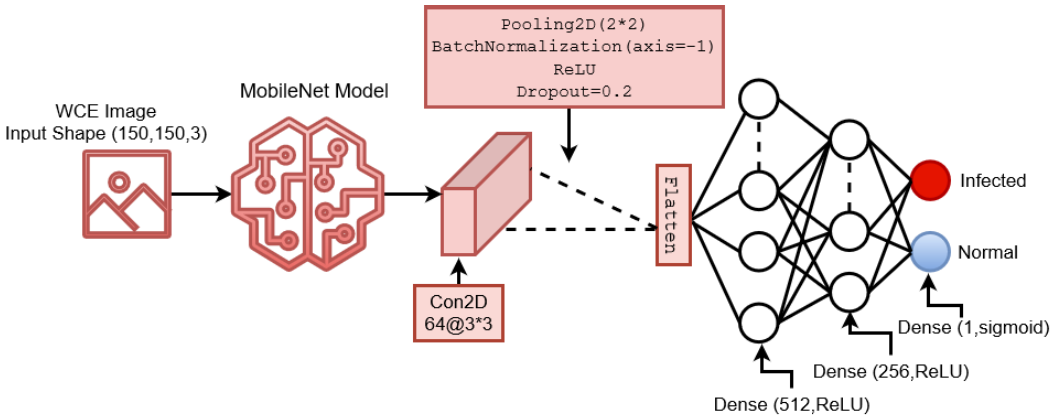


FIGURE 5. Architecture of BIR.

TABLE 2. BIR layers description.

Layer	Values	Output Shape
Mobilenet	-	(None ,4,4,1024)
Conv2D	Activation =ReLU, Filters=64, kernel=3*3	(None, 4, 4, 64)
MaxPooling2D	2*2	(None, 2, 2, 64)
BatchNormalization	axis=-1	(None, 2, 2, 64)
Activation	ReLU	(None, 2, 2, 64)
Dropout	Rate = 0.2	(None, 2, 2, 64)
Flatten	-	(None, 256)
Dense	Neurons = 512	(None, 512)
Dropout	Rate = 0.5	(None, 512)
Dense	Neurons = 256	(None, 256)
Dropout	Rate = 0.5	(None, 256)
Dense	Activation = Sigmoid	(None, 1)

The activation layer is used with the ReLU function and dropout layer with a 20% dropout rate after the Conv2D layer. The output from the pooling layer is 3 dimensional and a flattened layer is used to convert it into a vector before passing it to three fully connected layers. All layers and their operations are shown in Table 2 and the architecture of the proposed model shown in Figure 5.

The proposed model is to solve the binary classification problem so, 'binary_ crossentropy' loss function is used to measure the performance of a classification model whose output is a probability value between 0 and 1. It finds the error between the predicted class and actual class. Binary crossentropy loss function can be defined as in equation 2:

$$H_p(q) = -\frac{1}{N}\sum_{i=1}^{N} y_i . \log(p(y_i)) + (1-y_i) . \log(1-p(y_i)), \quad (2)$$

where y is the label (1 for Infected and 0 for Normal) and $p(y)$ is the predicted probability of the point being Infected for all N points. 'Adam' optimizer is used for training which changes the attribute weight and learning rate to reduce the loss of the learning model. Adam stands for adaptive moment estimation which is an update to the RMSProp optimizer.

MobileNet is a well-known deep neural network model that utilizes CNN architecture to offer high classification accuracy for several tasks. It is especially suitable for embedded devices concerning its low computational resources requirement. Despite its high classification accuracy, it does not ensure the same efficacy for different image classification tasks on divergent datasets. For example, when used with the newly collected bleedy endoscopy images, it does not provide high accuracy. However, its being lightweight and availability of two hyperparameters to control accuracy and latency, it can be leveraged to classify the bleedy endoscopy images. For this purpose, we use MobileNet for features extraction and features are passed to a custom-built CNN model for training. Compared to simple convolution and pooling layers in a CNN architecture, MobileNet provides higher classification accuracy for endoscopy bleedy images when ensembled with a custom-built CNN model.

V. EVALUATION METRICS

The performance of the deep learning model is analyzed in terms of accuracy, precision, recall, F1 score, and Cohen's kappa. As this study solves the binary classification problem so the confusion matrix is also used to evaluate the model. The confusion matrix is a table used to describe the classification model performance on the test data. It visualizes the performance of the learning model, Figure 6 shows the structure and details of the confusion matrix.

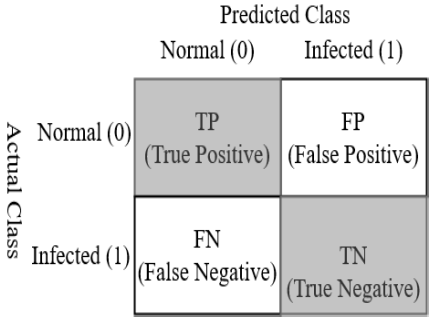


FIGURE 6. Confusion matrix.

A. ACCURACY

The accuracy score shows how much accurate model prediction is. In other words, the accuracy score is a ratio of the number of correct predictions given by a model. An accuracy score can be found by dividing the total number of correct

predictions by the total number of predictions. A model maximum accuracy score can be 1 and a minimum of 0. We can define it as:

$$Accuracy = \frac{correct\ predictions}{total\ predictions} \tag{3}$$

We can also define it as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

Here,

- **TP** is the value when the model predicts the image as Normal and the actual label of the image is also Normal.
- **TN** is the value when the model predicts the image as Infected and the actual label of the image is also Infected.
- **FP** is the value when the model predicts the image as Normal but the actual label of the image is Infected.
- **FN** is the value when the model predicts the image as Infected but the actual label of the image is Normal.

