IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 1 February 2024, accepted 13 May 2024, date of publication 3 June 2024, date of current version 13 June 2024.

Digital Object Identifier 10.1109/ACCESS.2024.3408224

RESEARCH ARTICLE

Improving Endoscopic Image Analysis: Attention Mechanism Integration in Grid Search Fine-Tuned Transfer Learning Model for Multi-Class Gastrointestinal Disease Classification

MOHAMED A. ELMAGZOUB[1], SWAPANDEEP KAUR[2], SHEIFALI GUPTA[2], ADEL RAJAB[3], KHAIRAN D. RAJAB[3], MANA SALEH AL RESHAN[4], HANI ALSHAHRANI[3], (Senior Member, IEEE), AND ASADULLAH SHAIKH[4], (Senior Member, IEEE)

[1]Department of Network and Communication Engineering, College of Computer Science and Information Systems, Najran University, Najran 61441, Saudi Arabia

[2]Chitkara University Institute of Engineering and Technology, Chitkara University, Chandigarh, Punjab 140401, India

[3]Department of Computer Science, College of Computer Science and Information Systems, Najran University, Najran 66462, Saudi Arabia

[4]Department of Information Systems, College of Computer Science and Information Systems, Najran University, Najran 66462, Saudi Arabia

Corresponding author: Mana Saleh Al Reshan (msalreshan@nu.edu.sa)

This work was supported by the Deanship of Graduate Studies and Scientific Research at Najran University through the Elite Funding Program under Grant NU/EP/SERC/13/82.

ABSTRACT Due to a continuous change in people's lifestyle and dietary habits, gastrointestinal diseases are on the increase, with dietary changes being a major contributor to a variety of bowel problems. Around two million people around the world die due to gastrointestinal (GI) diseases. Endoscopy is a medical imaging technology helpful in diagnosing gastrointestinal diseases like polyps and esophagitis. Its manual diagnosis is time-consuming; hence, computer-aided techniques are now widely used for accurate and fast GI disease diagnosis. In this paper, the Kvasir dataset of 4000 endoscopic images, comprising 500 images of each of the eight gastrointestinal tract disease classes have been classified using seven grid search fine-tuned transfer learning models. The fine-tuned transfer learning models employed in this paper are ResNet101, InceptionV3, InceptionResNetV2, Xception, DenseNet121, MobileNetV2, and ResNet50. The grid search algorithm has been used to determine the architectural and fine-tuning hyperparameters. The fine-tuned ResNet101 model performed the best, with a learning rate 0.001 and a batch size of 32 for the SGD optimizer at 40 epochs. These hyperparameters were optimized through grid search along with new set of layers added to the model. The newly added layers include one flatten layer, two dropout layers and five dense layers optimized using grid search. The grid search fine-tuned ResNet101 model obtained an accuracy of 0.90, a precision of 0.92, a recall of 0.92, and an f1-score of 0.91. Further, the grid search fine-tuned ResNet101 model was integrated with an attention mechanism to enhance performance by focusing on essential image features, notably in medical imaging where some regions may contain vital diagnostic information. The proposed grid search fine-tuned and attention mechanism integrated ResNet101 model achieved an accuracy of 0.935, precision of 0.93, recall of 0.94 and an f1-score of 0.93.

INDEX TERMS DenseNet121, endoscopy, gastrointestinal diseases, grid search, hyperparameter optimization, InceptionResnetV2, InceptionV3, MobileNetV2, ResNet50, ResNet101, Xception, attention mechanism.

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

I. INTRODUCTION

The gastrointestinal (GI) system is one of the most important systems in the body and is prone to a number of illnesses.

If the anomalies are not promptly detected and treated appropriately, they may develop into malignant cells. Colorectal cancer is the third most common cause of cancer-related deaths [1], and GI cancer (stomach, esophagus, and colon) accounts for around 2.8 million of these cases yearly, with a noteworthy mortality rate of about 65 percent [2].

Recent advances in imaging technologies have made it possible to see parts of the human body that were previously inaccessible. Endoscopy is one of these methods; to examine the GI tract, a tube with a camera is inserted [3]. Due to the high reliance on gastroenterologists' judgement in the endoscopic assessment of illness classification, results may vary from one expert to another [4]. Manually examining endoscopic data is time-consuming, demands significant concentration, and can occasionally be erroneous depending on the experience level of the clinicians involved. As a result, automatic recognition might be useful for expediting this process in terms of cost, duration, and classification accuracy [5].

Computer-assisted diagnosis is an essential step in the classification of large medical images. Taking one or more examination images as input, predicting them using the trained model, and then outputting the result produces a diagnostic result that indicates whether a particular disease is present and its severity. Patients obtain images using a variety of examination apparatus such as x-ray and ultrasound [6]. Additionally, endoscopic pictures and other pathological imaging are available when a physician is looking for sickness in the intestine.

In order to diagnose the patient, an endoscope is typically required to view the intestines' outer features. Polyps, inflammation, and malignancy are the three primary symptoms of various gastrointestinal lesions. Polyps are round or oval pedicled lumps that protrude from the large intestine's mucosal surface. Under colonoscopy, inflammation appears as significant hyperaemia, edema, erosion, and easy bleeding when touched, as well as pus and blood exudate on the surface of the intestinal mucosa [7]. A typical malignant tumour of the digestive system is cancer. The surface of the cancer is covered in necrosis and bleeding and protrudes into the intestinal lumen.

Numerous sophisticated intelligent classification techniques have surfaced in recent years, and classification accuracy has steadily increased [8]. The problem with image classification was that, the images were of poor quality and also the feature extraction techniques applied to different images were not explored fully. Deep learning uses human brain principles to analyse data and imitate the human brain for analytical learning. The most popular network model for deep learning is convolutional neural networks (CNN) [9]. It classifies image data based on image features and delivers the same into the network for training. To improve classification outcomes while training network architectures with deep learning, a large number of data sets are necessary in order to avoid model overfitting [10]. The criteria for doctors and patients throughout the examination process are particularly high due to the intricacy and inconvenience of endoscopy, which makes it challenging to gather colonoscopy data sets.

The motivation behind this research stems from the imperative to advance the field of medical image analysis, particularly in the context of endoscopy images. Endoscopy plays a pivotal role in diagnosing various medical conditions, and leveraging deep learning techniques can significantly enhance the accuracy and efficiency of image classification in this domain.

The use of transfer learning models, such as ResNet101, InceptionV3, InceptionResNetV2, Xception, DenseNet121, MobileNetV2, and ResNet50, is motivated by their proven success in a wide range of computer vision tasks. Fine-tuning these models through a meticulous grid search optimization of architectural hyperparameters aims to tailor their capabilities specifically for the nuanced challenges presented by endoscopy image classification. In this paper, classification of the endoscopic images into eight classes have been done using seven transfer learning models.

The key contributions made by this paper are:

1. This research explores the fine-tuning of seven transfer learning models—ResNet101, InceptionV3, InceptionResNetV2,Xception, DenseNet121, MobileNetV2, and ResNet50 by employing grid search optimization to adjust architectural hyperparameters. Different combinations of dropout layers and dense layers were explored for the multi-class classification of endoscopy images.
2. A comprehensive comparison of the performance of these fine-tuned transfer learning models was conducted, evaluating their accuracy, precision, recall, and F1-score. Among the models, the grid search fine-tuned ResNet101 exhibited superior performance, outperforming the other models.
3. Building on the success of the best-performing ResNet101 model, further optimization was carried out using fine-adjustment hyperparameters such as learning rate, batch size, and epochs. This optimization process involved grid search optimization to enhance the model's overall effectiveness.
4. To augment the classification performance of the optimized ResNet101 model, an attention mechanism was integrated. This fusion strategy leveraged the attention mechanism's ability to highlight key aspects, combining it with ResNet101's iterative feature map concatenation. The result was a more precise and interpretable categorization approach leading to an improvement in the overall classification performance of the model.

The paper structure comprises literature review in section II, a proposed methodology in section III, results and discussion in section IV, and a conclusion and future scope in section V.

II. LITERATURE REVIEW

Endoscopy is essential for diagnosing and treating GI tract diseases. The use of real-time AI image processing for diagnosing upper gastrointestinal cancers is still in experimental research and engineering. A comparison comparing endoscopic modalities, image counts, models, validation techniques, and outcomes for automated upper gastrointestinal cancer diagnosis and assessment was carried out on 65 studies. In order to improve performance, maturity, and potential for real-time upper gastrointestinal cancer diagnosis, this study compared and evaluated various AI approaches. According to the report, GI image processing for machine learning frequently uses support vector machines, or SVM. The study revealed that deep learning (DL) for GI image analysis frequently uses CNN-based supervised learning object detection models [11].

Recently, CADx (Computer Aided Diagnosis) systems are being used to reduce operator variation in conventional endoscopic procedures and provide guidance for precise illness diagnosis [9]. The training and testing feature sets are used by the CADx system to categorise GI tract illnesses. Results of classification tasks typically depend on techniques like preprocessing and image augmentation techniques that aid in the diagnosis of GI tract illnesses [10]. The system computation is sustained and improved via feature extraction [7]. A model called GastroNet was proposed which was obtained after fine-tuning of the YOLOv5 model. The model was used for determination of polyps and other abnormalities. The model comprised a single neural network for analyzing the whole image which were further split into parts. The probabilities were calculated for each of the part individually [12]. A lightweight deep Convolutional neural network was proposed to obtain the most important features from the endoscopic images. The obtained features were further reduced by the Cosine similarity-based method. The classification time was reduced due to the reduction in features [13]. For the KVASIR dataset, Khan et al. [14] suggested a DL model for the identification and categorization of gastrointestinal tract (GIT) anomalies. For the KVASIR dataset, the accuracy was 86.4%. Edge removal, contrast enhancement, filtering, color mapping, scaling, and color mapping are all provided for each image by MAPGI, an automated modular preprocessing framework for images of the gastrointestinal tract. Gamma correction values for images are generated automatically, adjusting mean pixel values within the range of 0–255 to 90±1. The Kvasir dataset is used to train three state-of-the-art neural networks: Inception-ResNet-v2, Inception-v4, and NASNet. Validation data is used to compare the three networks. Each example uses 15% for validation and 85% for training from the Kvasir dataset photographs [15].

