RESEARCH ARTICLE

Classification of Lung and Colon Cancer Histopathological Images Using Global Context Attention Based Convolutional Neural Network

MD. AL-MAMUN PROVATH[1], **KAUSHIK DEB**[1], **PRANAB KUMAR DHAR**[1,2],
AND TETSUYA SHIMAMURA[3], **(Senior Member, IEEE)**

[1]Department of Computer Science and Engineering, Chittagong University of Engineering and Technology (CUET), Chattogram 4349, Bangladesh
[2]Faculty of Science and Engineering, Waseda University, Shinjuku-ku, Tokyo 169-8555, Japan
[3]Department of Information and Computer Sciences, Saitama University, Sakura-ku, Saitama 338-8570, Japan

Corresponding author: Kaushik Deb (debkaushik99@cuet.ac.bd)

ABSTRACT The malignant neoplastic malady known as cancer appears to exhibit a significantly elevated rate of mortality owing to its virulence and pronounced propensity for metastasis. To augment the diagnostic efficacy, research endeavors have been undertaken utilizing complex deep learning architectures. However, the performance of these efforts remains circumscribed by smaller dataset size, quality of the data, the interclass variations present between lung adenocarcinoma and lung squamous cell carcinoma, and the complexity of deploying to mobile devices and failure to address both image and patient level accuracy measurements. To surmount these obstacles, the present study proposes a stage-based method for enhancing the images, in conjunction with utilizing a global context attention-guided convolutional neural network that effectively captures both channel and spatial information and semantic information extracted from the input image. Implementing the proposed methodology increased total image level accuracy to 99.76% and a patient level accuracy of 96.5%, a metric that has yet to be previously quantified. The addition of the global context attention module decreases the model's parameter count by 0.47 million, reduces the computational costs by saving 10.54 million floating point operations per second (FLOPs) and 10.72 million multiply-accumulate operations (MACs), and results in a 0.03s improvement in inference time. Furthermore, this module enhances both image level and patient level accuracy, boosting them by 2.84% and 3.17%, respectively, compared to using only the convolutional block attention module in the baseline convolutional neural network. Consequently, this modification renders the model highly suitable for deployment on mobile devices due to its adaptability. Our findings provide supporting evidence for the potential of this method to serve as a noninvasive screening tool capable of reliably classifying lung and colon cancer subtypes.

INDEX TERMS Stage-based method, global context attention module, convolutional neural networks, noninvasive screening tool, lung and colon cancer.

I. INTRODUCTION

Cancer encompasses a heterogeneous array of maladies, defined by the aberrant proliferation of malignant cells that can infiltrate and disseminate to other parts of the organism. The International Agency for Research on Cancer (IARC) reports that cancer remains a leading cause of death

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

worldwide, with a staggering 19 million new cases and 10 million deaths recorded in the year 2020 [1]. The phenomenon of metastasis, whereby cancer cells transmigrate from the primary site to an additional organ. In the year 2020, cancer in the lung and colorectal cancer emerged as the most widespread forms of cancer that caused death among both male and female individuals globally, with the appearance of 2.21 million fresh incidences of lung cancer and 1.93 million instances of colorectal cancer reported around the world,

ultimately leading to 1.80 million fatalities due to lung cancer and roughly 1 million deaths resulting from colorectal cancer [2]. Cancer formation can be attributed to various causes, including behavioral traits like a high body mass index, a habit of consuming alcohol excessively, and smoking. Furthermore, physical toxins such as radiation, UV rays, and genetic factors can also contribute to the formation of cancer [2]. Although weight loss, fatigue, nausea, persistent coughing, shortness of breath, muscle pain, bleeding, and bruising are symptoms commonly associated with cancer [3], they are not universal among all patients. They are not specific to cancer, making early detection difficult without a comprehensive set of diagnostic procedures. With 90% of cancer deaths resulting from metastatic disease, understanding and studying metastasis is a critical aspect of cancer research [4]. Metastasis of colon and lung cancers commonly occurs in the liver, lungs, peritoneum, brain, liver, bones, and lungs for lung cancer, respectively [5]. At the microscopic level, metastatic cells appear to be diseased primary cells, although symptoms are often associated with cancer cells in the affected organ [5]. The most effective method for reducing cancer mortality rate is through early detection and proper treatment [6]. For example, if colon cancer is detected at Stage 0, 92% of patients between 18 and 73 can survive with appropriate therapy [7]. The survival rates for lung cancer from Stage 1 to Stage 4 are 69%, 50%, 29%, and 8%, respectively. The unavailability of affordable screening systems is a hindrance to early detection of cancer, particularly in nations where approximately 70% of cancer-related deaths occur, according to the World Health Organization [2].

Deep learning (DL) offers a plausible remedy as a possible answer to this predicament. Pathology has utilized machine learning (ML) in various applications, such as disease diagnosis [8]. Automating the classification of histopathological images using ML algorithms can accurately diagnose disorders while reducing the workload of human experts and meeting the requirements of high accuracy, large datasets, and other criteria [9]. Additionally, transfer learning can be employed in the implementation of a different method to tackle cross-disciplinary knowledge acquisition challenges by leveraging information from existing datasets [9]. The application of scratch convolutional neural networks (CNN) presents certain advantages in comparison to transfer learning techniques.

Diagnostic techniques employed for the detection of lung cancer and colon cancer may include computed tomography (CT), X-rays, and biopsies, among other approaches. Biopsies are commonly utilized in the diagnosis of lung and colon cancer due to their high accuracy in identifying cancerous cells or tissues. Pathologists examine histology slides under a microscope to make their diagnoses. However, traditional manual diagnosis poses a significant burden on qualified experts, and pathologists who lack diagnostic experience are prone to errors in diagnosis. For those reasons, the classification of pathological images of lung and colon cancer necessitates a more intricate representation of features.

The histopathology image classification for lung and colon cancer faces several persistent challenges that have yet to be resolved. Initially, one of the major obstacles is the potential for misclassification due to the overlapping histopathological characteristics of different lung cancer subtypes (lung adenocarcinoma, lung squamous cell carcinoma), which can make it challenging even for experienced pathologists to differentiate between them. Accordingly, CNN, which can accurately distinguish between such subtypes, is critical for reliable diagnosis. Secondly, the scarcity of accessible datasets presents a formidable challenge in developing CNN models for histopathological image analysis, as the model may not accurately learn the differentiating features of various lung and colon cancer images. Moreover, focusing solely on lung cancer excludes benign nodules. Excluding the benign nodules can also lead to a bias in the model, as the model is not exposed to benign nodules, which may differ in shape, size, and texture from malignant nodules. The exclusion of benign can lead to overfitting the model by making it less accurate. Furthermore, focusing on the shape-based diversity of colon cancer images is the lack of standardization in defining and measuring shape-based features. Thirdly, the absence of sharpness within the image impairs its ability to discern intricate details, primarily attributed to the inherent blurriness present in the images. Fourthly, other authors have emphasized the holistic examination of the entire image, neglecting to allocate due consideration to localized regions. This oversight engenders heightened computational intricacies within the analysis framework. Fifthly, network interpretability falls short in elucidating the specific regions of interest that the model prioritizes, thus obscuring the precise focal areas within the analyzed images. Finally, incorporating patient-level value into the analysis can provide a more comprehensive understanding of the relationship between clinical factors and imaging features and ultimately lead to more accurate classification methods.

To overcome these issues, we develop a CNN model from scratch, with the ultimate aim of achieving higher accuracy in histopathological image classification. Specifically, we have incorporated the convolutional block attention module (CBAM) [10] into the network, which enables it to selectively focus on crucial features while disregarding irrelevant ones, thus improving the adaptive refinement of tissue feature extraction from each layer of the network. The integration of the global context (GC) module with the CBAM can enhance the CBAM's ability to capture semantic information from input images. The concatenated module can capture fine-grained details, attend to relevant image regions, and consider global contextual information. In addition, we have proposed a generative adversarial network (GAN) architecture and also labeled the CRAG dataset by expert histopathologists to generate supplementary images that can

be used in conjunction with the original dataset to mitigate the issue of image scarcity. We further extended our scope to encompass benign nodules, employing advanced image enhancement techniques to enhance the visual clarity of our dataset. Additionally, we integrated GradCAM visualization for its capacity to illuminate and accentuate areas indicative of malignancy within the images and significantly enhanced the precision and interpretative depth of our method.

This work describes a scratch CNN model developed by incorporating the global context attention block. Our contributions are as follows:

1) The proposed scratch CNN model is combined with the global context attention block, which enables the model to focus on the most relevant features for classification that help disambiguate between similar-looking classes. The proposed module reduces 0.47 million parameters and 10.54 million FLOPs,10.72 million MACs, and 0.03s inference time, and also increases the image level and patient level accuracy by 2.84% and 3.17%, respectively, than integrating CBAM [10] alone to CNN baseline.

2) We created a dataset, GGLCI, using the proposed GAN architecture and also labeled the CRAG dataset by expert histopathologists to increase samples. These additional images were able to address similar cancer subtypes and benign nodules incorporated in lung cancer to remove bias and improve the model's accuracy [11].

3) We developed a stage-based image enhancement method to enhance the quality of histopathological images.

4) To the best of our knowledge, this is the first work that addresses both patient and image levels in classifying lung and colon cancer. We also evaluate the proposed model on a more challenging Kather Multiclass Dataset that allows us to validate the robustness of our approach on a different dataset.

The remaining parts of this study are organized as follows: Section II will provide a brief yet comprehensive review of the current state-of-the-art techniques. The finer points of our proposed methodology shall be explicated in detail in Section III. The results of our empirical investigations shall be presented in Section IV, and the conclusion will be provided in Section V.

II. RELATED WORK

For over four decades, researchers in the field of medical imaging have been actively investigating the viability of automating the diagnosis of cancer through the classification of histopathological images into either benign or malignant patterns, with the primary objective of facilitating a more efficient and reliable image analysis system. Despite these efforts, the intricate nature of histological images has presented significant challenges in the image analysis

process, impeding the full realization of a comprehensive and sophisticated diagnostic tool. The study [8] delves into the evolution, present-day implementation, and future prospects of the application of computer-aided diagnosis (CAD) in medical imaging. The paper [9] introduces a sophisticated technique for the automated identification of malignancies within mammograms. The efficacy of the method was evaluated on a dataset of 200 mammograms, resulting in a classification accuracy of 85%. The paper [10] proposes a sophisticated method for the categorization of lung needle biopsy images utilizing sparse representation and a combination of multiple modalities, including texture, shape, and intensity, in order to enhance the accuracy of classification.