B. PRECISION

Precision is the ratio of correctly predicted positive observations of the total predicted positive observations. A model maximum precision score can be 1 and a minimum of 0. We can define precision as:

$$Precision = \frac{TP}{TP + FP} \tag{5}$$

C. RECALL

The recall is also known as the True Positive Rate (TPR) and Sensitivity. The recall score is intuitively the ability of the classifier to find all the positive samples. It is a ratio of TP divide by the sum of TP and FN. A model maximum recall score can be 1 and a minimum of 0. We can define it as:

$$Recall = \frac{TP}{TP + FN} \tag{6}$$

D. F1 SCORE

F1 score is the harmonic mean between precision and recall. It also knows as F Measure. It shows the balance between precision and recall and a model maximum F1 score can be 1 and minimum 0. We can define it as:

$$F_1 score = 2 * \frac{Precision\ *\ Recall}{Precision\ +\ Recall} \tag{7}$$

E. COHENS KAPPA VALUE

This value needs to be compared to the value that you would expect if the two raters were independent. The numerator represents the discrepancy between the observed probability of success and the probability of success under the assumption of an extremely bad case. According to Cohen's original article, values ≤ 0 indicates no agreement, 0.01–0.20 as none to slight, 0.21–0.40 as fair, 0.41– 0.60 as moderate, 0.61–0.80 as substantial, and 0.81–1.00 as almost perfect agreement.

VI. EXPERIMENT SETUP

A. PROBLEM STATEMENT

The use of machine learning techniques in medical imaging and medical decision making has been an attractive area of research. Using machine learning algorithms for diagnosing and predicting diseases is now a trending area on account of cost-effectiveness, less time-consumption, and high accuracy. Although not at par, many machines and deep learning models can perform diagnosis and disease prediction, competitive to human experts. The current study aims at detecting bleeding in the esophagus, stomach, small intestine, large intestine, rectum, and anus using WCE images and developing an automatic system using deep convolutional neural networks. With the human medical experts, it is a time-consuming process because the WCE procedure generates approximately 60,000 images and human experts cannot perform analysis on this large amount of data accurately and timely. To solve this problem we propose an approach which can detect infected images from a large amount of data accurately and help medical expert to perform fast-paced diagnosis. The architecture of the proposed approach is shown in Figure 7.

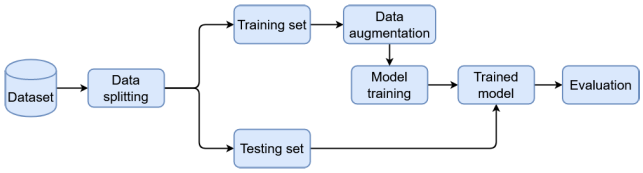


FIGURE 7. Our proposed approach.

B. EXPERIMENT FLOW

For experiments, first, the data is collected through WCE technology which is later checked by medical experts to segregate infected frames/images (see Section III). The number of infected and normal images is not equal and the dataset is imbalanced. The ratio of infected images is less as compared to the normal images in the dataset. The total images are 5000, out of which, 450 are infected and 4550 are normal. First, experiments are performed on the original dataset by splitting it into training and testing data as shown in Table 3. Using the training data we train the BIR model and then evaluate the performance of the trained model using the test data. The performance of the model is poor due to the highly imbalanced data.

To resolve dataset imbalance, data augmentation is performed on the minority class (infected) data. For data augmentation, we use only flip image techniques because the use of multiple data augmentation techniques can cause data redundancy [32], [33]. Data augmentation is performed only on 375 training samples of infected class. Augmentation generates 375 images for the infected class

TABLE 3. Training and testing count in the original dataset.

Dataset	Infected	Normal	Total
Total images	450	4550	5000
Training data	375	4475	4850
Testing data	75	75	150

TABLE 4. Training and testing count after data augmentation.

Dataset	Infected	Normal	Total
Total images	825	825	1650
Training data	750	750	1500
Testing data	75	75	150

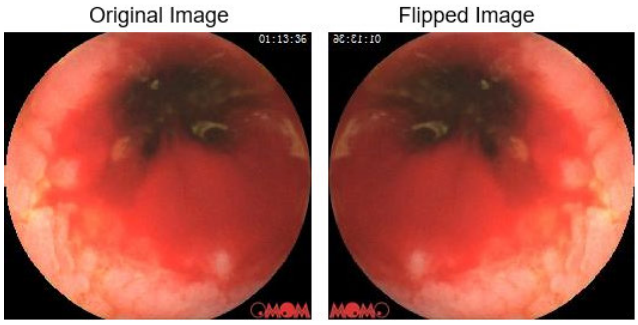


FIGURE 8. Flipped image sample.

thus making it a total of 750 images for the infected class as shown in Table 4. Figure 8 illustrates a sample image after flipped.

By generating flipped images, the number of infected images reaches 750 (original images + flipped images). Consequently, more data and features can be fed to the BIR model for training and testing. Moreover, to make the dataset balanced for training, an equal number of images from the normal and infected classes are used. After data augmentation and balancing of data, it is split into training and testing as shown in Table 4.

The experiments have been conducted using Intel Core i7 7th generation machine with 8GB random access memory (RAM), on Microsoft Windows 10 platform using Python language with Anacondas 3 software, and Jupyter notebook.

VII. RESULTS

Several experiments are performed to evaluate the efficacy and performance of the proposed BIR for WCE bleedy image classification. For this purpose, experiments are carried out using an imbalanced dataset, weighted cross-entropy, and balanced dataset using augmentation.

A. EXPERIMENTS USING IMBALANCED DATASET

The current study focuses on the use of MobileNet and CNN to perform the classification of bleedy images captured using WCE technology. The WCE generates a large number of images during its operation and medical experts need to spend a substantial amount of time to perform diagnosis. Previous research on automatic detection utilizes machine learning, as well as, deep learning approaches, yet, the provided accuracy can still be enhanced. We contribute to this classification problem by enhancing the accuracy of the classification model. For this purpose, the BIR model is proposed that performs very well for bleedy image classification.

First, the BIR model is tested on an imbalanced dataset in which 450 of the images are infected while 4550 normal images. Of 5,000 images, we used 4,850 images for training

TABLE 5. Results of BIR without augmentation.