Computer-aided diagnosis algorithms have produced promising results in medical imaging in the last several years [16], [17]. The research showed that the detection and classification of gastrointestinal tract illnesses has often been accomplished by automatic techniques based on deep learning and handcrafted models. Using the KVASIR dataset, Liu et al. [18] assessed the GI sickness recognition system. Six visual elements were combined with Haralick features and Local Binary Patterns to create the image texture. Following feature selection, kernel discriminant analysis and logistic regression were used to train the model. The obtained F1-score was 0.75. Using Bidirectional Marginal Fisher Analysis (BMFA), the author of [19] retrieved picture features and fed them to SVM for classification.

On KVASIR, [20] used transfer learning with data augmentation. After the dataset was refined using a pre-trained network called InceptionV3, the accuracy of the model was found to be 91.5%. CNN-based ulcer, erosion, and polyp categorization for stomach precancerous anomalies was presented by Agrawal et al. [21]. SqueezeNet with iterative reinforcement learning decreased the size and computing time of the model. 88.90% accuracy was attained overall. Good results were obtained in [22] when features were extracted using Inception V3 and VGGNet pre-trained models on the ImageNet dataset. SVM was then used to categorize the features. Pogorelov et al. [23] experimented with 17 different approaches, but the pre-trained ResNet50 and Logistic Model Tree (LMT) classifier yielded the best accuracy. The study's objectives were to identify eight classes as disease conditions, medical procedures, or anatomical landmarks while minimizing model performance and computation time and resources [23]. It was suggested to use a group of pre-trained models, including DenseNet201, InceptionV3, and ResNet50, to reliably classify endoscopic images. The accuracy of the ensemble model was 0.929 [24].

In this paper, seven grid search fine-tuned transfer learning models have been employed for classification of endoscopic images into eight classes and the best results were obtained by the ResNet101 model.

III. PROPOSED METHODOLOGY

Figure 1 shows the proposed methodology for the classification of the endoscopic images into eight classes of gastrointestinal abnormalities. Firstly, the preprocessing of KVASIR dataset was performed in which the images were resized into a size of 224 X 224. This size is required for the functioning of pre-trained models. The seven selected TL models are: ResNet101 [25], InceptionV3 [26], InceptionResNetV2 [27], Xception [28], DenseNet121 [29], MobileNetV2 [30] and ResNet50 [31]. The architectural hyperparameters like number of dense layers, dropout rate and activation function have been selected by grid search optimization for tuning of TL models. These new set of layers have been added to the seven TL models through grid search optimization. The optimized layers include three dense layers of size 1024, 512 and 256 respectively. A dropout layer of value of 0.5 succeeds these dense layers. These layers are further followed by a dense layer of size 128 and a dropout layer of 0.5. Finally, the last layer is a dense layer of size 8 and softmax activation function. After the addition of these layers, all the seven grid search fine-tuned models were compared and it was inferred that the grid search fine-tuned ResNet101 model performed best for these hyperparameters.

Similarly, fine-tuned hyperparameters like batch size, learning rate, optimizer and number of epochs have been selected by grid search optimization for fine-tuning of the ResNet101 models. The best fine-tuned hyperparameters i.e., batch size of 32, learning rate of 0.001, SGD optimizer and 40 epochs were selected by grid search optimization for the Resnet101 model. Finally, the most efficient model i.e., the grid search fine-tuned ResNet101 model classifies the images into the eight categories of GI diseases.

A. INPUT DATASET

In this paper, the KVASIR dataset [22] has been used which comprised endoscopic images of the human gastrointestinal tract. This dataset has been used for the detection of different types of abnormalities. The dataset comprises 4000 images of the gastrointestinal tract which is divided into 8 different classes (different anomalies) and each class has 500 images. The eight classes (3 normal and 5 diseases) are dyed-lifted-polyps, normal-cecum, normal-pylorus, normal-z-line, esophagitis, polyps, ulcerative colitis and dyed-resection-margins as shown in figure 2. The dataset has been divided into 2240 training images, 960 validation images and 800 test images.

B. IMAGE RESIZING

Pre-processing [32] is a crucial step in image processing as it improves the endoscopic images' characteristics and eliminates the image's superfluous data. The images used in this paper were of varying sizes. Hence, image resizing has been performed as the data preprocessing step. Moreover, image resizing in transfer learning models helps to maintain consistency and compatibility. Hence, the endoscopic images of the GI diseases have been resized to 224*224.

C. GRID SEARCH OPTIMIZATION OF HYPERPARAMETERS

Grid search is an optimization technique [33] which helps in automating the process of finding the hyperparameters i.e., it helps one to select the optimum hyperparameters to optimize problems through a given alternative parameter list. This technique is generally used to optimize deep learning models so that accurate results are obtained.

Hyperparameters [34] are variables that are set by the user before to the training process rather than being learned by a machine learning algorithm during training. They have significant impact over the performance and behaviour of the machine learning system and have control over a number of training process variables. In most cases, hyperparameters are pre-set and held constant throughout the training process. Architectures for CNN models are fairly complex and contain a lot of hyper-parameters. These hyperparameters can typically be divided into two categories: fine adjustment hyperparameters and architectural hyperparameters. Here, both type of hyperparameters i.e. architectural hyperparameters and fine-adjustment hyperparameters are tuned with the help of grid search optimization technique.

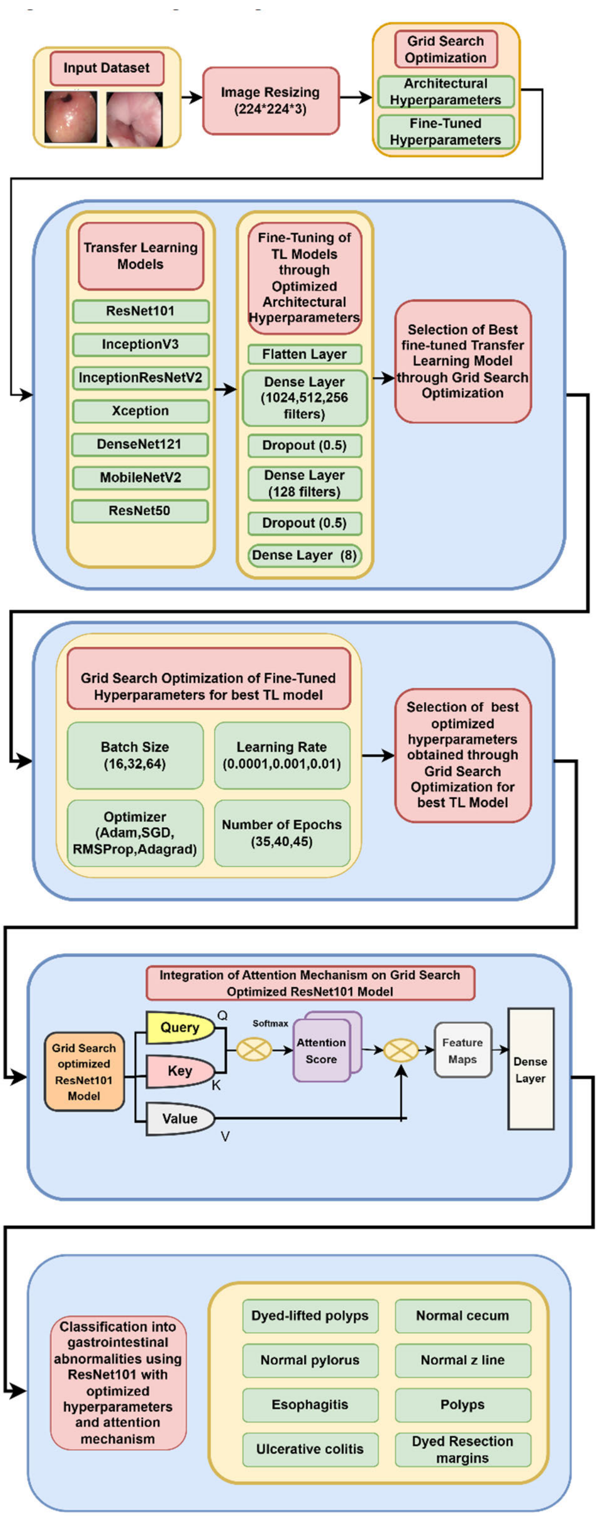


FIGURE 1. Proposed Methodology for classification of GI diseases.