In [12], the method was put to the test using a dataset consisting of 38 tissue samples. It achieved a classification accuracy of 95% for normal and cancerous tissue, as well as an accuracy of 80% for grading the cancer tissue. The paper [13] presented an algorithmic approach to identifying polyps within colonoscopy footage. The efficacy of the method was gauged using a dataset of 400 colonoscopy videos, resulting in 92% sensitivity. The authors of the paper [14] present a computational method for identifying and categorizing instances of pulmonary cancer within CT scans. The efficacy of the method was tested on a set of 100 CT scans, yielding a 95% accuracy rate for cancer detection and 85% for stage classification. The study in [15], which tested a dataset of 1115 patients, exhibited a remarkable accuracy of 97.64%. The method proposed in the study [16] was experimentally validated on a dataset, yielding a diagnostic accuracy of 95.44% for lung cancer and 94.44% for colon cancer. The findings in [17] showed a high level of accuracy, with a 97.20% success rate in identifying lung cancer. A recent investigation [11] introduced a novel and tailored deep convolutional neural network (DNN) architecture derived from the VGG16 model, specifically devised to enable the noninvasive detection and differentiation of lung cancer and latent tuberculosis (LTB). Through extensive training and testing on sizable CT image datasets, the DNN exhibited an impressive accuracy rate of 90.4%. However, the study acknowledged certain shortcomings, notably the limited number of representative examples in relation to the vast number of adjustable parameters present within the DNN architecture. Additionally, the exclusion of benign nodules from the dataset gave rise to inquiries regarding their potential influence on classification accuracy as well as the DNN's inherent capability to segment such nodules autonomously. Despite the DNN's noteworthy performance, occasional misclassification of nodules occurred due to overlapping features between LTB and malignant nodules, underscoring the innate complexities associated with this classification predicament.

The study [18] endeavored to devise an artificial intelligence-based diagnostic tool for the autonomous discernment of normal, unclear, and tumor images from colonoscopy videos. The methodology encompassed the

acquisition of 47,555 jpg images from colonoscopy videos of 24 patients, which were subsequently partitioned into three distinct classes by an accomplished gastroenterologist. Subsequently, a singular shot detector, a deep learning framework specifically tailored for object detection, underwent training utilizing 47,255 images and underwent validation employing two sets of 300 images. The obtained outcomes exhibited an average accuracy of 0.9067, precision of 0.9744, recall of 0.9067, and F1 score of 0.9393. Nonetheless, the investigation identified constraints in identifying certain tumors within the validation set, thereby proposing the necessity for expanding the training set and greater diversification of colonoscopy image classes, encompassing tumor morphology, color characteristics, and severity, to heighten the model's efficacy.

Furthermore, the integration of [19] has facilitated the generation of artificial images and stimulated extensive explorations aimed at refining the foundational elements. The integration of CBAM [20], which employs a bifocal attention strategy to optimize the features garnered by CNN, represents a remarkable effort to elevate its efficacy. Therefore, it can be deduced that the categorization of lung and colonic neoplasms has been of paramount importance for a considerable interim and the fusion of profound neural networks.

In paper [35], the proposed CVCnet uses a cross-view block to capture diverse feature embeddings from the views in stereo vision, and a cascaded spatial perception module is proposed to redistribute each location in feature maps according to the weight it occupies, which makes the extraction of features more effective and achieves the best performance for stereo image super-resolution tasks. Moreover, in [36], a lightweight image super-resolution network was developed that uses an expectation maximization attention mechanism to capture the long-range dependencies of HR-size feature maps. The EMASRN achieves state-of-the-art performance on several image super-resolution benchmarks while being significantly lighter than the existing methods. Besides that, in [37], a novel SISR method was used, an interactive memory to store the intermediate results of the SISR process, and a semantic similarity learning module to implicitly capture the semantic similarity between different views. The IML SISR method achieves better performance on several SISR benchmarks by effectively exploiting the complementary information of different views and refining the features of the intermediate results.

Therefore, it can be deduced that the categorization of lung and colonic neoplasms has been a matter of paramount import for a considerable interim. The fusion of profundal neural nettles and a plenitude of configurations has surpassed the current apogee of art methods. This fledgling discipline of inquiry presents a wealth of potential for growth and progress to confront these difficulties.

TABLE 1. Enhancement result of our method.

Metrics	Value
Entropy	7.78
PSNR	12.98
SI	31.74
IQI	85.78

A. PREPROCESSING

Preprocessing is crucial for enhancing the accuracy of cancer histopathological image classification techniques. The preprocessing phase involves several components, including image enhancement, resizing and scaling, and normalization.

Stage Based Image Enhancement Method: The selection of a distribution for the contrast transform function in enhancement methods depends on the type of input image. For histopathology images, a uniform distribution is often used to create a flat histogram. On the other hand, a bell-shaped histogram can be achieved using a rayleigh distribution in Fig. 1. In the first stage, a flat-shaped histogram is used to enhance the overall brightness and contrast of the image, and the formula for this is in Equation 1:

$$g(r) = 255 \times \frac{f(r) - min(f)}{max(f) - min(f)} \tag{1}$$

where, $g(r)$ is the equalized intensity value, $f(r)$ is the original intensity value, $min(f)$ and $max(f)$ are the minimum and maximum intensities in the image.

In the second stage, a bell-shaped histogram is used to further refine the enhancement. This can be achieved through Gaussian filtering, which results in a bell-shaped histogram, and the formula for Gaussian filtering is in Equation 2:

$$g(x, y) = \frac{1}{2}\pi\sigma^2 e^{-\frac{(x^2+y^2)}{2\sigma^2}} \tag{2}$$

where, $g(x, y)$ is the filtered intensity value, x and y are the coordinates of the pixel, and σ is the standard deviation of the gaussian distribution.

The uniform distribution indicates a repetition of similar intensities, while the Rayleigh distribution usually represents a normal distribution with a cluster of high-density data. In the proposed approach, there are two stages involved in achieving the desired outcomes (Fig. 2).

The entropy of the input image is 6.80, whereas ours is 7.78. Table 1 enumerates the results of various metrics for evaluating the efficacy of histopathological slides. It can be deduced from Table 1 that the stage-based image enhancement method yields commendable results in terms of the quality assessment metrics of the images.

In order to minimize computational expenses and optimize model performance, the images are resized to a compact size of 64×64 pixels. This resizing operation significantly reduces the computational burden and processing time, resulting in improved model performance. Furthermore, the original
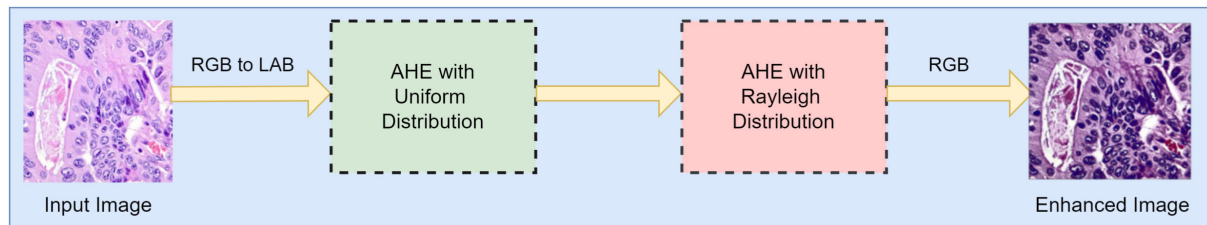
FIGURE 1. Stage-based image enhancement method.
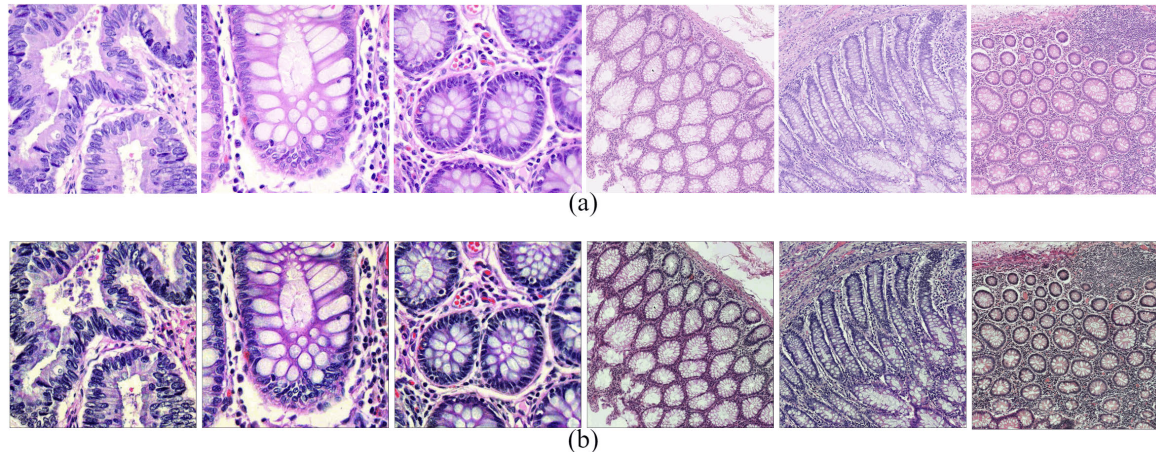
FIGURE 2. Image enhancement examples. (a) : Input image. (b): Enhanced image.

images contain RGB coefficients ranging from 0 to 255, which are excessively high for efficient processing. As a solution, we have transformed the pixel values by scaling them with a factor of 1/255.0, which effectively maps the pixel values to the normalized range of [0, 1]. This step ensures that all pixel values are brought within a standardized interval, facilitating faster convergence during network training.

The addition of normalization aims to standardize the raw input pixels, resulting in their transformation to have a mean of 0 and a standard deviation of 1. Without normalization, the dataset may contain numerical data points that vary significantly in magnitude, with some being very high and others very low. Consequently, during training, unnormalized data can lead to the dominance of large values over smaller ones for unevenly distributed input importance. However, normalizing the training data brings all data points to a uniform scale, enabling faster convergence during network training.