Evaluation Parameter	Score
Accuracy	0.800
Precision	1.000
Recall	0.600
F1 score	0.750
Cohens Kappa	0.600

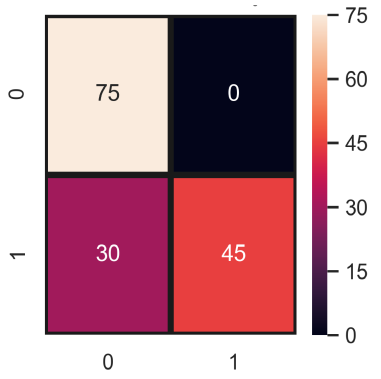


FIGURE 9. Confusion matrix of BIR before data augmentation, Here 0 representing the normal class and 1 representing the infected class.

and 150 images for testing. After training of the model, we pass test data containing 150 images (75 infected and 75 normals) to evaluate the model, and results are shown in Table 5.

The model performs inconsistently in terms of evaluation metrics, where the values for various metrics vary significantly. The confusion matrix in Figure 9 shows that the model predicts 120 images correctly and 30 images incorrectly. These 30 wrong predicted images belong to the infected class but the model predicts them as normal images.

The underlying reason for this poor performance is the over-fitting of the model on single class images data because we have 4,550 normal class images and only 450 infected class images in the dataset which shows the characteristic of a highly imbalanced dataset. So the model gets more training for normal class images than those of the infected class. Consequently, the model gives wrong predictions for the infected class. Another reason is that several infected images in the dataset have similar features to that of normal image features as shown in Figure 10. Many images that reflect the early stage of infection, look similar to normal images. Lacking distinguished features, such images make correct classification difficult for the model which degrades the performance.

B. BALANCING DATASET USING DATA AUGMENTATION

In light of the experiment results, it is highly desirable to make the dataset balanced. To this end, each infected image is flipped to increase the number of infected images. We apply the data augmentation technique only on infected images because we already have a large number of normal images to train the model. So through data augmentation, we increase the number of images of the infected class and make a balanced dataset. After data augmentation, the total infected

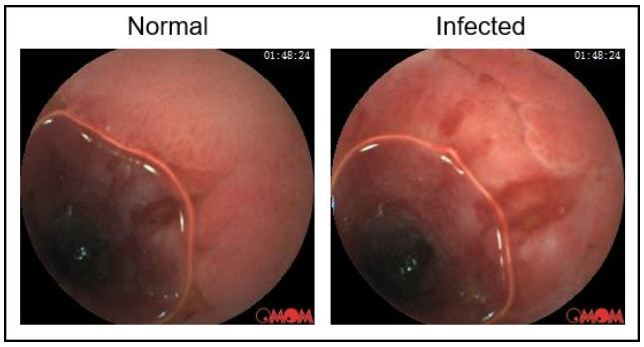


FIGURE 10. Similarity between infected image and normal image feature cause for the wrong prediction.

TABLE 6. Results of BIR after data augmentation.

Evaluation Parameter	Score
Accuracy	0. 993
Precision	1.000
Recall	0.994
F1 score	0.997
Cohen's Kappa	0.995

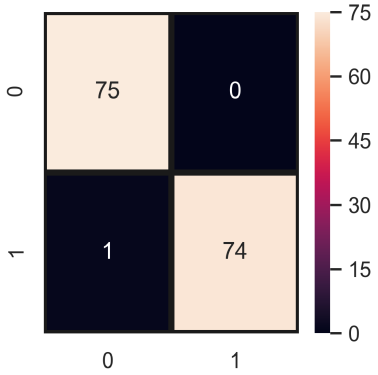


FIGURE 11. Confusion matrix of BIR after data augmentation. Here, 0 representing the normal class and 1 representing the infected class.

images reach 750. So, an equal number of normal images are taken to make the training data balanced which makes a total of 1500 images in the dataset. A total of 1500 images are used for training the model while 150 are used for model testing (75 for each class). The results of the proposed approach with the balanced dataset are shown in Table 6. There is no doubt that the values for evaluation metrics are in agreement.

The proposed approach obtains 99.3% accuracy and 100% precision score and predicts 149 images correctly out of 150 as shown in Figure 11. According to the confusion matrix, the model predicts only one image as normal but its actual class was infected.

The only wrong predicted image by the model is shown in Figure 12. Post examining reveals that the image is actually of an area that is an early stage of infection where the bleeding is not too clear so the feature of this image does not correctly represent the infected image.

We also include the cross-validation in our experiment to verify the effectiveness of the proposed approach. We apply K fold cross-validation after data augmentation and balancing

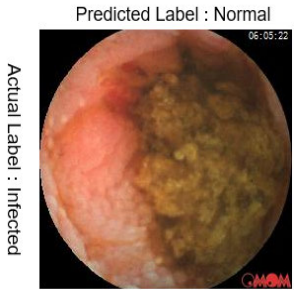


FIGURE 12. Single wrong prediction by our approach.

TABLE 7. Models performance with 10 fold cross-validation.

Iteration	SVM	LR	RF	ETC	DTC	BIR
Iteration 1	99.41	99.41	99.41	99.41	99.01	100
Iteration 2	98.24	96.15	100	100	96.51	100
Iteration 3	100	100	99.41	99.41	96.17	99.41
Iteration 4	97.04	100	100	98.02	100	100
Iteration 5	98.24	96.15	99.41	98.02	96.19	100
Iteration 6	98.24	96.15	98.82	99.41	95.42	100
Iteration 7	98.24	96.15	97.65	98.02	97.54	99.41
Iteration 8	97.04	96.15	97.56	99.41	95.39	99.41
Iteration 9	97.04	96.15	100	100	96.37	100
Iteration 10	97.04	96.15	98.22	98.02	97.21	100
Average	98.05	97.24	99.04	98.97	96.98	99.82
Confidence interval	+/-1.12	+/-1.25	+/-0.67	+/-0.53	+/-0.29	+/-0.09

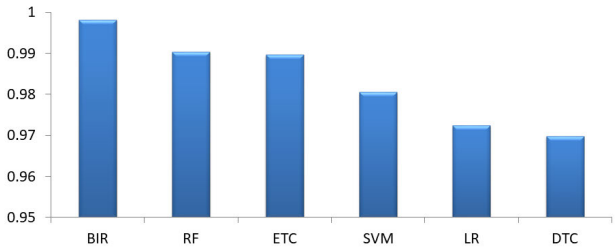


FIGURE 13. Models average accuracy score comparison with 10 fold cross-validation.

of data. The BIR model shows its effectiveness in this scenario and outperforms other models as shown in Table 7. The results show that BIR achieves the highest average accuracy of 99.82% with a +/− 0.09 confidence interval score.