1) ARCHITECTURAL HYPERPARAMETERS OPTIMIZATION

Architectural hyper-parameters include the number of dense layers,, the dropout rate and activation function. For a fully connected or dense layer, every neuron in the layer has a

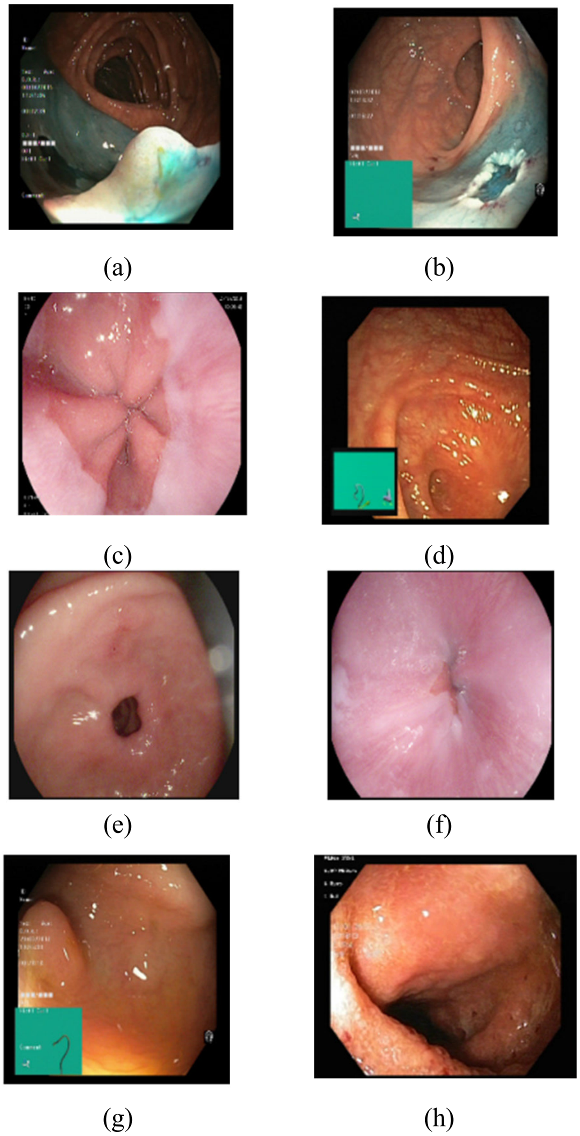


FIGURE 2. (a) Dyed-lifted Polyps (b) Dyed-resection margins (c) Esophagitis (d) Normal cecum (e) Normal -Pylorus (f) Normal z-line (g) Polyps (h) Ulcerative colitis.

connection with every neuron in the layer preceding it. Dense layers at the network's final end transform the gathered or flattened outputs into the required output.

The drop-out approach can be used to prevent overfitting in deep neural networks. By arbitrarily setting each update's output to zero, the dropout training method ''drops out'' the neurons. The dropout therefore lowers the interdependencies between the neurons by forcing the network to develop more stable and generalizable features. Spreading the weights over more neurons makes the network more robust to noise and improves its ability to generalize to new input.

In this paper, optimization of the architectural hyper-parameters is performed first using the steps as shown in table 1. The grid search algorithm essentially attempts every possible parameter value amalgamation and gives the output for the one having the greatest accuracy. Three parameters are required for optimization to get the best accuracy as shown in table 1.

TABLE 1. Grid search for architectural hyperparameters optimization.

Grid Search for opimization of architectural hyperparameters	
Step 1 : **Set a 3-dimensional grid for optimization of three hyperparameters**	Number of dense layers Dropout Rate Activation function
Step 2 : **Set a potential value interval that corresponds to each dimension**	Number of dense layers = [1,2,3,4,5] Dropout Rate = [0.3,0.4, 0.5] Activation function = [Tanh, ReLu, Leaky ReLu]
Step 3 : **See to all the candidate combinations and select the one with best optimized overall accuracy**	e.g. C1 = (3,0.3, Tanh)- Accuracy = 0.80 C2= (4,0.4, Leaky reLu)- Accuracy = 0.82 C3= (5,0.5, ReLu) – Accuracy=0.90

2) FINE-ADJUSTMENT HYPERPARAMETERS OPTIMIZATION

Fine-adjustment hyper-parameters include optimizers, batch size [35], learning rate, and number of epochs. A crucial hyperparameter that shouldn't be either too large or too tiny is the learning rate (LR). It is employed to determine the suggested models' rate of learning. If the LR is too little or too high, the model would take much longer to obtain the lowest loss because overshooting the low loss areas is possible. Steps used for optimization of the fine adjustment hyper-parameters are shown in table 2. The optimization of four parameters need to be done for obtaining the best accuracy.

TABLE 2. Grid search for fine-adjustment hyperparameters optimization.

Grid Search for optimization of fine-adjustment hyperparameters	
Step 1 : **Set a grid of 4-dimensions for optimization of 4 hyperparameters**	Batch size Learning Rate Optimizer Number of Epochs
Step 2 : **Set a potential value interval that corresponds to each dimension**	Batch Size = [16,32,64] Learning rate = [0.0001,0.001,0.01] Optimizer = [Adam , SGD, RMSProp, Adagrad] Number of Epochs = [35,40,45]
Step 3 : **See to all the candidate combinations and select the one with best optimized overall accuracy**	e.g. C1= (16, 0.0001, Adam,35) = 0.84 C2= (16, 0.001, SGD,45)=0.88 C3=(32,0.001,SGD,40)=0.90

The table 3 summarizes the optimal hyperparameter values obtained through a comprehensive grid search for transfer learning models. The goal was to fine-tune these models for enhanced performance in a specific application. The selected hyperparameters, their respective ranges, and the optimized values are detailed below:

The optimized values reflect the configurations that yielded the best results during the grid search optimization

TABLE 3. Optimum results obtained through grid search for transfer learning models.

Hyperparameters	Hyperparameter Range	Optimized Values
Number of Dense Layers	[1,2,3,4,5]	5
Dropout Rate	[0.3,0.4,0.5]	0.5
Activation Function	[Tanh, ReLu, Leaky ReLu]	ReLu
Batch Size	[16,32,64]	32
Learning Rate	[0.0001,0.001,0.01]	0.001
Optimizer	[Adam,SGD,RMSProp,Adagrad]	SGD
Number of epochs	[35,40,45]	40

process. These results indicate that, for the multi-class gastrointestinal disease classification, a transfer learning model with five dense layers, a dropout rate of 0.5, ReLu activation function, a batch size of 32, a learning rate of 0.001, trained with the SGD optimizer over 40 epochs, demonstrated optimal performance. These values contribute to the efficient convergence of the model during training, striking a balance between computational efficiency and the model's ability to learn from the data.

The fine-tuning process helps the model in adapting the particular characteristics and requirements of the endoscopy image classification task. By adjusting the number of dense layers, dropout rates, and other hyperparameters, the model is tailored to extract relevant features and patterns from the medical images. These findings provide valuable insights for configuring transfer learning models, emphasizing the importance of fine-tuning specific hyperparameters to achieve optimal results. The conclusions drawn from this grid search contribute to the establishment of best practices in model optimization for the gastrointestinal disease classification.

D. GRID SEARCH ARCHITECTURAL HYPERPARAMETERS TUNED TRANSFER LEARNING MODELS

Transfer learning uses the knowledge acquired from completing a source task and applies it to a target task instead of building a model from scratch, typically with fewer data or training resources needed. Figure 3 shows the architectural hyperparameters tuned fully connected head that has been added to the TL models. The layers were optimized through grid search optimization. The transfer learning models are followed by the flatten layer, three dense layers with 1024,512 and 256 number of filters respectively. These layers are followed by a dropout layer of 0.5. The dropout layer is further followed by a dense layer of 128 filters, dropout layer of 0.5 and a dense layer that classifies the endoscopic images into the eight GI disease classes.

Table 4 presents the description with regard to the architecture of the seven TL models. The ResNet101 models has 101 layers and 44.5 million parameters. The InceptionV3 model comprises of 42 layers and 24 million parameters. The InceptionResNetV2 model is the densest with 164 layers and 56 million parameters. The Xception model consists of

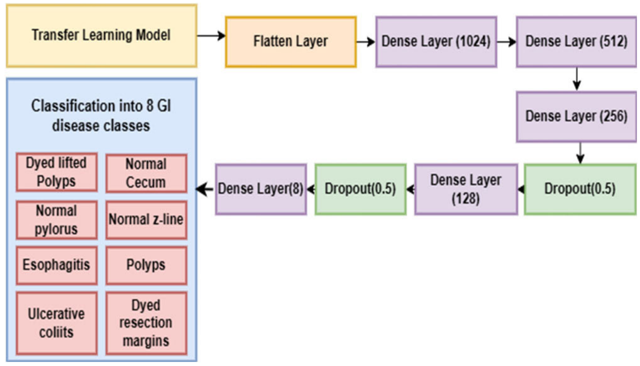


FIGURE 3. Architectural hyperparameters tuned fully connected head.

71 layers and 22.8 million parameters. The DenseNet121 model has 121 layers and 8 million parameters. The ResNet50 model is 50 layers deep with 25.6 million parameters while the MobileNetV2 model has 53 layers with the least number of parameters i.e. 3.4 million parameters. The detailed description of all the models has been given in the following sections.

TABLE 4. Architectural description of transfer learning models.

Model Name	Number of Layers	Parameters (Millions)	Input Layer Size
ResNet101	101	44.5	(224,224,3)
IncpetionV3	42	24	(224,224,3)
InceptionResNetV2	164	56	(224,224,3)
Xception	71	22.8	(224,224,3)
DenseNet121	121	8	(224,224,3)
MobileNetV2	53	3.4	(224,224,3)
ResNet50	50	25.6	(224,224,3)

1) GRID SEARCH FINE-TUNED RESNET101 MODEL

ResNet-101 is a deep CNN model that belongs to the ResNet (Residual Network) family [36]. This model includes pooling layers, convolutional layers and dense layers. This model also includes ReLU activation functions, batch normalization and shortcut connections. The training of very deep networks can be done due to the presence of shortcut connections, often referred to as skip connections, which allow the gradient to pass directly through the network without fading too rapidly.

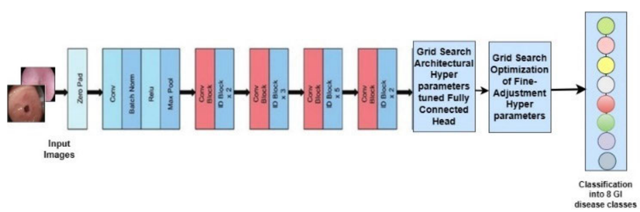


FIGURE 4. Grid search fine-tuned ResNet101 model.