B. FEATURE EXTRACTION

CNNs are powerful tools for feature extraction in lung and colon cancer classification. They enable accurate and automated analysis of medical images, utilizing their hierarchical architecture, adaptive filtering, and ability to capture complex patterns. CNNs have significantly transformed cancer classification, specifically lung and colon cancer, by serving as advanced tools for extracting important features. These models automatically learn and extract meaningful characteristics from raw histopathological images, capturing intricate patterns and structures that indicate different types of cancer. CNNs use convolutional layers that apply filters to the input image, analyzing it to produce maps of features that represent local patterns. These filters act as detectors for edges, textures, and other visual elements relevant to cancer classification. Pooling layers are then used to condense the feature maps while retaining the most significant information. The hierarchical arrangement of convolutional and pooling layers allows the CNN to progressively capture and understand complex features present in the input images. Moreover, the learned features in CNNs go beyond simple visual attributes and can represent more abstract concepts. As the network learns from large sets of labeled data, it automatically discovers distinctive features that are crucial for distinguishing between different types of cancer. This innate ability of CNNs to identify informative features is essential for developing highly accurate cancer classification models.

C. CLASSIFICATION USING SOFTMAX

The utilization of the softmax activation function in the intricate histopathological image classification task for five classes of lung and colon cancer enables the conversion of raw model outputs into a probability distribution, thereby facilitating precise inference and decision making and

expressed by the provided Equation 3:

$$S(i)_k = \frac{e^{i_k}}{\sum_{t=1}^{5} e^{i_t}} \tag{3}$$

where i and k denote the input vector and index of the current element, respectively.

D. GLOBAL CONTEXT ATTENTION BLOCK INCORPORATION WITH SCRATCH CNN MODEL

The methodology employed in Fig. 3 for the CNN initiates with an input image of size $64 \times 64 \times 3$, which undergoes two sequential convolutional blocks, kernel of size 3, and 32 output channels in each, designed to extract features through the use of filters and *GeLU* activation function, with the filters traversing across the input image to generate feature maps that emphasize specific traits of the image and the activation functions introducing non-linearity to the model, thus enabling it to comprehend intricate relationships within the data. The max pooling layer is of pool size 2×2 with stride 2, and the stride dimension is (1,1) with the same padding. To promote faster convergence, the *GeLU* activation function is employed. It can be defined as the product of the input value and the probability that a random variable sampled from the standard normal distribution is less than or equal to the input value. This probability is calculated using the cumulative distribution function of the standard normal distribution, denoted as P, and Equation 4 can be expressed as follows:

$$GeLU(x) = x \cdot P(X \leq x) \tag{4}$$

Here, P is the cumulative distribution function of the standard normal distribution. Subsequently, the model integrates three residual units meticulously devised to facilitate the training of deeper networks by circumventing the vanishing gradient problem. The model then employs a separable convolution layer that performs spatial convolution on the input followed by pointwise convolution on the output with the kernel of size 3 and various sizes of output channels. The model includes a GC attention block featuring a self-attention mechanism to enable the model to focus on varying parts of the image during the forward pass. Additionally, the activations of the preceding layer are normalized for each mini-batch of data using batch normalization to maintain the stability of the model during training and prevent activations from exploding or vanishing. During training, the dropout layer randomly drops out activations, thus preventing overfitting. Finally, after passing through the fully connected layer, the Softmax activation layer provides a probability distribution over the final prediction. To enhance a model's overall performance, combining batch normalization and dropout techniques is advantageous by incorporating an IC layer [21]. The IC layer serves the purpose of decreasing the mutual information between the outputs of any two neurons within the model. This reduction in mutual information is achieved by a factor of p^2, where p represents the probability

of dropout. Additionally, the IC layer also reduces the correlation coefficient between neurons by a factor of p. Thus IC layer effectively reduces the amount of shared information between neurons and decreases the interdependence among them, resulting in improved model performance.

This work presents an attempt to implement the residual unit by integrating {-IC-seperable_conv-GC attention block-*GeLU*-} layers. The study focuses on three distinct types of residual units and analyzes their unique short paths, with the objective of determining the optimal residual unit types: 1. $3\times$\{*GeLU*-IC-seperable_conv-GC attention block\} 2. $3\times$\{IC-seperable_conv-GC attention block-*GeLU*\} 3. $3\times$\{seperable_conv-GC attention block-*GeLU*-IC\}.

In order to enhance the effectiveness of CBAM's attention mechanism by incorporating semantic information, we fused the GC module, creating the GC attention block. The primary purpose of the GC module is to capture the holistic understanding and interpretation of the input image. Semantic information holds significant importance in the task of image classification as it empowers the model to grasp the underlying meaning and contextual relevance of objects depicted within the image. This goes beyond relying solely on basic visual features like edges and textures. Integrating the gc module into CBAM's attention mechanism enables the model to incorporate high-level semantic cues and global context while attending to specific regions of interest. This combination enhances the model's ability to comprehend the overall semantics of the image, leading to improved image classification performance.

Our proposed GC attention block is composed of two primary components: the attention module(CBAM) and GC module. CBAM's channel attention mechanism computes the channel-wise importance of features in an input feature map by rescaling them using a learned weight vector W and the spatial attention mechanism computes the spatial-wise importance of the features. It generates a spatial attention map A. Given input feature maps $X \in R^{H \times W \times C}$, where H represents the height, W represents the width, and C represents the number of channels in Equations 5-8:

1) GLOBAL CONTEXT MODULE

$$F_{gc} = f_{pool}(X) \in R^{1 \times 1 \times C} \tag{5}$$

where f_{pool} denotes the global pooling operation applied to X, resulting in feature maps F_{gc} with spatial dimensions reduced to 1×1 while preserving the number of channels, the dimensions of the feature maps are represented using the real number space R.

2) CBAM MODULE

$$F_{cbam} = f_{cbam}(X) \in R^{H \times W \times C} \tag{6}$$

where, the function f_{cbam} represents the application of channel-wise and spatial attention mechanisms to X, resulting in feature maps F_{cbam} with the same spatial dimensions (H, W) and number of channels (C).
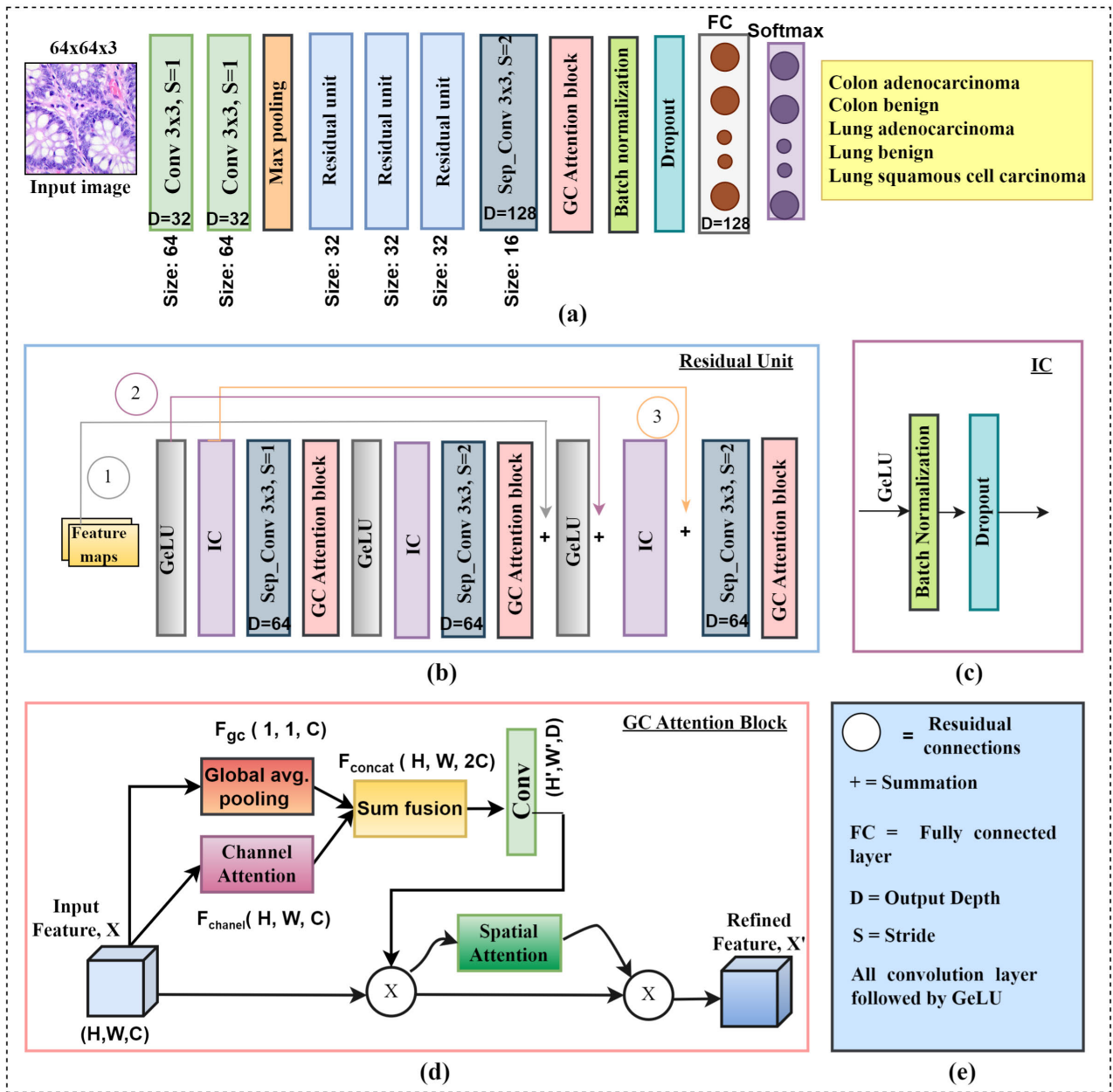
FIGURE 3. The overall framework of the proposed method. (a) : CNN baseline architecture. (b): Reformulating residual unit. (c): IC layer. (d): GC attention block. (e): Key showing important components.

3) CONCATENATION

$$F_{concat} = [F_{gc}, F_{cbam}] \in R^{H \times W \times 2C} \quad (7)$$

The operator $[\cdot, \cdot]$ denotes the concatenation along the channel axis, combining feature maps F_{gc} and F_{cbam} to form F_{concat} with dimensions $(H, W, 2C)$. The first C channels of F_{concat} correspond to F_{gc}, and the subsequent C channels correspond to F_{cbam}.

4) CONVOLUTIONAL LAYER

$$F_{fused} = f_{conv}(F_{concat}) \in R^{H' \times W' \times D} \quad (8)$$

The function f_{conv} represents the application of a convolutional layer to F_{concat}, resulting in fused feature maps F_{used} with dimensions (H', W', D). The effect of GC attention block on performance, parameter, MACs, inference time, and FLOPs count is explained in this paper.