C. PERFORMANCE OF BIR USING WEIGHTED CROSS-ENTROPY

One potential solution to the data imbalance problem is to follow the weighted cross-entropy approach where more weights are put on the class with fewer samples. The weighted cross-entropy function is used to put more weight on the minority class (bleedy images) than the majority class (normal images). As the dataset contains 450 infected and 4,550 normal images, the ratio of normal class images is approximately 10 times the infected images. The weight for the infected class is set to 10 in the cross-entropy loss. Experiments are performed to evaluate the performance of BIR with cross-entropy loss function with the imbalanced dataset. Results indicate that the accuracy of the BIR has been degraded severely when cross-entropy is used on the imbalanced dataset. Figures 14a and 14b show the graphs

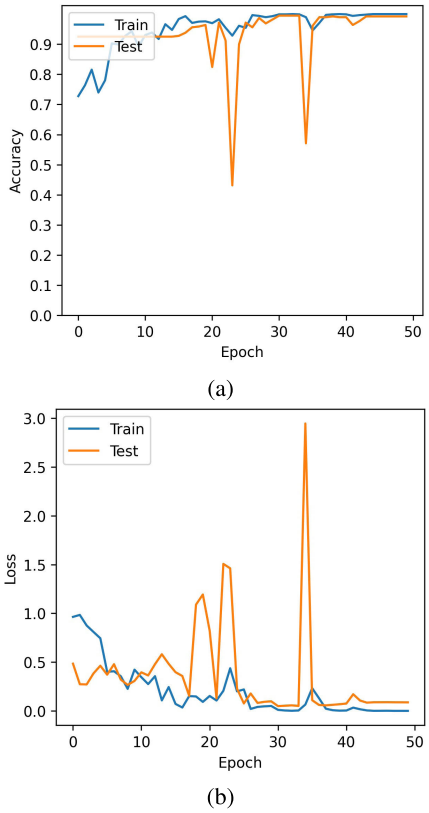


FIGURE 14. Accuracy and loss of BIR model when trained using the weighted cross-entropy, (a) BIR accuracy, and (b) BIR loss.

for model accuracy and loss when cross-entropy is used with BIR. Experimental results indicate that BIR classification accuracy is reduced to 0.62 with the cross-entropy.

D. BIR COMPARISON WITH OTHER MODELS

To compare the results of BIR with other models we used the CNN model and MobileNet model individually. The detailed specification of the structure of used CNN is given in Figure 15.

The performance of both the CNN and MobileNet is given in Table 8, along with the results of the BIR. Similarly, Figure 16 shows the performance of BIR concerning the accuracy, precision, recall, F1 score, and Cohen's Kapp in comparison to the CNN model and MobileNet. BIR outperforms the CNN model and MobileNet.

MobileNet model is used with the same setting as given in [31] while the CNN model is developed using 3 convolutional layers, 3 max-pooling layers, 5 dropout layers, 3 activation function layers, 3 batch Normalization layers, 1 flatten layer, and 3 fully connected layers.

We also used machine learning models for performance comparison with the BIR and deep neural networks after the data augmentation technique. Machine learning models such as RF [34], [35], ETC [36], SVM [37], LR [38], and DTC [39] are used on WCE images. We implement RF with three hyper-parameters including n_estimators, max_depth, and random_state. The n_estimator is used with 300 indicating that 300 decision trees will be constructed under RF criteria to

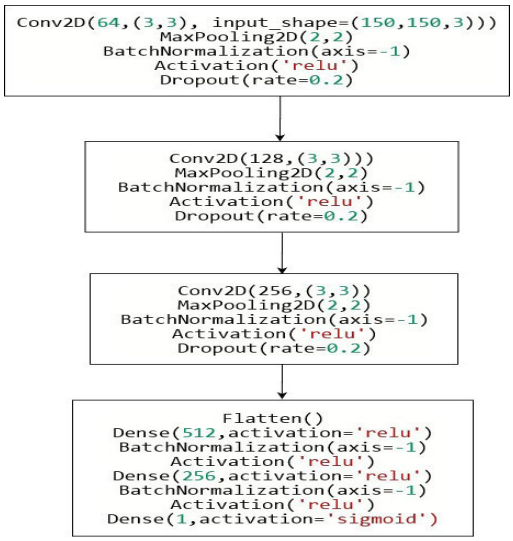


FIGURE 15. The CNN model layers architecture.

TABLE 8. CNN and MobileNet and BIR model results comparison.

Model	Accuracy	Precision	Recall	F_1 Score	Cohen's Kappa
CNN	0.954	0.973	0.950	0.961	0.940
MobileNet	0.980	0.980	0.970	0.975	0.973
BIR	0. 993	1.000	0.994	0.997	0.995

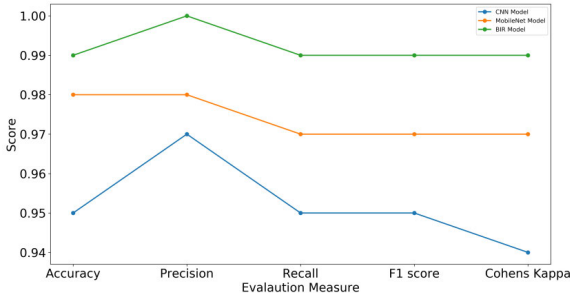


FIGURE 16. BIR, CNN and MobileNet model performance comparison.

make a final prediction and the max_depth hyper-parameter is also used with value 300 which will restrict each decision to maximum 300 level depths [40]. Similarly, for ETC we also used 300 n_estimators. An SVM model is applied to the dataset using the 'linear' kernel because the 'linear' kernel is better in binary classification problems when data is linearly separable. While LR is used with 'liblinear' solver and 'ovr' parameter which is considered better in binary classification problems. DTC is used only with one hyper-parameter which is max_depth and with the same value 300 as in RF, and ETC classifiers. The hyper-parameters of machine learning models are given in Table 9.