Here, the ResNet101 model has been modified, firstly by adding fine-tuned architectural hyperparameter fully connected head and second, by addition of fine-adjustment hyperparameters as shown in figure 4. These hyperparameters have been obtained by grid search optimization. One flatten layer, five dense layers and two dropout layers have been added in the fully connected head and also the model is trained with optimized fine-adjustment parameters

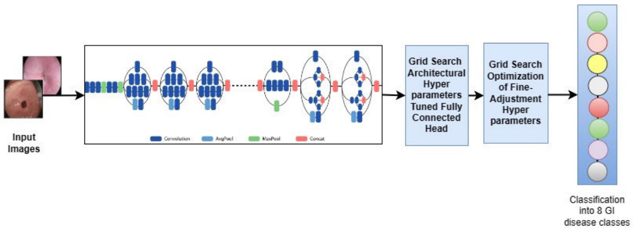


FIGURE 5. **Grid search fine-tuned InceptionV3 model.**

2) GRID SEARCH FINE-TUNED INCEPTIONV3 MODEL

In this model, inception modules [37] are used to increase the productivity and efficacy of deep CNNs. The third iteration of Inception models, denoted by the "V3" in InceptionV3, was created with the goal of improving accuracy while yet keeping a manageable computing cost. The utilisation of so-called "Inception modules," which are created to effectively capture multi-scale information, is the main innovation in InceptionV3. These modules use pooling techniques and numerous convolutional filters of various sizes (e.g., 1×1, 3×3, and 5×5) to capture various patterns at varying spatial resolutions. As a result, the network can effectively learn complicated characteristics by extracting both fine-grained and broad contextual information. To decrease the number of huge convolutions, InceptionV3 also adds other methods including batch normalisation and factorization.

Here, the InceptionV3 model has been modified, firstly by adding fine-tuned architectural hyperparameter fully connected head and second, by addition of fine-adjustment hyperparameters as shown in figure 5. These hyperparameters have been obtained by grid search optimization. One flatten layer, five dense layers and two dropout layers have been added and also the model is trained with optimized fine-adjustment parameters.

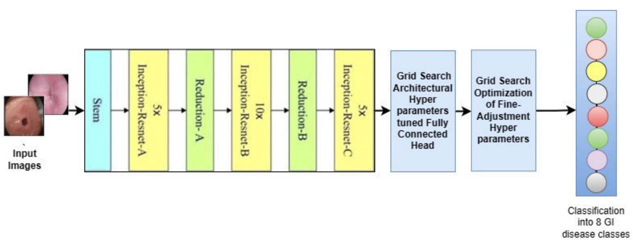


FIGURE 6. **Grid search fine-tuned InceptionResNetV2 model.**

3) GRID SEARCH FINE-TUNED INCEPTIONRESNETV2 MODEL

Both the Inception and ResNet models' components are combined in the InceptionResNetV2 model. Utilizing the benefits of both the Inception and ResNet architectures is the major goal of InceptionResNetV2 [38]. Inception modules are employed to effectively capture multi-scale information, and residual connections help in eliminating the drawback of vanishing gradient. Compared to InceptionV3, Inception-ResNetV2 has a more intricate network topology and more layers. It is made up of reduction blocks that use max pooling to minimise the spatial dimensions and an array of inception blocks with residual connections. The network architecture makes it more powerful for a variety of computer vision tasks by facilitating better feature extraction and representation learning. InceptionResNetV2 is computationally more intensive because of its depth and complexity.

In this paper, the InceptionResNetV2 model has been modified, firstly by adding fine-tuned architectural hyperparameter fully connected head and second, by addition of fine-adjustment hyperparameters as shown in figure 6. These hyperparameters have been obtained by grid search optimization. One flatten layer, five dense layers and two dropout layers have been added and also the model is trained with optimized fine-adjustment parameters.

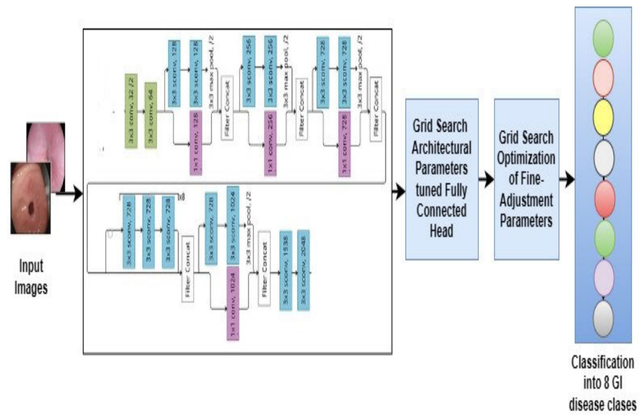


FIGURE 7. **Grid search fine-tuned xception model.**

4) GRID SEARCH FINE-TUNED XCEPTION MODEL

The Inception design served as inspiration for the Xception model, which adds a novel idea known as "depthwise separable convolutions" to increase the model's effectiveness and computing efficiency. The primary goal of Xception [39] is to boost deep neural networks' effectiveness and efficiency. It attempts to preserve or even improve the model's accuracy while minimising the amount of computations and parameters. In order to do this, Xception divides the conventional convolutional process into the depthwise convolution and the pointwise convolution steps. Each input channel is given its own independent convolutional filter during the depthwise convolution step by Xception. Without any mixing, it records spatial correlations between the input channels. Comparing this procedure to conventional convolutions, where each filter interacts with each input channel, the number of parameters is drastically reduced.

Here the Xception model has been modified, firstly by adding fine-tuned architectural hyperparameter fully connected head and second, by addition of fine-adjustment hyperparameters as shown in figure 7. These hyperparameters have been obtained by grid search optimization. One flatten layer, five dense layers and two dropout layers have been added and also the model is trained with optimized fine-adjustment parameters.

5) GRID SEARCH FINE-TUNED DENSENET121 MODEL

The model family that includes DenseNet-121 [40] places a high emphasis on feature reuse and promotes direct connections between network levels. The utilisation of dense blocks, which are made up of numerous densely connected layers, is the main innovation in DenseNet. Every layer present in the dense block obtains the characteristic maps from all preceding layers and transfers the same feature maps to all the layers that follow it. The network's dense connection architecture facilitates significant reuse of features and direct information flow through the complete network. Both of these factors improve parameter efficiency and lessen the chance of disappearing gradients.

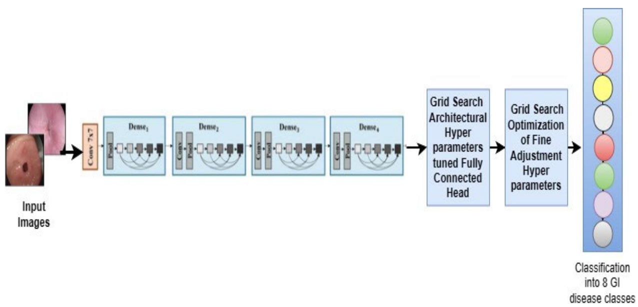


FIGURE 8. Grid search fine-tuned DenseNet121 model.

In this paper, the DenseNet121 model has been modified, firstly by adding fine-tuned architectural hyperparameter fully connected head and second, by addition of fine-adjustment hyperparameters as shown in figure 8. These hyperparameters have been obtained by grid search optimization. One flatten layer, five dense layers and two dropout layers have been added and also the model is trained with optimized fine-adjustment parameters.

6) GRID SEARCH FINE-TUNED MOBILENETV2 MODEL

The MobileNetV2 [41] deep learning model is effective and portable, making it well suited for mobile and embedded devices. This model has a small size, better computational efficiency, and is more accurate, these factors being MobileNetV2's key objective. Depthwise separable convolutions are widely used in MobileNetV2. These are made up of a pointwise convolution followed by a depthwise convolution which greatly lowers the amount of parameters and calculations while maintaining the model's expressive power. The idea of inverted residuals was also introduced by this model, in which the input and output layers have a higher dimension

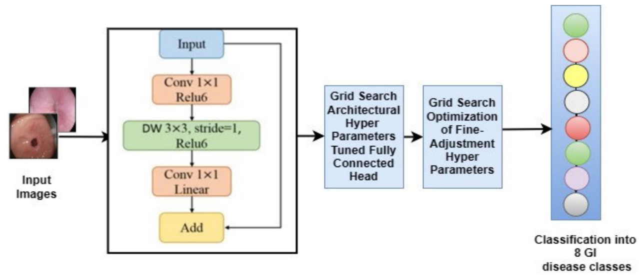


FIGURE 9. Grid search fine-tuned MobileNetV2 model.

than the intermediate levels. This lowers the model's computational expense. In the middle, linear bottlenecks are also utilized.

In this paper, the MobileNetV2 model has been modified, firstly by adding fine-tuned architectural hyperparameter fully connected head and second, by addition of fine-adjustment hyperparameters as shown in figure 9. These hyperparameters have been obtained by grid search optimization. One flatten layer, five dense layers and two dropout layers have been added and also the model is trained with optimized fine-adjustment parameters.

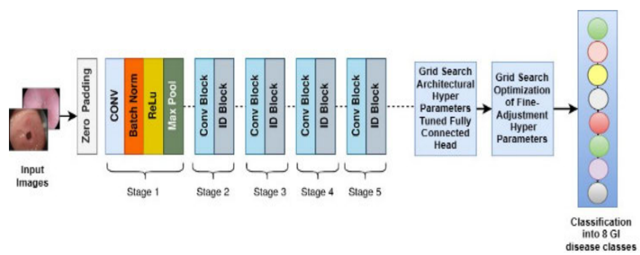


FIGURE 10. Grid search fine-tuned ResNet50 model.

7) GRID SEARCH FINE-TUNED RESNET50 MODEL

The issue of disappearing gradients in extremely deep networks is addressed by ResNet-50 [42]. ResNet-50 employs the idea of residual blocks, which are intended to make it possible to train very deep networks without encountering the degradation issue. Due to the difficulties of training very deep networks, the degradation problem emerges when adding more layers to a neural network causes a decline in accuracy.

Here, the ResNet50 model has been modified, firstly by adding fine-tuned architectural hyperparameter fully connected head and second, by addition of fine-adjustment hyperparameters as shown in figure 10. These hyperparameters have been obtained by grid search optimization. One flatten layer, five dense layers and two dropout layers have been added and also the model is trained with optimized fine-adjustment parameters.