III. EXPERIMENTS
A. IMPLEMENTATION ENVIRONMENT

The experiments for this work were conducted using Python v3.7 on a Colab Pro notebook, which is a cloud-based platform that provides faster access to backend tensor processing unit (TPU) runtime and offers up to 100 GB of

storage capacity for creating and saving notebooks. Google Colab Pro's CPU clock rate is 2.30 gigahertz, signifying a rapid processing capability of 2.30 billion cycles per second. The hardware utilized in the experiments included an Asus TUF Gaming A15 FA506ICB Ryzen 5 4600H RTX 3050 4GB Graphics with 8GB RAM, running on the Microsoft Windows 11 operating system. The deep learning model was developed using the Keras API with TensorFlow backend using FastAI framework, while the scikit learn library, Pandas, Numpy, cv2, Matplotlib, os were utilized for model evaluation, scoring strategy, and numerical computation.

B. DATASET DESCRIPTION

In order to establish the validity of our proposed model, we conducted experiments on multiple datasets, including the LC25000 dataset [22], labeled colorectal adenocarcinoma gland (CRAG) dataset [23], the self-generated GGLCI dataset of lung and colon cancer images created using GANs and validated by an expert histopathologist, the GLaS dataset [24], and the Kather Multiclass dataset [25].

1) LC25000 DATASET

In this current research, the LC25000 dataset is utilized in [22]. The dataset consists of 25,000 images of lung and colon tissues, which are classified into five categories. The lung tissue images are divided into three categories: adenocarcinoma, squamous cell carcinoma, and benign, while the colon images are classified into two categories: adenocarcinoma, and benign tissues. The dataset was increased from 1,250 pathology images to 25,000 images through image augmentation techniques such as flipping and rotating. The images were resized to a square format of 768×768 by cropping. The images used have been validated for privacy and comply with the Health Insurance Portability and Accountability Act regulations. Some sample images from the dataset are shown in Fig. 4.

2) GLaS DATASET

The dataset [24], encompasses a total of 165 images, which have been procured from the examination of 16 histological sections stained utilizing the Hematoxylin and Eosin (A&H) technique. These sections, sourced from distinct individuals and subjected to processing at different times in the laboratory, exhibit a pronounced inter-patient diversity in staining patterns and tissue morphological composition. Patient No-15's samples are shown in Fig. 5.

3) CRAG DATASET

The CRAG [23] corpus was instituted with the intent of facilitating segmentation; however, within the confines of our investigative foray, we sought to tackle a challenge that pertained to categorization. The visages within the CRAG dataset were systematically partitioned and designated by seasoned histopathologists into two distinct classes: Adenocarcinomas of the colon and innocuous colonic tissue.

The compendium consisted of one hundred and forty-eight specimens of malignant colonic neoplasms and forty-five instances of innocuous colonic tissue. A few samples are shown in Fig. 6.

4) THE KATHER MULTICLASS DATASET

The Kather Multiclass Dataset, an opus published in [25], is comprised of two distinct subsets, both of which consist of H&E stained tissue patches meticulously extracted from 86 colorectal tissue slides with the aid of manual delineation of tissue regions. The KMI subset encompasses a total of 100,000 histopathological tissue patches obtained at 20x magnification.

5) GGLCI DATASET

With the aim of addressing the challenge of image scarcity and increasing the quantity of available images, the GGLCI dataset was created by utilizing GAN. The configuration of the GAN utilized to synthesize a new dataset of images is represented in Fig. 7. The design of the generator encompasses three convolutional layers and deconvolutional layers and transforms a 100-dimensional random number array derived from the input data into a 64×64 pixel image. The deconvolutional layer featuring a 5×5 filter is interspersed with subsampling layers, including max pooling, *GeLU* activation, and batch normalization. The discriminator, consisting of three convolutional layers with 5×5 filters, assesses the authenticity of the generated image. The training of the GAN was accomplished by employing the Jensen-Shannon divergence metric to quantify the divergence between probability distributions. Adversarial networks serve as a means of gauging the disparity between the probabilities assigned to the generator's outputs and the actual data. This disparity assessment facilitates the training of the GAN, propelling it toward the production of images that bear a greater resemblance to the real data. The Jensen Shannon divergence [26] proves to be a valuable tool in the training of GANs due to its efficiency in optimization and its ability to furnish a smooth metric of similarity, thereby circumventing the obstacles posed by other divergence measures, such as Kullback Leibler divergence. The formula for Jensen-Shannon divergence is in Equation 9:

$$(X, Y) = \frac{(KL(X||M) + KL(Y||M))}{2} \tag{9}$$

where, X and Y are two probability distributions, $M = \frac{X+Y}{2}$ is the average of the two distributions, $KL(X||Y)$ is the Kullback Leibler divergence between X and Y.

To optimize the efficiency of the neural network, we set the learning rate at 0.001 and performed training for 1000 epochs utilizing a Lion optimizer, and we generated 2500 images. These images were produced by the GAN generator, which is trained to generate synthetic data samples that closely resemble the real data distribution. The generated samples are depicted in Fig. 8.
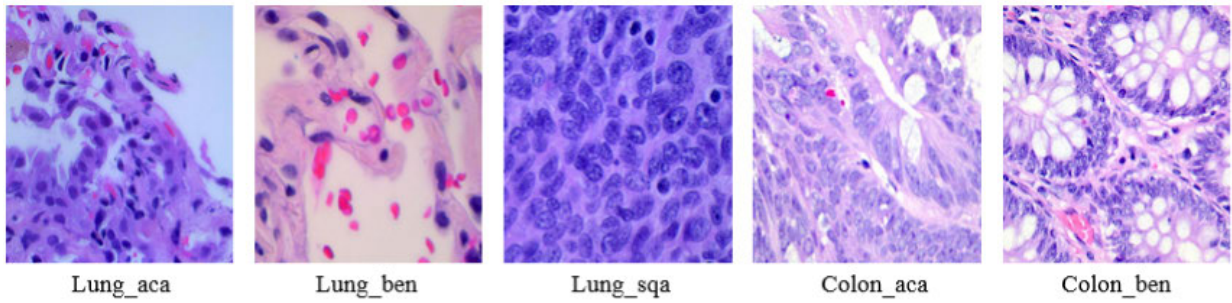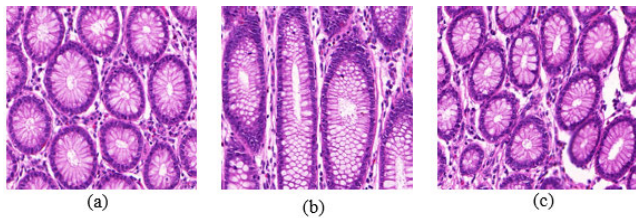
FIGURE 4. LC25000 dataset sample images.

FIGURE 5. GLaS dataset patient no-15's samples. (a) : Sample 1. (b): Sample 2. (c): Sample 3.

C. PERFORMANCE METRICS

Evaluation metrics play a crucial role in evaluating the resilience and effectiveness of deep learning or machine learning models. In order to comprehensively assess the performance of both scratch-trained and fine-tuned pre-trained models, the evaluation process was meticulously conducted by leveraging the five foremost evaluation metrics outlined in Equations 10-14:

$$A = \frac{TP + TN}{TP + TN + FP + FN} \tag{10}$$

$$P = \frac{TP}{TP + FP} \tag{11}$$

$$S = \frac{TP}{TP + FN} \tag{12}$$

$$\text{F1 score} = \frac{2 \times P \times S}{P + S} \tag{13}$$

$$\text{MCC} = \frac{(TP \times TN) - (FP \times FN)}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \tag{14}$$

where A, P, S, and MCC represent the accuracy, precision, sensitivity, and Matthews correlation coefficient, respectively and TP, FP, TN, and FN refer to true positive, false positive, true negative, and false negative. In multiclass classification, TP represents the number of instances correctly classified as belonging to a specific class. TN represents instances correctly classified as not belonging to that specific class for all other classes. FP are instances incorrectly classified as belonging to the specific class, while FN are instances incorrectly classified as not belonging to the specific class.

Inference time is the amount of time it takes for a model to process a new sample. FLOPs and MACs are
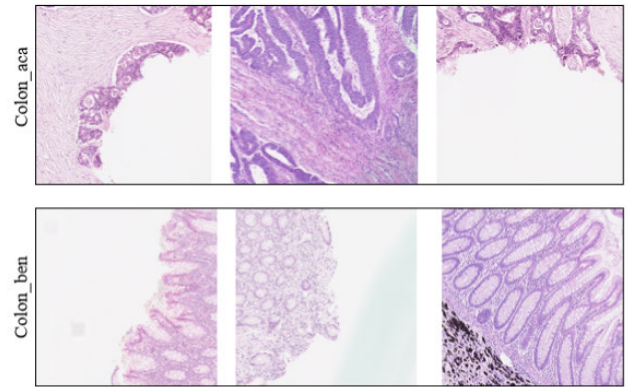

FIGURE 6. CRAG dataset samples.

essential metrics for evaluating computational complexity. FLOPs quantify the speed of arithmetic operations, including additions and multiplications with floating-point numbers. In contrast, MACs focus solely on counting multiply-accumulate operations, a critical element in various deep learning computations. Using Equations 15-16 FLOPs and MACs can be calculated:

$$\text{FLOPs} = (\text{Number of multiply-add operations per element})$$
$$\times (\text{Number of elements processed}) \tag{15}$$

$$\text{MACs} = (\text{Number of multiply operations per element})$$
$$\times (\text{Number of elements processed}) \tag{16}$$

Patient level accuracy is a commonly utilized evaluation criterion in medical image analysis [27] and is calculated as the fraction of patients that were correctly classified in a dataset. Patient level accuracy was measured using the Equations 17-18:

$$\text{Patient\_score} = \frac{P_C}{P_T} \tag{17}$$

$$\text{Patient\_accuracy} = \frac{\sum \text{Patient\_score}}{T_T} \tag{18}$$

where, P_C, P_T and T_T represent respectively image of patient P correctly identified, the total images of Patient P, and the total number of patients.