Machine learning models performed well on the WCE image dataset for image classification, especially RF, ETC, and DTC performed better than those of SVC and LR which shows the efficiency of tree-based models on WCE image data over linear models. The results of machine learning models are shown in Table 10 and a comparison between machine learning models and BIR has shown in Figure 17.

TABLE 9. Hyper-parameters for machine learning algorithms.

Models	Hyper-parameters
RF	n_estimators=300, random_state=5, max_depth=300
ETC	n_estimators=300, random_state=5, max_depth=300
LR	solver='liblinear', multi_class='ovr'
SVM	kernel='linear', C=2.0
DTC	max_depth=300

TABLE 10. Machine learning model performance on same dataset used in our proposed approach.

Model	Accuracy	Precision	Recall	F1 Score	Cohen's Kappa
RF	0.97	0.97	0.97	0.97	0.96
ETC	0.93	0.94	0.93	0.93	0.93
LR	0.88	0.91	0.88	0.88	0.87
SVM	0.89	0.91	0.89	0.89	0.89
DTC	0.91	0.92	0.91	0.91	0.91

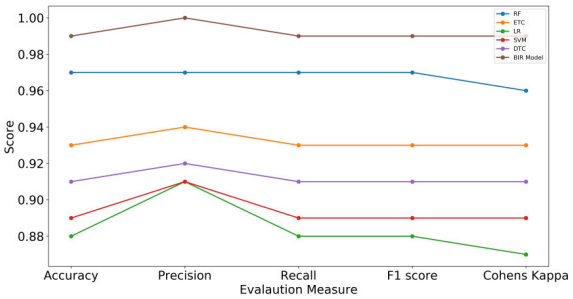


FIGURE 17. Machine learning models performance comparison with BIR model.

E. ANALYSIS OF PROPOSED APPROACH WITH ADDITIONAL DATASET

Besides the performance evaluation on the selected dataset and with other state-of-the-art approaches, the proposed model is tested on an additional dataset. The dataset is gathered using the Google search engine where bleedy and normal WCE images are collected. Due to the poor quality of the downloaded images, image enhancement is carried out to increase the quality of WCE images. This helps to achieve an accuracy similar to that of our own collected WCE images. Few samples of normal and bleedy images are shown in Figure 18. The dataset contains both normal and infected images in a similar proportion and comprises 22 and 25 images for normal and infection, respectively.

The collected data is used to validate the performance of the proposed model for detecting the bleedy WCE images. Experiments are performed and the results are shown in Table 11. Results indicate that the performance of the proposed approach is good even with the newly collected data whose images' quality is not as good as that of our own WCE image dataset. Despite that the accuracy score and other performance evaluation measures are slightly reduced, the performance is still competitive with CNN and MobileNet models. A slight reduction in the accuracy is caused due to the quality of the WCE images collected from Google. Although image enhancement is carried out, the quality is still inferior to that of images in the original dataset. Newly collected dataset is publicly available at [41].

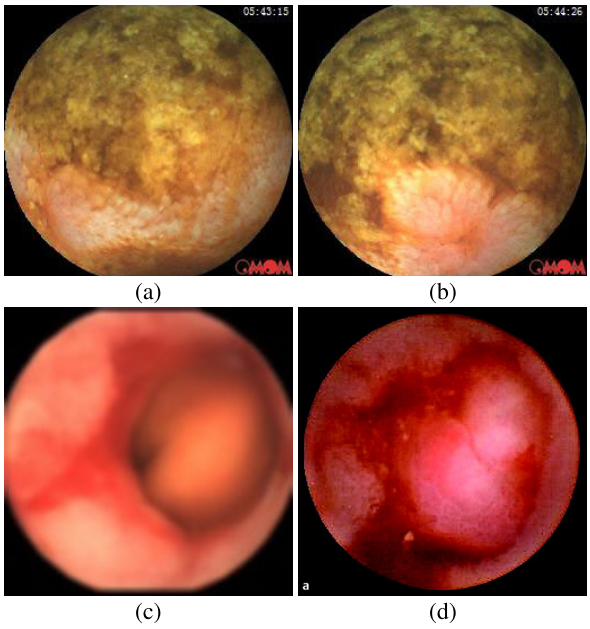


FIGURE 18. Sample images from the new dataset, (a,b) Normal WCE images, and (c,d) Bleedy WCE images.

TABLE 11. Performance with BIR and other models with two datasets.

Model	Accuracy	Precision	Recall	F_1 Score	Kappa
CNN	0.954	0.973	0.950	0.961	0.940
MobileNet	0.980	0.980	0.970	0.975	0.973
BIR-dataset1	0. 993	1.000	0.994	0.997	0.995
BIR-dataset2	0. 978	0.990	0.960	0.979	0.957

VIII. DISCUSSION

This section presents an in-depth analysis of experimental results through the proposed technique based on accuracy measures: accuracy, precision, recall, F1 score, and Cohen's Kappa. This study is composed of four steps: data collection, dataset preprocessing, classification model, and results analysis. The data are collected using WCE technology from 33 patients in a hospital. In preprocessing, data augmentation helps to generate more images for the infected class to make the dataset balanced for training the model and help avoid the model overfit. The classification of the proposed BIR model is elevated by 20% as a result of dataset balancing. The proposed model BIR is a combination of CNN and MobileNet model and proves to be efficient than those of individual CNN model and MobileNet models, as shown in Table 8. The BIR model performs significantly on visual RGB features and achieves the accuracy of 0.993 and 1.000 precision score on the original dataset. The performance of the BIR model on the Google collected dataset is slightly reduced and it achieves an accuracy of 0.978 and precision of 0.99 which is also better than the state-of-art approaches except [42].