E. GRID SEARCH OPTIMIZED RESNET101 MODEL WITH ATTENTION MECHANISM

The proposed methodology introduces an attention mechanism within the grid search optimized ResNet101 model. The methodology emphasizes the potential of attention mechanisms in enhancing the model's ability to discern relevant patterns, contributing to the interpretability and performance of ResNet101 model and has been shown in figure 11. The Attention Layer is an essential component of the model that is designed to dynamically weigh input features, allowing the model to focus on relevant information during training.

The Attention Layer is constructed with trainable weights (W_query, W_key, W_value) that enable the model to adaptively learn relationships between features. By employing matrix multiplication and softmax activation, the attention mechanism calculates attention scores, highlighting the significance of different input elements. Integrating this layer into the model contributes to the broader exploration of attention mechanisms, which have shown promise in capturing contextual information and improving model accuracy. In this case, it creates three trainable weight matrices (W_query, W_key, W_value) that will be used to transform the input data during the attention mechanism. It performs matrix multiplications with the input data x using the learned weights (W_query, W_key, W_value) to create query (q), key (k), and value (v) matrices. Then the attention scores are computed by taking the matrix multiplication of q and k. The softmax activation is applied to the attention scores to obtain a probability distribution. Finally, the attention-weighted sum is calculated using the softmax scores and the value matrix (v). Subsequently, dense layers with rectified linear unit (ReLU) activations and dropout are employed for classification.

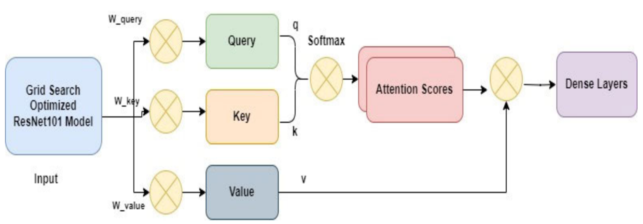


FIGURE 11. Attention mechanism on grid search fine-tuned ResNet101 model.

IV. RESULTS AND DISCUSSION

This section includes the results of the seven grid search fine-tuned transfer learning models that are ResNet101, InceptionV3, InceptionResNetV2, Xception, DenseNet121, MobileNetV2 and ResNet50.

A. EVALUATION OF THE GRID SEARCH FINE-TUNED TRANSFER LEARNING MODELS

The training performance of all the seven grid search fine-tuned transfer learning models in terms of training accuracy, validation accuracy, training loss and validation loss at 1st, 39th and 40th epoch has been given in table 5. At the 40th epoch, highest training accuracy of 0.9991 and validation accuracy of 0.8719 is obtained by the ResNet101 model. Lowest training loss of 0.0054 and validation loss of 0.6425 is also obtained by the ResNet101 model at the 40th epoch. Figure 12 shows the accuracy curves for all the transfer learning models.

TABLE 5. Training performance of all grid search fine-tuned transfer learning models.

Model Name	Epoch	Training Accuracy	Validation Accuracy	Training Loss	Validation Loss
DenseNet121	1	0.3705	0.6260	1.7555	1.0145
	39	0.9960	0.8219	0.0105	1.0543
	40	0.9955	0.7990	0.0162	1.3594
ResNet50	1	0.4205	0.4385	1.7242	1.5449
	39	0.9991	0.8833	0.0032	0.9169
	40	0.9897	0.8292	0.0449	1.1024
MobileNet V2	1	0.3210	0.1750	1.8478	2.0676
	39	0.9937	0.7521	0.0228	1.4815
	40	0.9964	0.7937	0.0161	1.2306
Xception	1	0.3103	0.4281	1.7632	1.4013
	39	0.9973	0.8406	0.0085	0.8992
	40	0.9982	0.8250	0.0066	1.0193
InceptionResNetV2	1	0.2763	0.4135	1.8637	1.6321
	39	0.9987	0.8562	0.0091	0.7395
	40	0.9978	0.8562	0.0097	0.6969
InceptionV3	1	0.3129	0.4167	1.8131	1.5029
	39	0.9978	0.7719	0.0116	1.2942
	40	0.9942	0.8042	0.0248	0.9949
ResNet101	1	0.4299	0.6187	1.7563	1.0050
	39	0.9982	0.8604	0.0054	0.7531
	40	0.9991	0.8719	0.0054	0.6425

Figure 13 presents the comparison of the seven grid search fine-tuned transfer learning models in terms of the confusion matrix parameters. The seven grid search fine-tuned TL models are: DenseNet121, ResNet50, MobileNetV2, Xception, InceptionResNetV2 and ResNet101. The models have been evaluated at an optimized arrchitectural and fine-tuned hyperparameters.

The lowest accuracy of 0.84 is obtained by the DensNet121 model while the highest accuracy of 0.90 is obtained by the ResNet101 model. An accuracy of 0.89 is obtained by both the InceptionV3 and InceptionResNetV2 model. The lowest precision of 0.83 is obtained by the DenseNet121 model while ResNet101 model achieved the highest precision of 0.92. ResNet101 also obtained the highest recall of 0.92 and highest f1-score of 0.91. From figure 12 and table 5, it can be concluded that the ResNet101 is the best performing model for the optimized hyperparameters.

B. BEST PERFORMING GRID SEARCH FINE-TUNED MODEL – RESNET101

Figure 14 shows the confusion matrix for the best performing TL model- ResNet101 at the optimized hyperparameters. From the figure, it can be seen that the normal pylorus class is the best performing class out of the eight classes of GI diseases.

Table 6 displays the assessment measures for a classification model based on a collection of imagery of the digestive tract displaying various GI disorders. According to the model, there should be eight separate classes. '

According to the performance measures, the grid search fine tuned ResNet101 approach ssems to be working

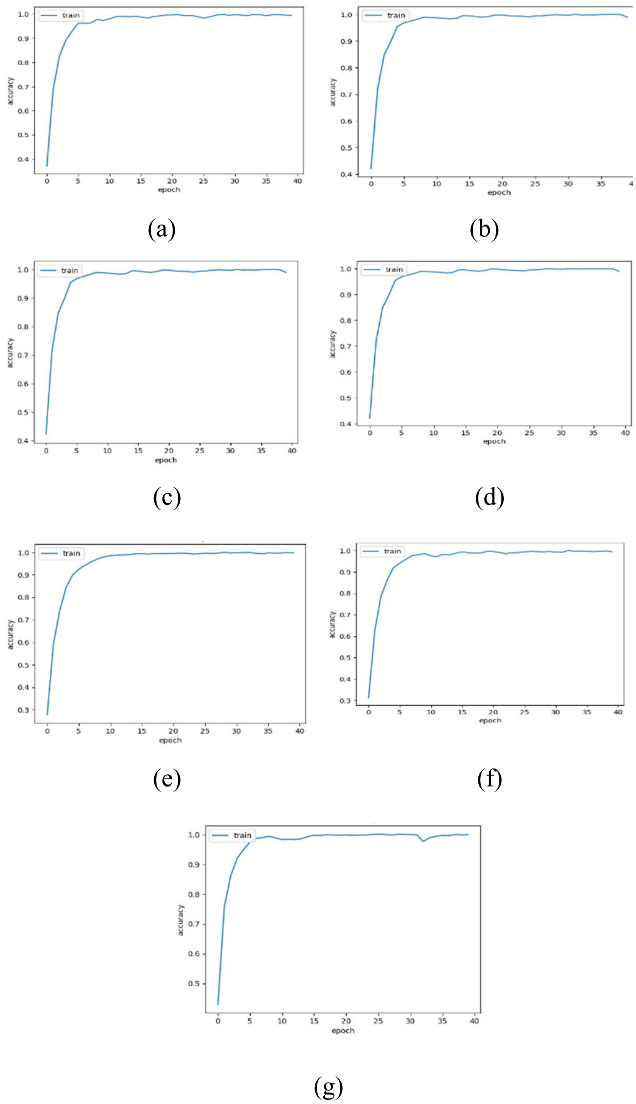


FIGURE 12. Accuracy curves of the TL models (a) DenseNet121 (b) ResNet50 (c) MobileNetV2 (d) Xception (e) InceptionResNetV2 (f) InceptionV3 (g) ResNet101.

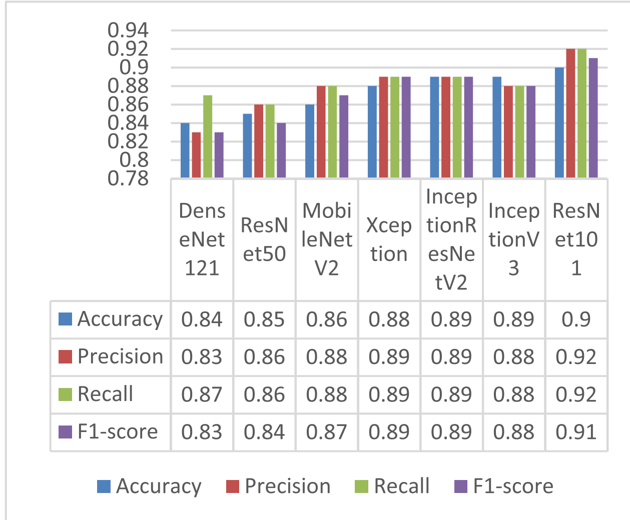


FIGURE 13. Comparison of grid search fine-tuned TL models in terms of confusion matrix parameters.