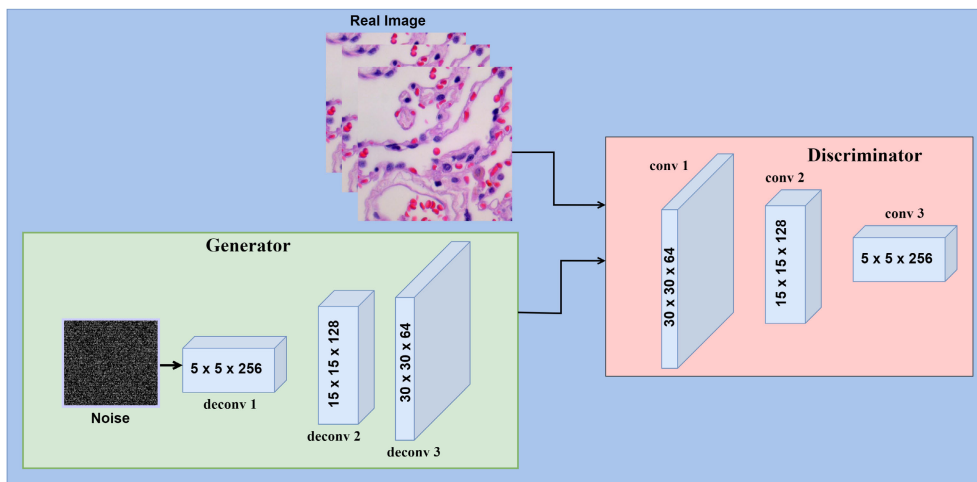In Equation 17, Patient_score is the ratio of correctly identified images of a patient and total images of that

FIGURE 7. Proposed generative adversarial network architecture consisting of generator and discriminator.
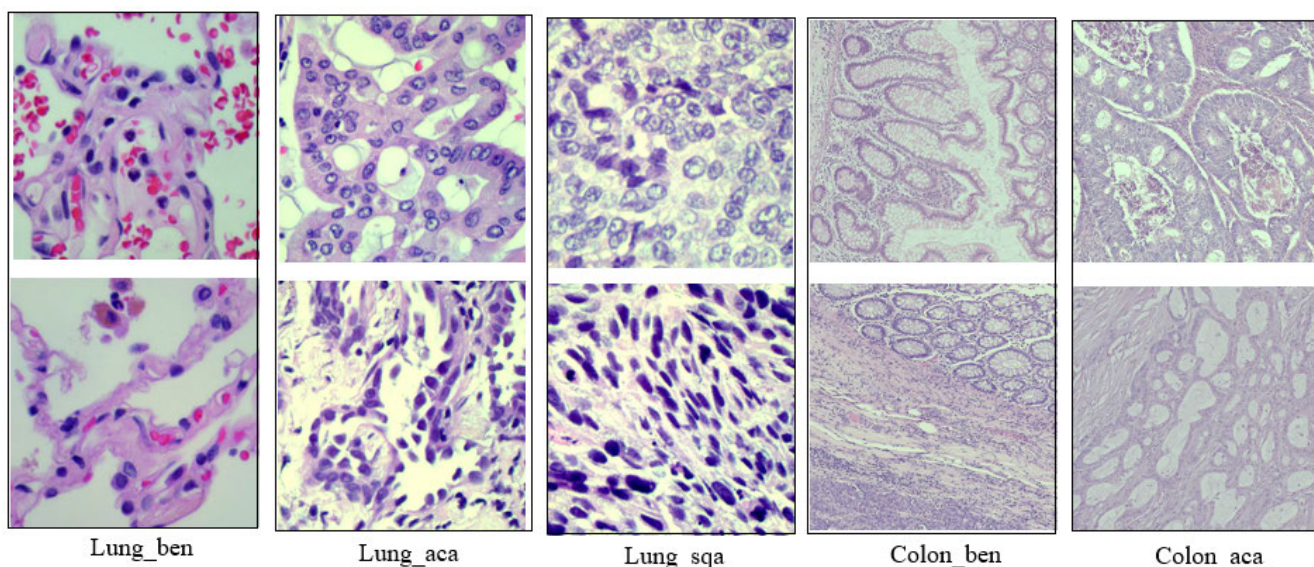

FIGURE 8. GAN generated image samples.

patient. In the Equation 18, Patient_accuracy is calculated by averaging patient scores.

IV. RESULTS AND DISCUSSION

In order to address the issue of limited availability of data in the medical domain, we generated additional images from our self-generated GGLCI dataset. We labeled the CRAG dataset to construct a more extensive and diverse dataset for the task of classifying lung and colon cancer histopathology images. The resulting combined dataset in Table 2 was then partitioned randomly into three distinct subsets, comprising training, testing, and validation data, in a ratio of 70:15:15. This enabled us to train and evaluate our proposed models using a sufficiently large and diverse set of labeled images, thereby enhancing the overall robustness and generalizability of our approach for accurate cancer classification.

In order to reduce the computational complexity of the framework, we modified the dimensions of the images to $64 \times 64 \times 3$. To illustrate the impact of this resizing operation, we provide representative examples of the rescaled images in Fig. 9.

After employing normalization to enhance the convergence speed of the model during its training, a performance comparison was conducted on both normalized and unnormalized data, as shown in Table 3. The table's results indicate superior performance was achieved in both image and patient level with batch normalization.

The categorical cross-entropy loss, with its intricate fusion of logarithmic transformations, multivariate probabilistic calculations, and intricate label differentiations, presents an arduous and labyrinthine endeavor that delves deep into the intricate nuances of image classification and is given by

TABLE 2. Combined dataset summary.

		Datasets			
Type	**Class Name**	LC25000	GGLCI	CRAG	**Total**
Colon benign	Col_ben	5000	500	45	5545
Colon adenocarcinoma	Colon_aca	5000	500	148	5648
Lung benign	Lung_ben	5000	500	-	5500
Lung adenocarcinoma	Lung_aca	5000	500	-	5500
Lung squamous cell carcinoma	Lung_squ	5000	500	-	5500

TABLE 3. Performance impact of batch normalization, where BN denotes batch normalization.

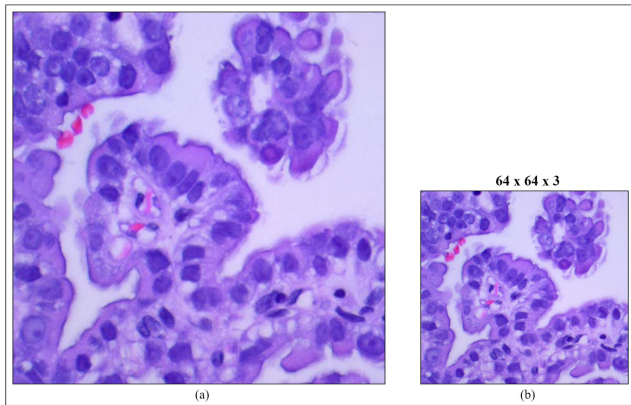|	Image level		Patient level	
BN	Training Accuracy	Testing Accuracy	Training Accuracy	**Epochs**
×	97.71	97.34	94.7	110
✓	**99.81**	**99.76**	**96.5**	**110**

FIGURE 9. Example of resizing. (a) : Original image. (b): Resized image.

Equation 19:

$$CCL = -log\ p(k) \tag{19}$$

where *CCL* denotes the categorical cross-entropy loss, and $p(k)$ refers to the probabilistic value of the class *k* that is fired up in the one hot vector.

We employ a comprehensive analysis to assess performance, juxtaposing the outcomes achieved through the conventional hyper-parameter training process with those realized using the one-cycle policy [43].

Table 4 depicted a thorough and comprehensive illustration of the intricately tuned hyperparameters using traditional fashion. It is evident that a batch size of 256 and an epoch value of 110 yielded the highest testing accuracy of 99.76%, showcasing their effectiveness in optimizing the neural network model. Additionally, the Lion optimizer and a dropout rate of 0.45 also achieved the same remarkable testing accuracy of 99.76%, highlighting their strong performance in enhancing the model's predictive capabilities.

A total of 110 epochs were used in order to evaluate the classification accuracy of the proposed model in Fig. 10(a) and revealed that the validation accuracy of the proposed model achieved the best outcome of 99.76% at the last

epoch. Furthermore, after the 83rd epoch, the accuracy remained stable, with the model consistently achieving the same maximum accuracy as the last epoch. Conversely, the training accuracy curve exhibited a gradual and almost steady increase, ultimately reaching the highest accuracy of 99.81% at the 85th epoch, and this value closely corresponds to the accuracy achieved in the final epoch. In contrast, the validation accuracy curve displayed occasional performance declines, indicating a level of instability. However, these declines were mitigated over time as the training process continued. In addition, the validation accuracy remained relatively stable after the 60th epoch, providing evidence that the proposed model is capable of providing robust and reliable classification outcomes. Fig.10(c) indicates minimized generalization error by avoiding overfitting.

Now, one cycle policy involves cycling the learning rate between specified bounds to avoid steep loss areas and find flatter minima. The optimal learning rate is chosen just before the loss begins to rise, while weight decay (L2 regularization) should be the largest that allows training at a high learning rate. Through a grid search of weight decays $(0.0001, 1e^{-5}, 1e^{-6})$ shown in Fig. 11(a), we select WD = 0.0001 for low loss and the highest learning rate before divergence, setting the initial learning rate at $1e^{-2}$, close to the loss bottom but still descending. Based on Fig. 11(b), following a learning rate range test, we opt for a maximum learning rate of $5e^{-4}$ and a minimum learning rate of $5e^{-5}$. In Fig. 11(c), a conspicuous learning rate pattern emerges: it begins modestly, surges to its zenith at the midpoint, then gracefully tapers off towards completion. This strategic maneuver entails a humble launch with a low initial learning rate, gradually escalating it to a loftier level. This elevated learning rate functions as a robust regulator, thwarting the model from settling into constrictive and precipitous local minima, thereby encouraging the exploration of broader, more resilient ones. As we traverse the midpoint of the learning rate cycle, we embark on a process of gradually diminishing the learning rate, signaling our entry into a promising, stable territory. Our objective at this stage
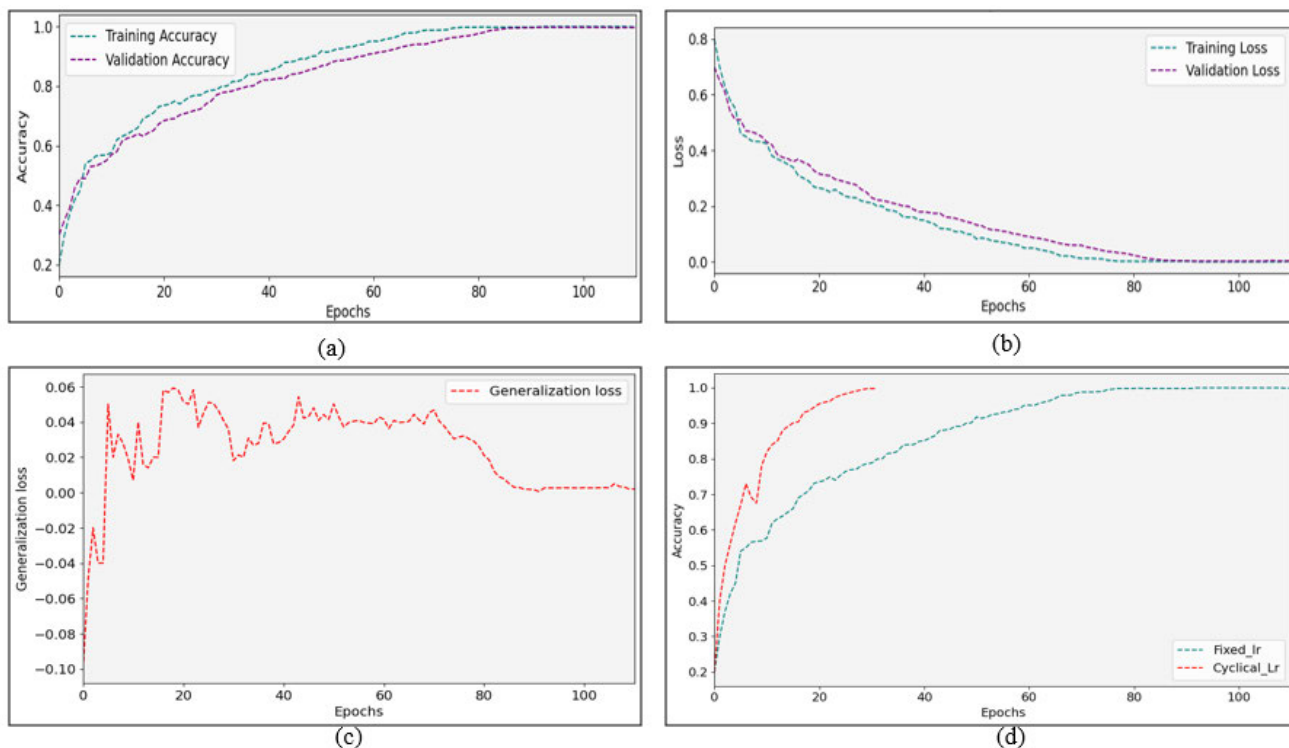
FIGURE 10. Combined dataset outcomes. (a) : Accuracy comparison curve (fixed lr). (b): Loss comparison curve (fixed lr). (c): Generalization loss curve (fixed lr), which is the gap between validation and training loss. (d): Observation of super convergence (validation accuracy).