Performance analysis of seven classifiers including RF, ETC, LR, SVM, DTC, CNN, and MobileNet with the proposed BIR model indicates that the proposed model outperforms these models. For further validation of the proposed model, we applied a ten-fold cross-validation technique and the results prove the significance of BIR as shown in Table 7.

TABLE 12. Previous studies result in comparison with BIR.

Reference	Disease	Year	Accuracy
[3]	Bleeding	2014	0.950
[10]	Bleeding	2016	0.957
[43]	Bleeding	2017	0.976
[42]	Bleeding, Ulcer	2018	0.983
[44]	Abnormality	2019	0.894
BIR-dataset1	Bleeding	2020	0.993
BIR-dataset1	Bleeding	2020	0.978

TABLE 13. Performance comparison with state-of-the-art approaches.

Reference	Disease	Year	Accuracy
[3]	Bleeding	2014	0.698
[10]	Bleeding	2016	0.719
[43]	Bleeding	2017	0.713
[42]	Bleeding, Ulcer	2018	0.480
[44]	Abnormality	2019	0.867
BIR	Bleeding	2020	0.993

the BIR achieves the highest average accuracy of 99.8% in 10 fold cross-validation. To corroborate its superiority over other approaches, its performance is compared with four other similar approaches from the literature. The results of the comparison are provided in Table 12.

Table 12 shows the reported accuracy of the classifications algorithms on different datasets and comparison seems inappropriate. For a fair performance analysis, we implemented selected models [3], [10], [42]–[44] on WCE dataset. Training and testing are performed following the same procedure that is used to evaluate BIR. Experimental results are shown in Table 13. Results indicate that the proposed BIR outperforms state-of-the-art approaches concerning accuracy. The selected approaches perform poorly on the WCE image dataset.

IX. CONCLUSION

This study proposes an approach for automatic classification of WCE images to detect the bleedy images. The dataset contains a total of 4550 and 450 normal and bleedy images, respectively, from 33 patients. The proposed model BIR, an ensemble of MobileNet and custom-defined CNN architecture, uses the MobileNet for feature extraction and the CNN for training. Several experiments are performed on both imbalanced and image flipping based balanced datasets. Results indicate that balancing the dataset increases the classification accuracy from 0.80 to 0.993 with precision, recall, F1 score, and Cohen's kappa values of 1.00, 0.994, 0.997, and 0.995, respectively. The BIR model performs well with the Google collected WCE images dataset also and achieves an accuracy of 0.978 which is better than the state-of-art approaches. Ensemble provides higher accuracy than using either of MobileNet or CNN architecture for classification. Four state-of-the-art approaches are selected from the literature for performance analysis. Experiments on the WCE dataset suggest that the proposed BIR model show superior performance than that of state-of-the-art approaches.

Results from five machine learning models applied on the WCE dataset suggest that the performance of machine learning models is inferior to that of deep learning models. Additionally, tree-based machine learning models like RF, ETC, and DTC tend to perform better than linear models like SVM and LR. In the end, we want to declare that the performance of BIR is tested only on a single WCE image dataset, and may vary when applied to other datasets. In the future, we want to conduct further experiments on more datasets to make their results generalizable. For external validation of the proposed approach, more data are required which we intend to collect and perform more experiments in the future.

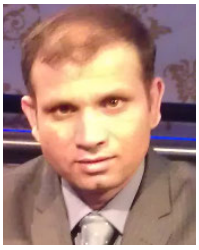
REFERENCES

[1] WHO. (2020). *Death Rate Because of Infection Disease*. Accessed: May 20, 2020. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/cancer
[2] R. L. Siegel, K. D. Miller, S. A. Fedewa, D. J. Ahnen, R. G. S. Meester, A. Barzi, and A. Jemal, "Colorectal cancer statistics, 2017," *CA, Cancer J. Clinicians*, vol. 67, no. 3, pp. 177–193, May 2017.
[3] Y. Fu, W. Zhang, M. Mandal, and M. Q.-H. Meng, "Computer-aided bleeding detection in WCE video," *IEEE J. Biomed. Health Informat.*, vol. 18, no. 2, pp. 636–642, Mar. 2014.
[4] M. Sharif, M. A. Khan, M. Rashid, M. Yasmin, F. Afza, and U. J. Tanik, "Deep CNN and geometric features-based gastrointestinal tract diseases detection and classification from wireless capsule endoscopy images," *J. Exp. Theor. Artif. Intell.*, pp. 1–23, Feb. 2019.
[5] T. Rahim, M. A. Usman, and S. Y. Shin, "A survey on contemporary computer-aided tumor, polyp, and ulcer detection methods in wireless capsule endoscopy imaging," 2019, *arXiv:1910.00265*. [Online]. Available: http://arxiv.org/abs/1910.00265
[6] X. Liu, J. Gu, Y. Xie, J. Xiong, and W. Qin, "A new approach to detecting ulcer and bleeding in wireless capsule endoscopy images," in *Proc. IEEE-EMBS Int. Conf. Biomed. Health Informat.*, Jan. 2012, pp. 737–740.
[7] Y. Yuan, B. Li, and M. Q.-H. Meng, "WCE abnormality detection based on saliency and adaptive locality-constrained linear coding," *IEEE Trans. Autom. Sci. Eng.*, vol. 14, no. 1, pp. 149–159, Jan. 2017.
[8] Y. Yuan, J. Wang, B. Li, and M. Q.-H. Meng, "Saliency based ulcer detection for wireless capsule endoscopy diagnosis," *IEEE Trans. Med. Imag.*, vol. 34, no. 10, pp. 2046–2057, Oct. 2015.
[9] J. Bernal, F. J. Sánchez, G. Fernández-Esparrach, D. Gil, C. Rodríguez, and F. Vilariño, "WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians," *Comput. Med. Imag. Graph.*, vol. 43, pp. 99–111, Jul. 2015.
[10] Y. Yuan, B. Li, and M. Q.-H. Meng, "Bleeding frame and region detection in the wireless capsule endoscopy video," *IEEE J. Biomed. Health Informat.*, vol. 20, no. 2, pp. 624–630, Mar. 2016.
[11] K. Pogorelov, S. Suman, F. A. Hussin, A. S. Malik, O. Ostroukhova, M. Riegler, P. Halvorsen, S. H. Ho, and K. Goh, "Bleeding detection in wireless capsule endoscopy videos—Color versus texture features," *J. Appl. Clin. Med. Phys.*, vol. 20, no. 8, pp. 141–154, Aug. 2019.
[12] M. Mathew and V. P. Gopi, "Transform based bleeding detection technique for endoscopic images," in *Proc. 2nd Int. Conf. Electron. Commun. Syst. (ICECS)*, Feb. 2015, pp. 1730–1734.
[13] D.-Y. Liu, T. Gan, N.-N. Rao, Y.-W. Xing, J. Zheng, S. Li, C.-S. Luo, Z.-J. Zhou, and Y.-L. Wan, "Identification of lesion images from gastrointestinal endoscope based on feature extraction of combinational methods with and without learning process," *Med. Image Anal.*, vol. 32, pp. 281–294, Aug. 2016.
[14] E. Tuba, M. Tuba, and R. Jovanovic, "An algorithm for automated segmentation for bleeding detection in endoscopic images," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, May 2017, pp. 4579–4586.
[15] P. Szczypiński, A. Klepaczko, M. Pazurek, and P. Daniel, "Texture and color based image segmentation and pathology detection in capsule endoscopy videos," *Comput. Methods Programs Biomed.*, vol. 113, no. 1, pp. 396–411, Jan. 2014.
[16] B. Li and M. Q.-H. Meng, "Computer-aided detection of bleeding regions for capsule endoscopy images," *IEEE Trans. Biomed. Eng.*, vol. 56, no. 4, pp. 1032–1039, Apr. 2009.
[17] S. Wang, Y. Xing, L. Zhang, H. Gao, and H. Zhang, "Deep convolutional neural network for ulcer recognition in wireless capsule endoscopy: Experimental feasibility and optimization," *Comput. Math. Methods Med.*, vol. 2019, pp. 1–14, Sep. 2019.