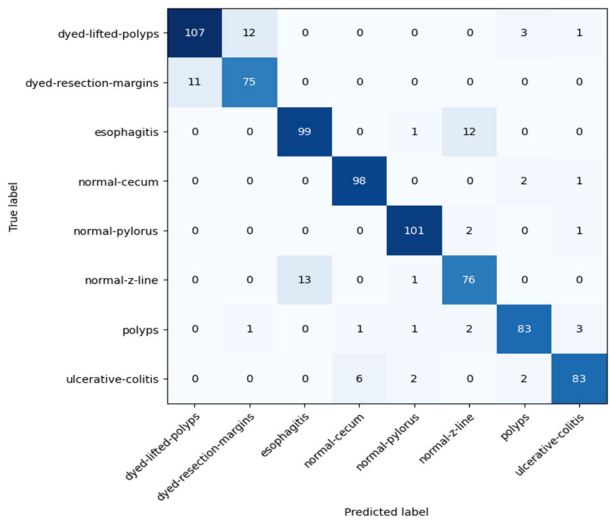


FIGURE 14. Confusion matrix for ResNet101 Model for the optimized hyperparameter combination.

effectively for few particular classes, such as ''Normal cecum'' and ''normal pylorus,'' where the model has achieved good values of precision, recall, and F1-scores. When applied to other classes, such as ''dyed-resection-margins'' and ''normal z-line,'' the model performs around average because these classes have lower values of the assessment measures.

Additionally, it's important to note that few of the classes have higher recall than precision, whilst some of the others have better precision value than recall. For instance, whereas the ''normal cecum'' class has higher recall than accuracy, the ''normal pylorus'' class has both good precision and recall. This suggests that the model might be more effective at detecting some classes of the GI diseases than the rest of the classes.

For optimal architectural and fine-tuned hyperparameters, the precision, recall, and f1-score for each of the eight classes are shown in table 6. The regular pylorus class had the highest precision, 0.95. The typical pylorus class achieved the highest recall (0.97) and f1-score (0.96).

C. ANALYSIS OF GRID SEARCH OPTIMIZED RESNET101 MODEL WITH ATTENTION MECHANISM

Figure 15 shows the confusion matrix for the grid search optimized ResNet101 model integrated with attention mechanism. An overall accuracy of 0.935 was obtained.

Table 7 displays the assessment measures for Grid search optimized ResNet101 model with attention mechanism based on a collection of imagery of the digestive tract displaying various GI disorders. There are total eight separate classes and the table displays the performance metrics for each category. '

TABLE 6. Confusion matrix parameters at optimized hyperparameters.

Class	Precision	Recall	F1-score
Dyed-lifted-polyps	0.91	0.87	0.89
Dyed-resection-margins	0.85	0.87	0.86
Esophagitis	0.88	0.88	0.88
Normal Cecum	0.93	0.97	0.95
Normal pylorus	0.95	0.97	0.96
Normal Z-line	0.83	0.84	0.84
Polyps	0.92	0.91	0.92
Ulcerative- colitis	0.93	0.89	0.91

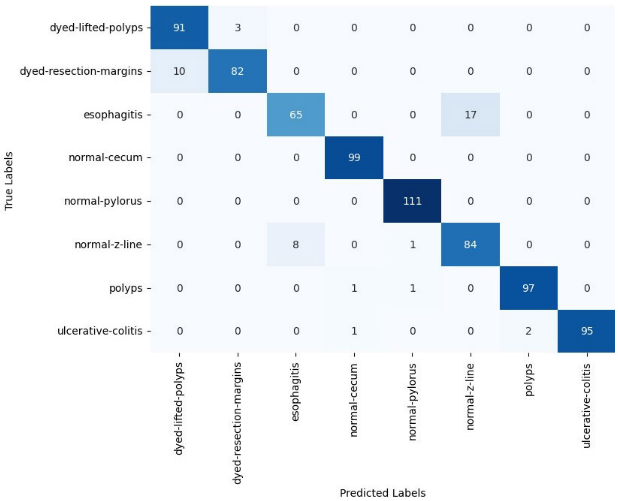


FIGURE 15. Confusion matrix for ResNet101 model for the optimized hyperparameter combination with attention mechanism.

According to the performance measures, the grid search fine tuned ResNet101 model integrated with attention mchanism works effectively for almost all the classes like "polyps", "Normal cecum" and "normal pylorus," where the model has achieved good values of precision, recall, and F1-scores. When applied to other classes, such as "esophagitis" and "dyed-resection margins," the model performs around average because these classes have lower values of the assessment measures.

Additionally, it's important to note that few of the classes have higher recall than precision, whilst some of the others have better precision value than recall. For instance, whereas the "dyed lifted polyps" class has higher precision than recall, the "normal pylorus" category has greater sensitivity than precision. This suggests that the model might be more effective at detecting some classes of the GI diseases than others, and that the ratio of precision to recall may change which is dependent on a specific class.

For optimal architectural and fine-tuned hyperparameters, integrated with attention mechanism, the performance metrics for all the eight classes are shown in table 7. The normal cecum class had the highest precision, 0.99 and highest f-score of 0.99. The polyps class, normal cecum and normal pylorus achieved the highest recall of 0.98.

TABLE 7. Confusion matrix parameters at optimized hyperparameters with attention mechanism.

Class	Precision	Recall	F1-score
Dyed-lifted-polyps	0.97	0.90	0.93
Dyed-resection-margins	0.89	0.95	0.92
Esophagitis	0.79	0.89	0.84
Normal Cecum	0.99	0.98	0.99
Normal pylorus	0.97	0.98	0.98
Normal Z-line	0.90	0.82	0.86
Polyps	0.96	0.98	0.97
Ulcerative- colitis	0.97	0.97	0.97

D. CLASSIFICATION AND MISCLASSIFICATION RESULTS

Figure 16 depicts the categorization results of different gastrointestinal tract disease classes. Figure 16 (a) depicts the Actual and Predicted class as "ulcerative colitis", Figure 16 (b) as "esophagitis", and Figure 16 (c) as "normal pylorus".

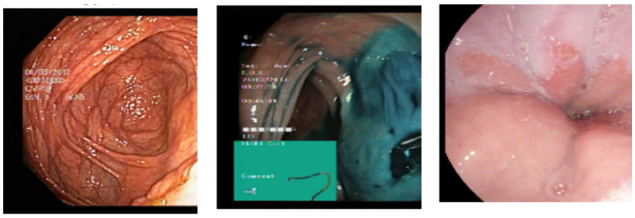
Actual : Ulcerative colitis Predicted: Ulcerative colitis (a)	Actual:Esophagitis Predicted : Esophagitis (b)	Actual : Normal pylorus Predicted : Normal pylorus (c)

FIGURE 16. Classification results.

Figure 17 depicts the misclassification results for various classes of gastrointestinal tract diseases. Figure 17 (a) depicts the Actual class as "normal-cecum" and the predicted class as "polyps". Figure 17 (b) depicts the Actual class as "dyed-lifted polyps" and the Predicted class as "dyed resection margins", and Figure 17 (c) depicts the Actual class as "normal z-line" and the Predicted class as "esophagitis"

E. COMPARISON WITH THE STATE-OF-THE -ART MODELS

Table 8 compares of the presented model with the state-of-the-art models [43], [44]. An ensemble model obtained an accuracy of 0.929 for 8000 images, [24]. An accuracy of 0.901 was obtained by a CNN for 8000 images of 720∗576 size [45]. An accuracy of 0.87 was obtained by another CNN [46]. The AlexNet model attained an accuracy of 0.85 for 4000 images [47]. CNN model [48] attained an accuracy of 0.88 for 825 images with an image size of 525∗525. The Support Vector Machine (SVM) model [49] achieved an accuracy of 0.88 for 8000 images. An accuracy of 0.78 was obtained by linked color imaging [50] for 208 images. The ResNet50 model [51] achieved an accuracy of 0.87 for 785 images. Another CNN [52] a4ttained an accuracy of 0.859 and a combination of CNN and SVM [53] achieved an accuracy of 0.85. The proposed model attained an



Actual : Normal-cecum Predicted : Polyps (a)

Actual : Dyed-lifted polyps Predicted : Dyed resection margins (b)

True : Normal z-line Predicted : Esophagitis (c)

FIGURE 17. Misclassification results.

accuracy of 0.935 for 4000 images when ResNet101 model was applied to the endoscopic images. s

TABLE 8. Comparison with the state-of-the-art models.

Reference No.	Year	No. of Images	Image Size	Technique	Accuracy
[24]	2023	8000	224*224	Ensemble Model of DenseNet201,InceptionV3 and ResNet50	0.929
[45]	2022	8000	720*576	Convolutional Neural Network	0.901
[46]	2021	8000	720*576	Convolutional Neural Network	0.87
[47]	2020	4000	224*224	AlexNet	0.85
[48]	2018	825	525 *525	Convolutional neural network	0.88
[49]	2019	8000	720 * 576	Support Vector Machine	0.88
[50]	2019	208	64*64	Linked Color Imaging	0.78
[51]	2019	785	--	ResNet50	0.87
[52]	2017	1.4 million	227*227	Convolutional Neural Network	0.859
[53]	2015	180	32*32	Convolutional Neural Network + Support Vector Machine	0.85
Proposed Model	2024	4000	224*224	Grid search optimized ResNet101 with Attention Mechanism	0.935

V. CONCLUSION AND FUTURE WORK

This study offers a reliable framework for categorising the disorders of the GI tract in the Kvasir dataset. By assisting in early detection, deep learning algorithms might decrease the likelihood of acquiring malignant diseases while minimising the needless removal of benign tumours. Seven TL models i.e. ResNet101, InceptionV3, InceptionResNetV2, Xception, DenseNet121, MobileNetV2 and ResNet50 were employed in this paper. These models can help in directing the focus of doctors' to the most important parts of the endoscopic images that might have been missed. Grid search optimization has been performed to obtain the optimized values of the architectural as well as the fine-adjustment hyperparameters. The ResNet101 model performed the best at the best optimized hyperparameters. Attention mechanism was applied to the best optimized ResNet101 model and highest accuracy of 0.935 was obtained. In the future, a refinement may be achieved in the performance parameters by use of hybrid models. Also, the model could be applied to a different dataset to make the models more generalizable.