TABLE 4. Hyperparameter configuration for fixed learning rate.

Hyperparameter	Range	Image level				Patient level
		Testing Accuracy	Training Accuracy	Testing Loss	Training Loss	Accuracy
Batch Size	128	97.23	97.81	2.77	2.19	95.31
	256	**99.76**	**99.91**	**0.24**	**0.09**	**96.5**
	512	96.52	97.32	3.48	2.68	94.19
Epoch	85	94.93	96.44	5.07	3.56	90.89
	100	97.81	99.89	2.19	0.11	94.56
	110	**99.76**	**99.91**	**0.24**	**0.09**	**96.5**
Optimizer	RMSProp	87.0	89.27	13.0	10.73	84.15
	Lion	**99.76**	**99.91**	**0.24**	**0.09**	**96.5**
	AdanW	97.3	98.67	2.7	1.33	93.88
Dropout	0.40	96.76	98.18	3.24	1.82	94.26
	0.45	**99.76**	**99.91**	**0.24**	**0.09**	**96.5**
	0.50	97.52	99.08	2.48	0.92	95.35
Learning rate	$5e^{-3}$	96.93	98.39	3.07	1.61	93.29
	$1e^{-3}$	**99.76**	**99.91**	**0.24**	**0.09**	**96.5**
	$2e^{-3}$	97.94	98.93	2.06	1.07	95.18

TABLE 5. Class wise performance of combined dataset.

Classes	Mean Accuracy	Precision	Sensitivity	F1 Score
Colon_aca		1.00	0.99	1.00
Colon_ben		1.00	1.00	1.00
Lung_aca	99.0	0.99	0.99	0.99
Lung_ben		1.00	1.00	1.00
Lung_sqa		0.99	0.99	0.99

pivots towards pinpointing the optimal minima within this dependable zone. Incorporating cyclical momentum (CM) into this strategy begins at its zenith. It gracefully wanes in tandem with the ascending learning rate, eventually

stabilizing at a value of 0.8 shown in Fig. 11(d). This infusion of cyclical momentum, paired with the LR range test, amplifies convergence stability, particularly when grappling with more substantial learning rate values, surpassing the benefits of a static momentum approach. In our performance comparison of different batch sizes, including 128, 256, and 512, it becomes apparent that the 512 batch size stands out as the superior choice within the one-cycle approach, achieving a superior balance in regularization that leads to enhanced performance. Observing Fig. 10(d), it is discernible that the one-cycle policy achieves comparable validation accuracy in a notably shorter timeframe, specifically within
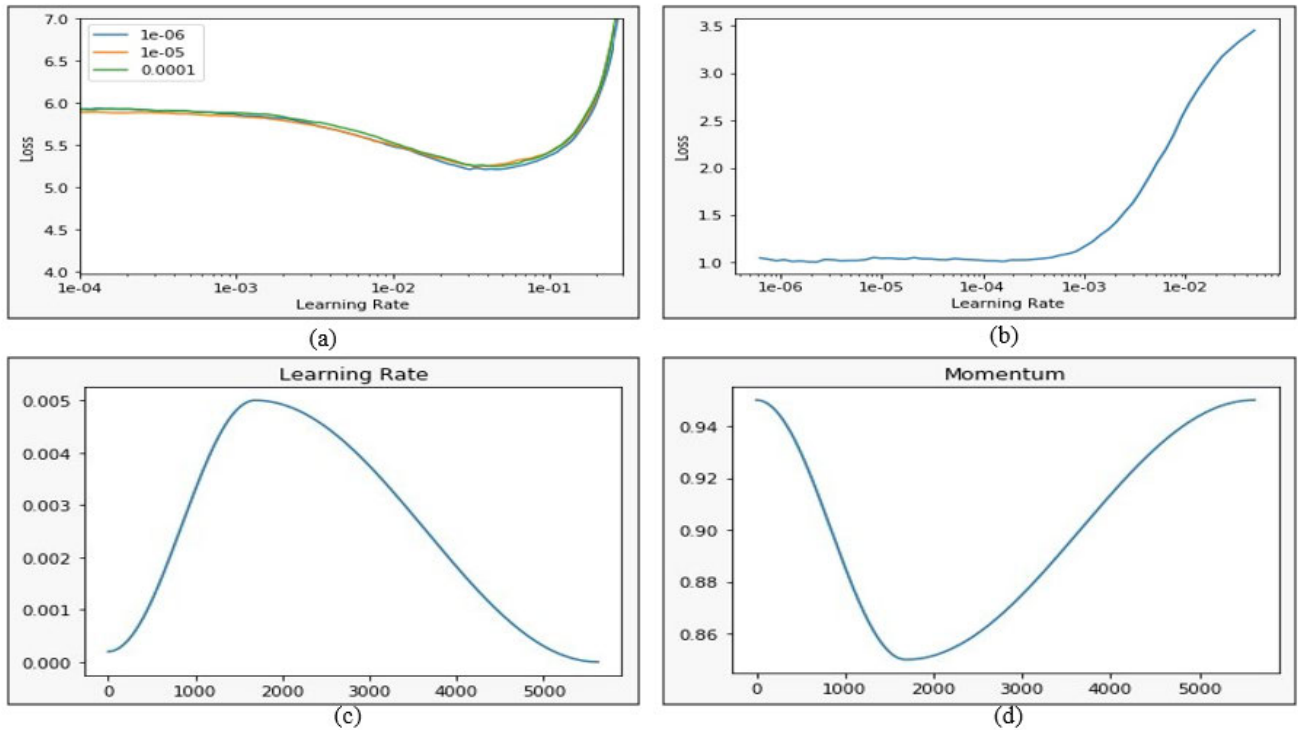
FIGURE 11. Hyperparameter tuning for one cycle policy. (a) : Tuning weight decay. (b): Learning rate range test. (c): Cyclical learning rate. (d): Cyclical momentum.
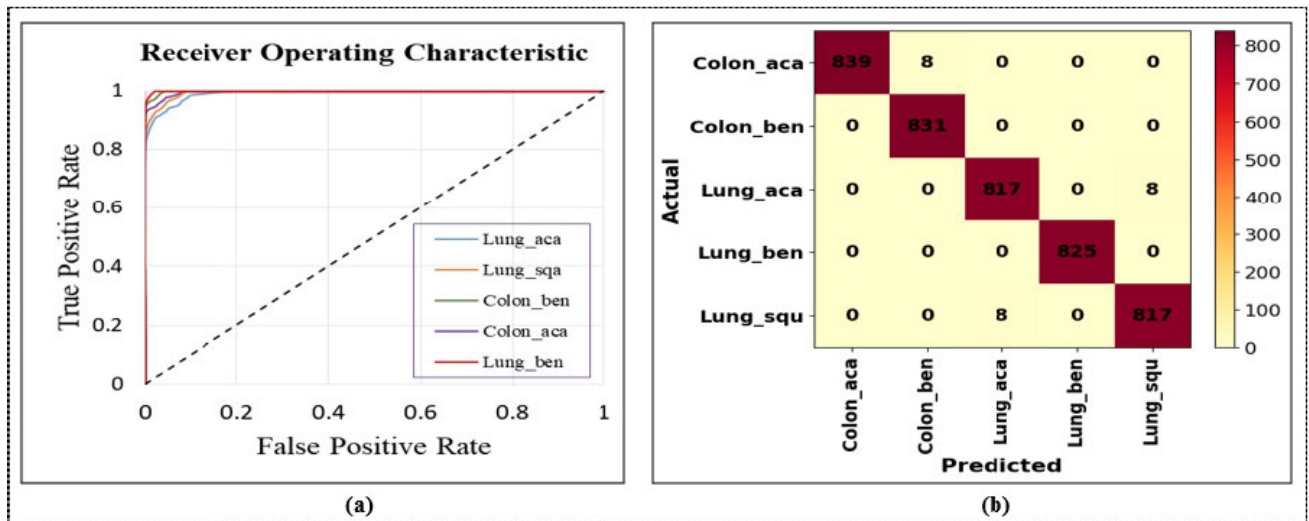

FIGURE 12. Combined dataset outcome. (a) : Receiver operating characteristic curve. (b): Confusion matrix.

the 32nd epoch. In contrast, the conventional hyperparameter training process attains the same accuracy level by the 110th epoch. Given the substantial reduction in training duration afforded by the one-cycle policy, we have opted to adopt this hyperparameter configuration for the remaining sections of our research paper.