[18] J. Y. Lee, S.-H. Choi, and J. W. Chung, "Automated classification of the tympanic membrane using a convolutional neural network," *Appl. Sci.*, vol. 9, no. 9, p. 1827, May 2019.
[19] S. S. Chaturvedi, K. Gupta, and P. S. Prasad, "Skin lesion analyser: An efficient seven-way multi-class skin cancer classification using mobilenet," in *Proc. Int. Conf. Adv. Mach. Learn. Technol. Appl.* Singapore: Springer, 2020, pp. 165–176.
[20] I. D. Apostolopoulos and T. A. Mpesiana, "COVID-19: Automatic detection from X-ray images utilizing transfer learning with convolutional neural networks," *Phys. Eng. Sci. Med.*, vol. 43, no. 2, pp. 635–640, Jun. 2020.
[21] R. V. M. D. Nobrega, S. A. Peixoto, S. P. P. da Silva, and P. P. R. Filho, "Lung nodule classification via deep transfer learning in CT lung images," in *Proc. IEEE 31st Int. Symp. Comput.-Based Med. Syst. (CBMS)*, Jun. 2018, pp. 244–249.
[22] G. Pan and L. Wang, "Swallowable wireless capsule endoscopy: Progress and technical challenges," *Gastroenterol. Res. Pract.*, vol. 2012, Oct. 2012, Art. no. 841691. [Online]. Available: https://europepmc.org/articles/PMC3255457
[23] Medtronic. (2020). *PillCam SB 3 System*. Accessed: Sep. 9, 2020. [Online]. Available: https://www.medtronic.com/covidien/en-us/products/capsule-endoscopy/pillcam-sb-3-system.html
[24] W. De Mulder, S. Bethard, and M.-F. Moens, "A survey on the application of recurrent neural networks to statistical language modeling," *Comput. Speech Lang.*, vol. 30, no. 1, pp. 61–98, Mar. 2015.
[25] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.
[26] M. Umer, S. Sadiq, M. Ahmad, S. Ullah, G. S. Choi, and A. Mehmood, "A novel stacked CNN for malarial parasite detection in thin blood smear images," *IEEE Access*, vol. 8, pp. 93782–93792, 2020.
[27] Q. Li, W. Cai, X. Wang, Y. Zhou, D. D. Feng, and M. Chen, "Medical image classification with convolutional neural network," in *Proc. 13th Int. Conf. Control Automat. Robot. Vis. (ICARCV)*, Dec. 2014, pp. 844–848.
[28] S. S. Yadav and S. M. Jadhav, "Deep convolutional neural network based medical image classification for disease diagnosis," *J. Big Data*, vol. 6, no. 1, p. 113, Dec. 2019.
[29] W. Rawat and Z. Wang, "Deep convolutional neural networks for image classification: A comprehensive review," *Neural Comput.*, vol. 29, no. 9, pp. 2352–2449, Sep. 2017.
[30] D. S. Kermany *et al.*, "Identifying medical diagnoses and treatable diseases by image-based deep learning," *Cell*, vol. 172, no. 5, pp. 1122–1131, 2018.
[31] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*. [Online]. Available: http://arxiv.org/abs/1704.04861
[32] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *J. Big Data*, vol. 6, no. 1, p. 60, Dec. 2019.
[33] W. A. Abbasi and F. U. A. A. Minhas, "Issues in performance evaluation for host–pathogen protein interaction prediction," *J. Bioinf. Comput. Biol.*, vol. 14, no. 03, Jun. 2016, Art. no. 1650011.
[34] A. Liaw and M. Wiener, "Classification and regression by randomforest," *R News*, vol. 2, no. 3, pp. 18–22, 2002.
[35] F. Rustam, A. Mehmood, M. Ahmad, S. Ullah, D. M. Khan, and G. S. Choi, "Classification of shopify app user reviews using novel multi text features," *IEEE Access*, vol. 8, pp. 30234–30244, 2020.
[36] M. Goetz, C. Weber, J. Bloecher, B. Stieltjes, H.-P. Meinzer, and K. Maier-Hein, "Extremely randomized trees based brain tumor segmentation," in *Proc. BRATS Challenge-MICCAI*, 2014, pp. 6–11.
[37] W. S. Noble, "What is a support vector machine?" *Nature Biotechnol.*, vol. 24, no. 12, pp. 1565–1567, Dec. 2006.
[38] D. G. Kleinbaum, K. Dietz, M. Gail, M. Klein, and M. Klein, *Logistic Regression*. New York, NY, USA: Springer, 2002.
[39] P. H. Swain and H. Hauska, "The decision tree classifier: Design and potential," *IEEE Trans. Geosci. Electron.*, vol. GE-15, no. 3, pp. 142–147, Jul. 1977.
[40] F. Rustam, A. Mehmood, S. Ullah, M. Ahmad, D. M. Khan, G. S. Choi, and B.-W. On, "Predicting pulsar stars using a random tree boosting voting classifier (RTB-VC)," *Astron. Comput.*, vol. 32, Jul. 2020, Art. no. 100404. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2213133720300585
[41] (2020). *WCE Image Dataset*. [Online]. Available: https://sites.google.com/view/ashimran/datasets
[42] A. Liaqat, M. A. Khan, J. H. Shah, M. Sharif, M. Yasmin, and S. L. Fernandes, "Automated ulcer and bleeding classification from WCE images using multiple features fusion and selection," *J. Mech. Med. Biol.*, vol. 18, no. 4, Jun. 2018, Art. no. 1850038.
[43] S. Suman, F. A. B. Hussin, A. S. Malik, K. Pogorelov, M. Riegler, S. H. Ho, I. Hilmi, and K. L. Goh, "Detection and classification of bleeding region in WCE images using color feature," in *Proc. 15th Int. Workshop Content-Based Multimedia Indexing*, Jun. 2017, pp. 1–6.
[44] X. Guo and Y. Yuan, "Triple ANet: Adaptive abnormal-aware attention network for WCE image classification," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2019, pp. 293–301.