REFERENCES

[1] K. Saluja, A. Bansal, A. Vajpaye, S. Gupta, and A. Anand, "Efficient bag of deep visual words based features to classify CRC images for colorectal tumor diagnosis," in *Proc. 2nd Int. Conf. Advance Comput. Innov. Technol. Eng. (ICACITE)*, Apr. 2022, pp. 1814–1818.

[2] R. Pillai, N. Sharma, and R. Gupta, "Detection & classification of abnormalities in GI tract through MobileNetV3 transfer learning model," in *Proc. 14th Int. Conf. Comput. Commun. Netw. Technol. (ICCCNT)*, Jul. 2023, pp. 1–6.

[3] J.-B. Liu, J. Cao, A. Alofi, A. AL-Mazrooei, and A. Elaiw, "Applications of Laplacian spectra for n-prism networks," *Neurocomputing*, vol. 198, pp. 69–73, Jul. 2016.

[4] A. Khan, M. A. Gul, A. Alharbi, M. I. Uddin, S. Ali, and B. Alouffi, "Impact of lexical features on answer detection model in discussion forums," *Complexity*, vol. 2021, pp. 1–8, Apr. 2021.

[5] Y. Sasaki, R. Hada, and A. Munakata, "Computer-aided grading system for endoscopic severity in patients with ulcerative colitis," *Digestive Endoscopy*, vol. 15, no. 3, pp. 206–209, Jul. 2003.

[6] P. Wang, T. M. Berzin, J. R. Glissen Brown, S. Bharadwaj, A. Becq, X. Xiao, P. Liu, L. Li, Y. Song, D. Zhang, Y. Li, G. Xu, M. Tu, and X. Liu, "Real-time automatic detection system increases colonoscopic polyp and adenoma detection rates: A prospective randomised controlled study," *Gut*, vol. 68, no. 10, pp. 1813–1819, Oct. 2019.

[7] M. A. Khan, M. S. Sarfraz, M. Alhaisoni, A. A. Albesher, S. Wang, and I. Ashraf, "StomachNet: Optimal deep learning features fusion for stomach abnormalities classification," *IEEE Access*, vol. 8, pp. 197969–197981, 2020.

[8] M. Mackiewicz, "Capsule endoscopy-state of the technology and computer vision tools after the first decade, new techniques in gastrointestinal endoscopy," in *Proc. ISBN*, vol. 4, 2011, pp. 953–978.

[9] T. Aoki, A. Yamada, K. Aoyama, H. Saito, A. Tsuboi, A. Nakada, R. Niikura, M. Fujishiro, S. Oka, S. Ishihara, T. Matsuda, S. Tanaka, K. Koike, and T. Tada, "Automatic detection of erosions and ulcerations in wireless capsule endoscopy images based on a deep convolutional neural network," *Gastrointestinal Endoscopy*, vol. 89, no. 2, pp. 357–363, Feb. 2019.

[10] T. D. Lange, P. Halvorsen, and M. Riegler, "Methodology to develop machine learning algorithms to improve performance in gastrointestinal endoscopy," *World J. Gastroenterology*, vol. 24, no. 45, pp. 5057–5062, Dec. 2018.

[11] M. Vania, B. A. Tama, H. Maulahela, and S. Lim, "Recent advances in applying machine learning and deep learning to detect upper gastrointestinal tract lesions," *IEEE Access*, vol. 11, pp. 66544–66567, 2023.

[12] F. Yasmin, M. M. Hassan, M. Hasan, S. Zaman, A. K. Bairagi, W. El-Shafai, H. Fouad, and Y. C. Chun, "GastroNet: Gastrointestinal polyp and abnormal feature detection and classification with deep learning approach," *IEEE Access*, vol. 11, pp. 97605–97624, 2023.

[13] M. Souaidi and M. E. Ansari, "Multi-scale analysis of ulcer disease detection from WCE images," *IET Image Process.*, vol. 13, no. 12, pp. 2233–2244, Oct. 2019.

[14] M. Souaidi, A. A. Abdelouahed, and M. El Ansari, "Multi-scale completed local binary patterns for ulcer detection in wireless capsule endoscopy images," *Multimedia Tools Appl.*, vol. 78, no. 10, pp. 13091–13108, May 2019.

[15] T. Cogan, M. Cogan, and L. Tamil, "MAPGI: Accurate identification of anatomical landmarks and diseased tissue in gastrointestinal tract using deep learning," *Comput. Biol. Med.*, vol. 111, Aug. 2019, Art. no. 103351.

[16] J. Naz, M. Attique Khan, M. Alhaisoni, O.-Y. Song, U. Tariq, and S. Kadry, "Segmentation and classification of stomach abnormalities using deep learning," *Comput., Mater. Continua*, vol. 69, no. 1, pp. 607–625, 2021.

[17] M. N. Noor, M. Nazir, I. Ashraf, N. A. Almujally, M. Aslam, and S. F. Jilani, "GastroNet: A robust attention-based deep learning and cosine similarity feature selection framework for gastrointestinal disease classification from endoscopic images," *CAAI Trans. Intell. Technol.*, Jun. 2023.

[18] Y. Liu, Z. Gu, and W. K. Cheung, "Hkbu at MediaEval 2017 medico: Medical multimedia task," in *Proc. MediaEval*, 2017, pp. 1–3.

[19] A. Asperti and C. Mastronardo, "The effectiveness of data augmentation for detection of gastrointestinal diseases from endoscopical images," 2017, *arXiv:1712.03689*.

[20] X. Zhang, W. Hu, F. Chen, J. Liu, Y. Yang, L. Wang, H. Duan, and J. Si, "Gastric precancerous diseases classification using CNN with a concise model," *PLoS ONE*, vol. 12, no. 9, Sep. 2017, Art. no. e0185508.

[21] T. Agrawal, R. Gupta, S. Sahu, and C. Y. Espy-Wilson, "SCL-UMD at the medico task-MediaEval 2017: Transfer learning based classification of medical images," in *Proc. MediaEval*, 2017, pp. 1–13.

[22] K. Pogorelov, K. R. Randel, C. Griwodz, S. L. Eskeland, T. de Lange, D. Johansen, C. Spampinato, D.-T. Dang-Nguyen, M. Lux, and P. T. Schmidt, "KVASIR: A multi-class image dataset for computer aided gastrointestinal disease detection," in *Proc. 8th ACM Multimedia Syst. Conf.*, 2017, pp. 164–169.

[23] Q. Zhang, "A novel ResNet101 model based on dense dilated convolution for image classification," *Social Netw. Appl. Sci.*, vol. 4, no. 1, pp. 1–13, Jan. 2022.

[24] S. S. A. Naqvi, S. Nadeem, M. Zaid, and M. A. Tahir, "Ensemble of texture features for finding abnormalities in the gastro-intestinal tract," in *Proc. MediaEval*, 2017, pp. 469–478.

[25] S. Kaur, S. Gupta, S. Singh, V. T. Hoang, S. Almakdi, T. Alelyani, and A. Shaikh, "Transfer learning-based automatic hurricane damage detection using satellite images," *Electronics*, vol. 11, no. 9, p. 1448, Apr. 2022.

[26] C. A. Ferreira, T. Melo, P. Sousa, M. I. Meyer, E. Shakibapour, P. Costa, and A. Campilho, "Classification of breast cancer histology images through transfer learning using a pre-trained inception Resnet v2," in *Proc. Int. Conf. Image Anal. Recognit.* Cham, Switzerland: Springer 2018, pp. 763–770.

[27] F. Baldassarre, D. González Morín, and L. Rodés-Guirao, "Deep koalarization: Image colorization using CNNs and Inception-ResNet-v2," 2017, *arXiv:1712.03400*.

[28] S. Sharma and S. Kumar, "The xception model: A potential feature extractor in breast cancer histology images classification," *ICT Exp.*, vol. 8, no. 1, pp. 101–108, Mar. 2022.

[29] A. Sulaiman, S. Kaur, S. Gupta, H. Alshahrani, M. S. A. Reshan, S. Alyami, and A. Shaikh, "ResRandSVM: Hybrid approach for acute lymphocytic leukemia classification in blood smear images," *Diagnostics*, vol. 13, no. 12, p. 2121, Jun. 2023.

[30] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4510–4520.

[31] I. Z. Mukti and D. Biswas, "Transfer learning based plant diseases detection using ResNet50," in *Proc. 4th Int. Conf. Electr. Inf. Commun. Technol. (EICT)*, Dec. 2019, pp. 1–6.

[32] S. Kaur, S. Gupta, S. Singh, D. Koundal, and A. Zaguia, "Convolutional neural network based hurricane damage detection using satellite images," *Soft Comput.*, vol. 26, no. 16, pp. 7831–7845, Aug. 2022.

[33] I. Syarif, A. Prugel-Bennett, and G. Wills, "SVM parameter optimization using grid search and genetic algorithm to improve classification performance," *TELKOMNIKA (Telecommunication Comput. Electron. Control)*, vol. 14, no. 4, p. 1502, Dec. 2016.

[34] P. Probst, A.-L. Boulesteix, and B. Bischl, "Tunability: Importance of hyperparameters of machine learning algorithms," *J. Mach. Learn. Res.*, vol. 20, no. 1, pp. 1934–1965, 2019.

[35] O. K. Oyedotun, K. Papadopoulos, and D. Aouada, "A new perspective for understanding generalization gap of deep neural networks trained with large batch sizes," *Appl. Intell.*, vol. 53, no. 12, pp. 15621–15637, Jun. 2023.

[36] A. Alshammari and R. C. Chabaan, "Sppn-rn101: Spatial pyramid pooling network with Resnet101-based foreign object debris detection in airports," *Mathematics*, vol. 11, no. 4, p. 841, Feb. 2023.

[37] Z. Zhou, X. Yang, J. Ji, Y. Wang, and Z. Zhu, "Classifying fabric defects with evolving inception v3 by improved L2, 1-norm regularized extreme learning machine," *Textile Res. J.*, vol. 93, nos. 3–4, pp. 936–956, 2023.