The investigation presented in this work shows the utilization of a proposed model, where the testing subset at the 32nd epoch was classified, and the resulting confusion matrix and receiver operating characteristic (ROC) curves were analyzed, as displayed in Fig. 12. The assessment of the confusion matrix exposed the presence of misclassification in Lung_aca, with some Lung_aca erroneously classified as Lung_squ. These observations were further validated by the ROC curves, as the curves for Lung_ben and Colon_ben approached the top-left corner, indicating a successful classification of their respective samples. Nevertheless, the classifier encountered difficulties in categorizing the samples

TABLE 6. Performance evaluation for different Network configurations of scratch CNN where R, IT, F, MCC for the residual unit, inference time, F1-score, Matthews correlation coefficient, respectively.

|			Network configuration						Image			Patient
	Model	R	Neurons	#Params	FLOPs	MACs	IT		Accuracy	F	MCC	Accuracy
Model 1	CNN	1		1.27M	29.36M	28.14M	1.02s		90.01	92.31	91.98	87.05
		2							89.30	91.7	91.36	86.85
		3							88.51	91.13	91.03	86.41
Model 2	CNN+ECA-Net	1		1.36M	42.45M	41.53M	0.49s		94.01	94.45	94.13	89.52
		2							93.67	93.91	93.62	89.11
		3							93.13	93.35	93.17	88.65
Model 3	CNN+CBAM	1	128	1.81M	91.45M	90.66M	0.98s		96.92	96.70	96.81	93.33
		2							96.15	96.21	96.11	92.84
		3							95.85	95.35	95.16	92.17
Model 4	CNN+ECA-Net+GC	1		0.93M	27.87M	26.52M	0.31s		95.70	95.91	96.02	93.11
		2							95.22	95.15	94.78	92.90
		3							94.70	94.9	94.57	92.36
Model 5	**CNN+GC Attention Block**	**1**		**1.34M**	**80.91M**	**79.94M**	**0.95s**		**99.76**	**99.70**	**99.59**	**96.50**
		2							99.13	98.90	98.61	95.81
		3							98.55	98.25	97.91	95.27

belonging to the remaining three classes. In sum, the proposed model demonstrated a high level of efficiency, as evidenced by its performance values. To gain a more comprehensive understanding of the proposed model's performance, classification is provided in Table 5. Upon analyzing the results presented in Table 6, it becomes evident that among the three types of residual units, specifically type 1, utilizing a combination of the $3\times\{GeLU$-IC-seperable_conv-GC attention block$\}$ architecture, exhibits superior performance compared to the other types. In other words, this particular configuration outperforms the rest in terms of achieving better results, according to the metrics evaluated. The assessment results demonstrate the integration of different attention modules and a GC module with a CNN baseline. The CNN baseline achieved image-level accuracy of 90.01% and patient-level accuracy of 87.05%, comprising 1.27 million parameters and 29.36 million FLOPs. Incorporating the ECA-Net module resulted in an improvement of 6.14% and 2.47% in image level and patient-level accuracy, respectively. This came with an increase of 0.09 million parameters and 13.09 million FLOPs. Including the CBAM module alone in the architecture led to an accuracy improvement of 6.91% in image-level and 4.28% in patient-level accuracy compared to Model 1. It also introduced an increase of 0.54 million parameters and 62.09 FLOPs. The addition of the GC module in Model 2 not only reduced the number of parameters and FLOPs but also increased both image level and patient level accuracy. Finally, incorporating the GC and attention module resulted in a reduction of 0.47 million parameters and 10.54 million FLOPs. This module also improved the image level and patient level accuracy by 2.84% and 3.17%, respectively, compared to using CBAM alone. The GC attention module effectively reduces the number of FLOPs through global pooling while also experiencing a smaller decrease in parameters due to the introduction of an additional convolutional layer. The global pooling operation $\mathbf{f}_{pool}$ is applied to the input feature maps X, resulting in feature maps $\mathbf{f}_{gc}$ with spatial

dimensions reduced to 1×1 while preserving the number of channels (C). This operation significantly reduces FLOPs since it computes the average or maximum value across all spatial locations. The GC module improves accuracy by incorporating global semantic information through the global pooling operation. By considering the entire feature map during pooling, the module captures important GC, which can aid in lung and colon cancer classification tasks. Reducing the number of parameters and FLOPs in Model 5 provides benefits such as improved efficiency, faster inference times, reduced overfitting, more accessible training and deployment, model compression, and cost reduction. However, upon meticulous examination of the tabulated results, it becomes conspicuously evident that the proposed Model 5, featuring the incorporation of a residual unit connection 1 and combining a scratch CNN with both CBAM and GC attention blocks, has yielded an exceptional F1 score of 99.70% and a commendable MCC of 99.59%. The utilization of F1 score and MCC as evaluation metrics is particularly advantageous due to their ability to account for both precision and recall, thereby offering a well-balanced assessment of the model's performance. Moreover, their resilience in the case of class imbalance guarantees a faithful reflection of the model's capacity to detect instances of the minority class while upholding a stringent standard of precision. The results unequivocally manifest a notable reduction in inference time when utilizing the GC attention block is significant.

Through meticulous statistical analysis of Table 7, it's evident that Model 5 outshines the state-of-the-art models in histopathology image classification. The Matthews correlation coefficient (MCC) of 99.59 achieved by Model 5 stands out significantly, indicating its superior ability to handle both true positives and true negatives. Moreover, the accuracy (A) of 99.4 attests to its exceptional overall classification performance. While some transfer learning models demonstrate strong results, they fall short of Model 5's performance, even with their larger parameter counts and

TABLE 7. Performance comparison of transfer learning models and our proposed model 5 on the test set where IT, A, MCC for inference time, accuracy, Matthews correlation coefficient respectively.

| | Neurons | #Params | FLOPs | MACs | IT | Image | | | | | Patient |
						Precision	Sensitivity	F1-score	A	MCC	A
DenseNet121		7.087M	2834.41M	2833.26M	1.3s	96.13	89.07	87.03	88.39	86.87	84.74
VGG16		14.78M	15355.26M	15347.11M	1.96s	95.76	85.12	90.3	89.73	90.05	84.96
ResNet50	128	22.80M	3858.50M	3856.19M	1.03s	97.3	89.23	90.80	90.91	90.91	86.53
InceptionV3		23.80M	2836.79M	2836.14M	1.35s	95.91	87.71	87.30	91.24	87.14	87.21
EfficientNet-B3		7.16M	1271.17M	1267.67M	1.07s	92.43	91.79	91.85	92.11	91.47	88.05
Xception		22.8M	19100.24M	19099.61M	0.72s	92.22	93.47	92.69	93.23	92.57	88.42
Model 5		**1.34M**	**80.91M**	**79.94M**	**0.95s**	**99.4**	**99.6**	**99.7**	**99.76**	**99.59**	**96.50**
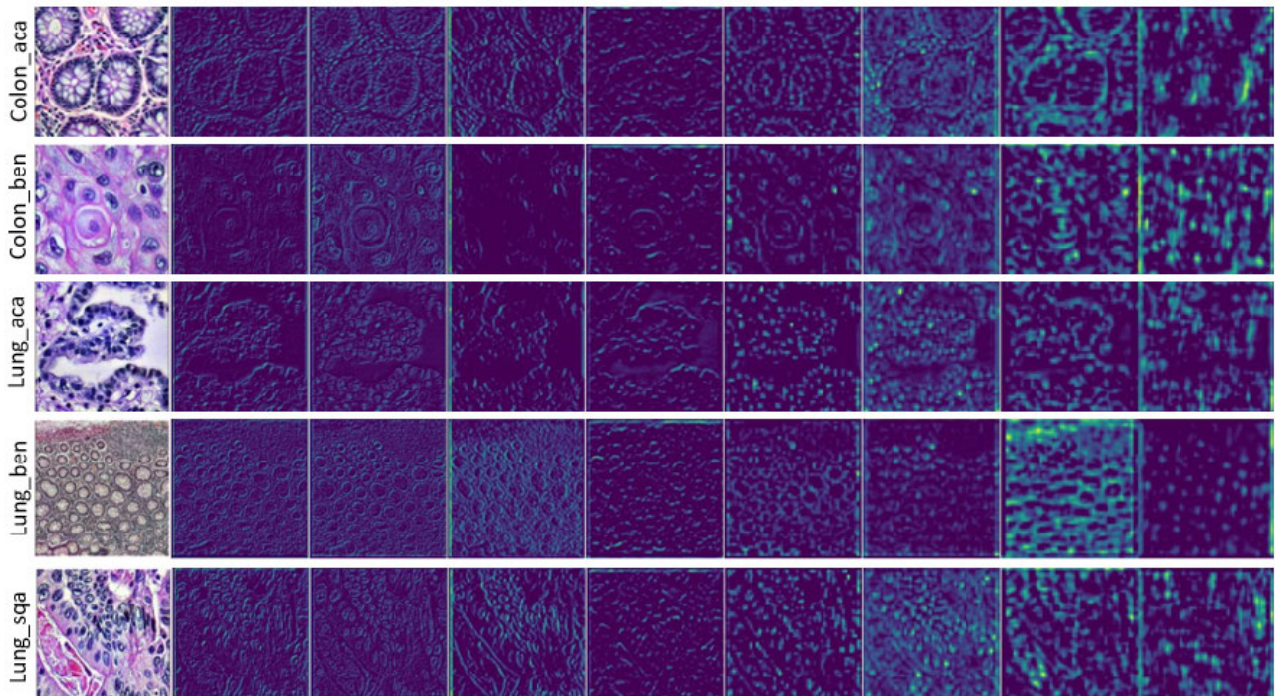
FIGURE 13. Each row showing low-to-high level convolutional feature maps using the proposed model 5.

TABLE 8. Ablation study for scratch CNN where AI and AP denote image level accuracy and patient level accuracy, respectively.

	Stage based enhancement	GGLCI+Labeled images	Attention module	GC Attention module	AI	AP
Baseline	✗	✗	✗	✗	86.4	83.5
Experiment 1	✓	✗	✗	✗	89.4	87.2
Experiment 2	✓	✓	✗	✗	92.3	89.6
Experiment 3	✓	✓	✓	✗	96.9	93.1
Experiment 4	✓	✓	✗	✓	**99.7**	**96.5**

computational requirements. Additionally, Model 5 manages to maintain a competitive inference time of 0.95s, making it a practical choice for real-world applications. This combination of high accuracy, balanced performance, and efficiency underpins the compelling case for Model 5's superiority in histopathology image classification. This disparity can be attributed to the incongruity in nature between the benchmark datasets and the histopathology slides, leading to the irrelevance of the low-level features retrieved through transfer learning methods that are reliant on benchmark datasets.