FURQAN RUSTAM received the master's degree in computer science from The Islamia University of Bahawalpur, Pakistan, in 2017, and the Master of Science degree in computer science from the Department of Computer Science, Khwaja Fareed University of Engineering and Information Technology (KFUEIT), Rahim Yar Khan, Pakistan, in 2020. He worked as a Research Assistant with the Fareed Computing and Research Center, KFUEIT. His research interests include data mining, mainly working on machine learning and deep learning-based IoT, and text mining tasks.

MUHAMMAD ABUBAKAR SIDDIQUE received the B.Sc. degree in computer science and the Master of Information Technology (MIT) degree from Bahauddin Zakariya University, Multan, Pakistan, in 2003 and 2005, respectively, and the Ph.D. degree from the College of Computer Science, Chongqing University, Chongqing, China, in 2015. He joined the COMSAT Institute of Information Technology, as an Assistant Professor, for a period of one year. He is currently working as an Assistant Professor with the Department of Computer Science and Information Technology, Khwaja Fareed University of Engineering and Information Technology, Rahim Yar Khan, Pakistan.



HAFEEZ UR REHMAN SIDDIQUI received the B.Sc. degree in mathematics from Islamia University Bahawalpur (IUB), Pakistan, in 1998, the M.Sc. degree in computer science from Bahauddin Zakariya University (BZU) Multan, Pakistan, in 2000, and the Ph.D. degree in electronic engineering from London South Bank University, in April 2016. From 2001 to 2006, he worked as a Lecturer in computer science with the Network Institute of Computer Education (NICE). His research interests include biomedical and energy engineering applications, data recognition, image processing, system embedded programming, and machine learning.

SALEEM ULLAH was born in Ahmedpur East, Pakistan, in 1983. He received the B.Sc. in computer science from The Islamia University of Bahawalpur, Pakistan, in 2003, the M.I.T. degree in computer science from Bahauddin Zakariya University, Multan, in 2005, and the Ph.D. degree from Chongqing University, China, in 2012. From 2006 to 2009, he worked as a Network/IT Administrator in different companies. From August 2012 to Feb 2016, he worked as an Assistant Professor with The Islamia University of Bahawalpur. Since February 2016, he has been working as an Associate Professor with the Khwaja Fareed University of Engineering and Information Technology, Rahim Yar Khan. He has almost 13 years of industry experience in the field of IT. He is an Active Researcher in the field of adhoc networks, congestion control, and security.

ARIF MEHMOOD received the Ph.D. degree from the Department of Information and Communication Engineering, Yeungnam University, South Korea, in November 2017. He is currently working as an Assistant Professor with the Department of Computer Science & IT, The Islamia University of Bahawalpur, Pakistan. His research interests include data mining, mainly working on AI and deep learning-based text mining, and data science management technologies.

IMRAN ASHRAF received the M.S. degree in computer science from the Blekinge Institute of Technology, Karlskrona, Sweden, in 2010, and the Ph.D. degree in information and communication engineering from Yeungnam University, South Korea, in 2018. He has worked as a Postdoctoral Fellow with Yeungnam University, Gyeongsan, South Korea. He is currently working as an Assistant Professor with the Information and Communication Engineering Department, Yeungnam University. His research interests include indoor positioning and localization, indoor location-based services in wireless communication, smart sensors for smart cars, and data mining.

GYU SANG CHOI received the Ph.D. degree from the Department of Computer Science and Engineering, Pennsylvania State University, University Park, PA, USA, in 2005. From 2006 to 2009, he was a Research Staff Member with the Samsung Advanced Institute of Technology (SAIT) for Samsung Electronics. Since 2009, he has been a Faculty Member with the Department of Information and Communication, Yeungnam University, South Korea. His research interests include non-volatile memory and storage systems.