[38] B. Das, A. Saha, and S. Mukhopadhyay, "Rain removal from a single image using refined inception ResNet v2," *Circuits, Syst., Signal Process.*, vol. 42, no. 6, pp. 3485–3508, Jun. 2023.

[39] R. O. Ogundokun, A. Li, R. S. Babatunde, C. Umezuruike, P. O. Sadiku, A. T. Abdulahi, and A. N. Babatunde, "Enhancing skin cancer detection and classification in dermoscopic images through concatenated MobileNetV2 and xception models," *Bioengineering*, vol. 10, no. 8, p. 979, Aug. 2023.

[40] S. Pappula, T. Nadendla, N. B. Lomadugu, and S. R. Nalla, "Detection and classification of pneumonia using deep learning by the dense net-121 model," in *Proc. 9th Int. Conf. Adv. Comput. Commun. Syst. (ICACCS)*, vol. 1, Mar. 2023, pp. 1671–1675.

[41] T. F. Mahdi, H. G. Daway, and J. Jouda, "White blood cell detection and classification using transfer DenseNet201 and MobileNetv2 learning models," in *Proc. AIP Conf.*, 2023, vol. 2830, no. 1, p. 40013, doi: 10.1063/5.0156771.

[42] C.-L. Lin and K.-C. Wu, "Development of revised ResNet-50 for diabetic retinopathy detection," *BMC Bioinf.*, vol. 24, no. 1, pp. 1–18, Apr. 2023.

[43] U. K. Lilhore, M. Poongodi, A. Kaur, S. Simaiya, A. D. Algarni, H. Elmannai, V. Vijayakumar, G. B. Tunze, and M. Hamdi, "Hybrid model for detection of cervical cancer using causal analysis and machine learning techniques," *Comput. Math. Methods Med.*, vol. 2022, pp. 1–17, May 2022.

[44] P. Dhiman, V. Kukreja, P. Manoharan, A. Kaur, M. M. Kamruzzaman, I. B. Dhaou, and C. Iwendi, "A novel deep learning model for detection of severity level of the disease in citrus fruits," *Electronics*, vol. 11, no. 3, p. 495, Feb. 2022.

[45] A. Ahmed, "Classification of gastrointestinal images based on transfer learning and denoising convolutional neural networks," in *Proc. Int. Conf. Data Sci. Appl. (ICDSA)*, vol. 1, Singapore: Springer, 2021, pp. 631–639.

[46] D. Ezzat, H. M. Afify, M. H. N. Taha, and A. E. Hassanien, "Convolutional neural network with batch normalization for classification of endoscopic gastrointestinal diseases," in *Machine Learning and Big Data Analytics Paradigms: Analysis, Applications and Challenges*. Wiesbaden, Germany: Springer, 2021, pp. 113–128.

[47] R. Liao, K. Qi, D. Che, and T. H. Zeng, "Exploration of the possibility of early diagnosis for digestive diseases using deep learning techniques," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Dec. 2020, pp. 2343–2350.

[48] A. M. Godkhindi and R. M. Gowda, "Automated detection of polyps in CT colonography images using deep learning algorithms in colon cancer diagnosis," in *Proc. Int. Conf. Energy, Commun., Data Analytics Soft Comput. (ICECDS)*, Aug. 2017, pp. 1722–1728.

[49] A. A. Pozdeev, N. A. Obukhova, and A. A. Motyko, "Automatic analysis of endoscopic images for polyps detection and segmentation," in *Proc. IEEE Conf. Russian Young Researchers Electr. Electron. Eng. (EIConRus)*, Jan. 2019, pp. 1216–1220.

[50] M. Min, S. Su, W. He, Y. Bi, Z. Ma, and Y. Liu, "Computer-aided diagnosis of colorectal polyps using linked color imaging colonoscopy to predict histology," *Sci. Rep.*, vol. 9, no. 1, pp. 2881–2888, Feb. 2019.

[51] A. Bour, C. Castillo-Olea, B. Garcia-Zapirain, and S. Zahia, "Automatic colon polyp classification using convolutional neural network: A case study at Basque country," in *Proc. IEEE Int. Symp. Signal Process. Inf. Technol. (ISSPIT)*, Dec. 2019, pp. 1–5.

[52] R. Zhang, Y. Zheng, T. W. C. Mak, R. Yu, S. H. Wong, J. Y. W. Lau, and C. C. Y. Poon, "Automatic detection and classification of colorectal polyps by transferring low-level CNN features from nonmedical domain," *IEEE J. Biomed. Health Informat.*, vol. 21, no. 1, pp. 41–47, Jan. 2017.

[53] R. Zhu, R. Zhang, and D. Xue, "Lesion detection of endoscopy images based on convolutional neural network features," in *Proc. 8th Int. Congr. Image Signal Process. (CISP)*, Oct. 2015, pp. 372–376.

MOHAMED A. ELMAGZOUB was born in Riyadh, Saudi Arabia, in 1985. He received the B.S. degree (Hons.) in electrical and electronic engineering from Omdurman Islamic University, Sudan, in 2007, and the master's degree in electrical communication engineering and the Ph.D. degree in electrical engineering (communications) from Universiti Teknologi Malaysia (UTM), Malaysia, in 2012 and 2016, respectively. He was a Telecommunication Engineer with the Communication Department, Dams Implementation Unit, Sudan, from 2008 to 2009. He is currently an Associate Professor with the Department of Network and Communication Engineering, Najran University. He has authored ten ISI articles, two of them highly cited review articles in the area of passive optical networks and hybrid optical and wireless communication systems. His research interests include passive optical networks, hybrid optical and wireless communication systems, optical OFDM communications, and MIMO and next generation access networks.

SWAPANDEEP KAUR received the B.E., M.Tech., Ph.D. degrees in electronics and communication engineering in 2012, 2016, and 2022, respectively. She is an Assistant Professor with Chitkara University, for the past seven years. She has published 13 research papers and filed seven patents. Her research interests include artificial intelligence, machine learning, and deep learning.

SHEIFALI GUPTA is a Professor with the Chitkara University Research and Innovation Network (CURIN), Chitkara University, Punjab Campus, India. She specializes in the area of digital image processing, pattern recognition, machine intelligence, biomedical image processing, agriculture based image processing, and deep learning. She has published more than 60 research papers and articles in reputed national and international journals and conferences. She has filed 15 patents in the field of image processing and mechatronics. She has conducted different workshops based on modelling real-time video processing system in simulink using the video and image processing blockset. Currently, she is working on an agriculture based project in collaboration with Dr. Y. S. Parmar University of Horticulture and Forestry, Nauni, Solan. She is guiding post graduate and doctoral students.

She has received the prestigious IRDP Award 2018 for remarkable achievements in teaching, research and publications.

ADEL RAJAB received the bachelor's degree in computer science and information system and the master's and Ph.D. degrees in computer science and engineering from the University of South Carolina, USA. He is currently an Associate Professor with the College of Computer Science and Information System (CSIS) and the Vice-Dean of graduate studies for academic affairs with Najran University, Najran, Saudi Arabia. His research interests include robotics, drones, machine learning, and bioinformatic OBS networks.

KHAIRAN D. RAJAB received the bachelor's degree from the University of South Carolina, USA, the master's degree from the University of South Florida, USA, and the Ph.D. degree in computer science and engineering from the University of Southern Florida, under the supervision of Prof. L. A. Piegl. He is currently a Full Professor with the College of Computer Science and Information Systems (CSIS), Najran University, Najran, Saudi Arabia. He has published over 54 research papers in high impact factor journals and reputable conferences. His research interests include geometric modeling, NURBS, data mining, network security, and cyber learning. He is a member of the College Council, CSIS. He is a reviewer and on editorial board of several research conferences and journals.

MANA SALEH AL RESHAN received the B.S. degree in information systems from King Khalid University, Abha, Saudi Arabia, in 2007, the M.S. degree (Hons.) in computer, information, and network security from DePaul University, Chicago, USA, in 2011, and the Ph.D. degree in computer science from The Catholic University of America (CUA), Washington, DC, USA, in 2019. He was a Teaching Assistant with the College of Computer Science and Information Systems, Najran University, Saudi Arabia, from 2007 to 2009. He was a Lecturer with the College of Computer Science and Information Systems, Najran University, in 2012, where he is currently an Associate Professor and the Head of the Network Engineering Department. His current research interests include computer network and security, system security, wireless and mobile security, body area networks, and cloud security.

HANI ALSHAHRANI (Senior Member, IEEE) received the bachelor's degree in computer science from King Khalid University, Abha, Saudi Arabia, the master's degree in computer science from California Lutheran University, Thousand Oaks, CA, USA, and the Ph.D. degree from Oakland University, Rochester, MI, USA. Currently, he is an Associate Professor of computer science and information systems with Najran University, Najran, Saudi Arabia. He has published over 60 research publications. His current research interests include smartphones, the IoT, crowdsourcing security, and privacy.

ASADULLAH SHAIKH (Senior Member, IEEE) received the B.Sc. degree in software development from the University of Huddersfield, England, the M.Sc. degree in software engineering and management from Goteborg University, Sweden, and the Ph.D. degree in software engineering from the University of Southern Denmark. He is currently a Professor, the Head of research and graduate studies, and the Coordinator of seminars and training with the College of Computer Science and Information Systems, Najran University, Najran, Saudi Arabia. He was a Researcher with UOC, Barcelona, Spain. He has vast experience in teaching and research. He has more than 200 publications in the area of software engineering in international journals and conferences. His current research interests include UML model verification, UML class diagrams verification with OCL constraints for complex models, formal verification, and feedback technique for unsatisfiable UML/OCL class diagrams. He is an Editor of the *International Journal of Advanced Computer Systems and Software Engineering* (IJACSSE) and an International Advisory Board of several conferences and journals. Further details can be obtained using www.asadshaikh.com.