The feature map produced by the convolutional layer of a CNN encapsulates complex representations of local patterns, textures, and visual attributes, exhibiting spatial hierarchies that encode high-level semantic information important for cancer classification in lung and colon histopathological images. The higher-level feature maps exhibit larger and less comprehensible features than those typically processed by humans. The arrangement of feature maps in Fig. 13, based on the convolutional part of Model 5 from the lower level to higher level layers, demonstrates that the convolutional architecture of the model primarily captures irregular shapes,

(a)

(b)
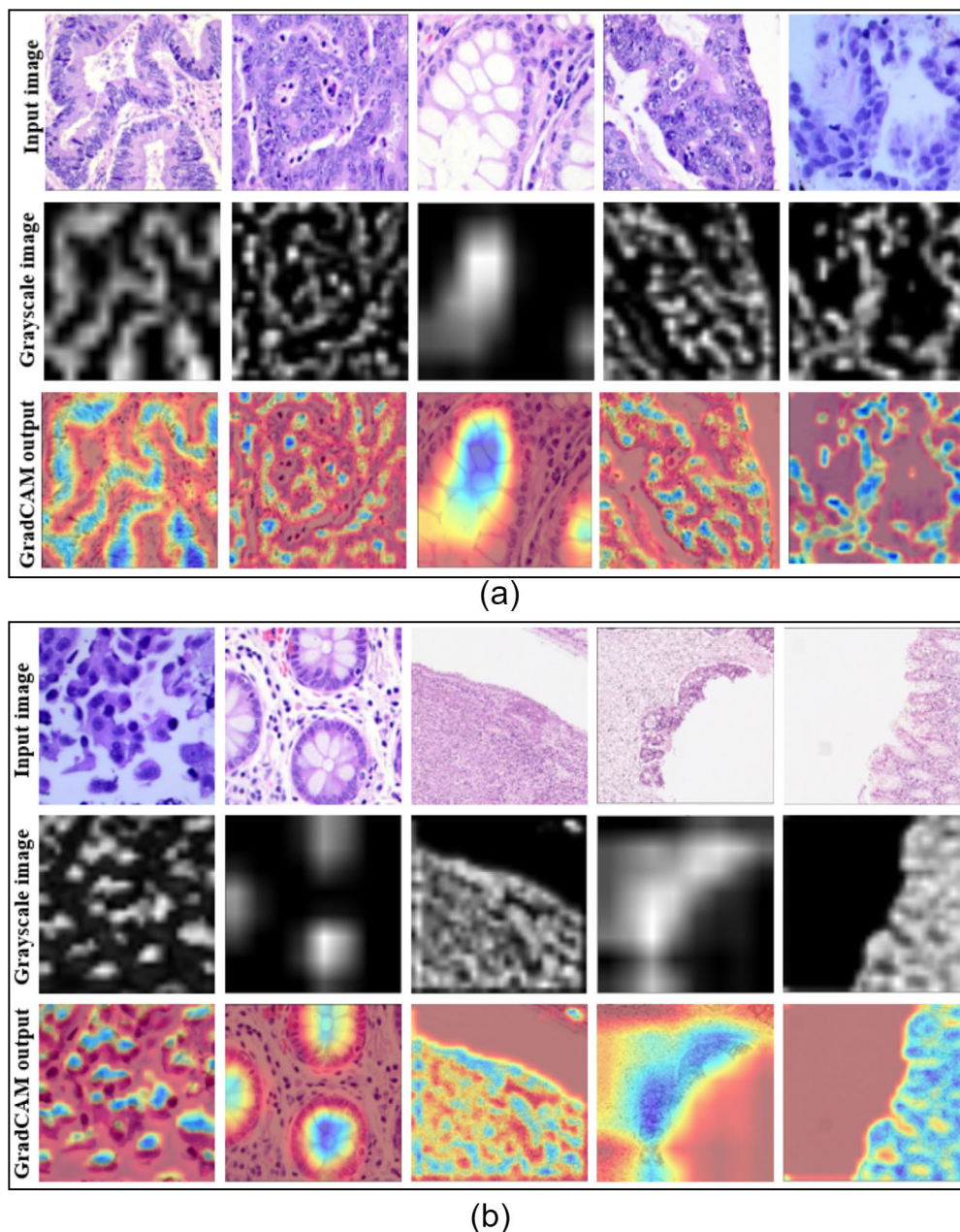
FIGURE 14. Examples of GradCAM visualization outcome. (a) : Lung cancer samples. (b): Colon cancer samples.

disorganized cell arrangements, abnormal cell size and shape, hyperchromatic nuclei, increased mitotic activity, and nuclear to cytoplasmic ratio as features in the lower-level layers, while deeper layers extract increasingly encoded features that lack human readability.

The GradCAM visualization, a sophisticated tool utilized by researchers to gain insight into the decision-making mechanisms of deep neural networks, provides a means to comprehend the inner workings of the network and discover any potential biases or deficiencies when combined with grayscale heatmaps, offering valuable insight into the network's prediction outcomes through the illumination of

the attention it pays to specific regions of the image, as exemplified by a few representative samples of GradCAM visualization and grayscale heat maps displayed in Fig. 14, thus allowing for the fine-tuning of the network and enhancement of its overall performance by directing attention to critical areas of the image.

By implementing the same network configuration and hyperparameters as used in the table of the combined dataset, the performance evaluation of the proposed method was conducted on the Kather Multiclass dataset. The dataset was divided into 70%, 15%, and 15% portions for training, testing, and validating the models, respectively. The obtained results
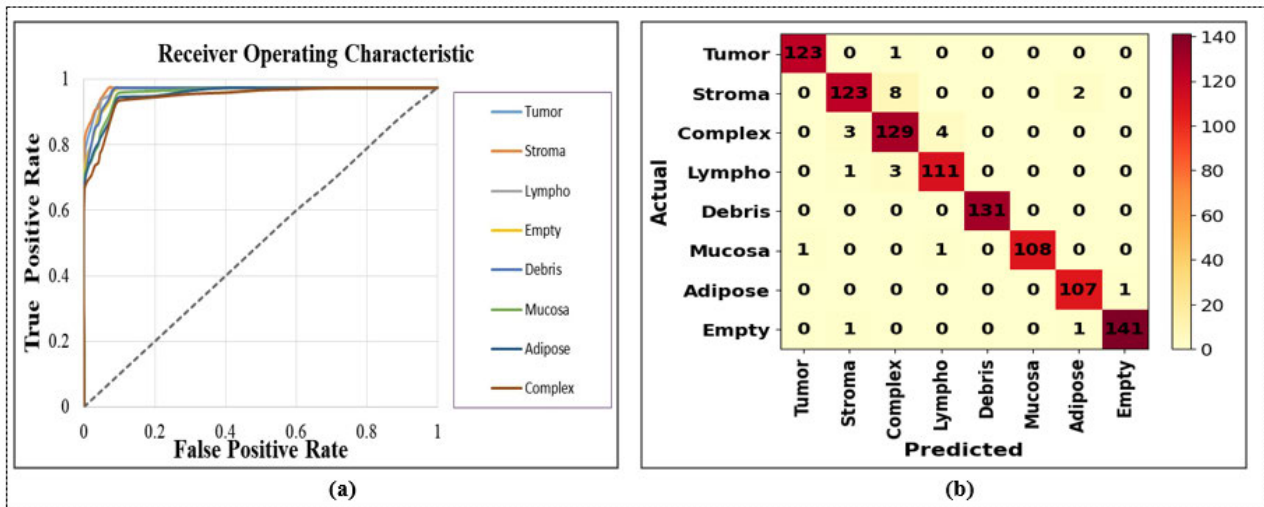
FIGURE 15. Kather Multiclass dataset outcome. (a) : Receiver operating characteristic curve. (b): Confusion matrix.

revealed that the proposed method achieved the highest F1-score of 97.5%, suggesting its competence not only on specific datasets but also on other datasets. Furthermore, the robustness of the proposed method's performance across various datasets is supported by the confusion matrix and ROC curve displayed in Fig. 15. Conducting a series of ablation experiments, we progressively incorporate enhanced modules into the model alongside the scratch CNN to assess their dependability and significance. These experiments aim to evaluate the model's performance based on different assessment criteria. The results of these rigorous investigations are presented in Table 8, providing a comprehensive summary of our findings. From the table, the baseline model, without any additional modules, achieved an accuracy of 86.4 at the image level and 83.5 at the patient level while experiment 1 introduces stage-based enhancement, resulting in 3% and 3.7% at image and patient levels, respectively improvement in performance compared to the baseline. In experiment 2, in addition to stage-based enhancement, GGLCI and labeled images are used, further improving the performance and in experiment 3 attention module is introduced, leading to a significant boost in performance with an accuracy of 96.9% and 93.1% at image and patient level respectively. Finally, experiment 4 combines stage-based enhancement, GGLCI, and GC attention module and achieves the highest performance among all experiments, with an accuracy of 96.9% and 93.1% at the image and patient level, respectively. at image and patient level, respectively. and the improvement over the baseline is substantial at 13.3% and 13% at image and patient levels, respectively.

In this work, the effectiveness of the proposed multiclass classification model was assessed and compared to the results obtained by previous researchers in the field, as depicted in Table 9. Despite the non-uniformity of the datasets utilized by these researchers, a comparison was drawn based on the common objective of the task. Our model exhibited

exceptional performance with a remarkable classification accuracy of 99.76%, effectively reducing the discrepancy between the two forms of cancer, lung adenocarcinoma, and lung squamous cell carcinoma, and signifying a relative improvement. Prior studies neglected to incorporate residual blocks, which are crucial components for constructing deeper network configurations and alleviating the vanishing gradient issue. The implementation of skip connections within residual networks also facilitates the efficient flow of information across the network, thereby enabling optimization. Despite this, the shallow network architecture of CNN models results in restricted feature extraction and an inclination towards overfitting. Our proposed model consists of three residual blocks and employs separable convolution layers. Furthermore, the IC layer in our model promotes greater autonomy among neurons and counteracts the occurrence of wandering optimization patterns that could be produced by the singular usage of batch normalization.

In the medical image analysis domain, patient-level accuracy is considered as a primary evaluation metric, and the results demonstrate exceptional performance, with a classification accuracy rate of 96.5% for the GLaS dataset. It is crucial to report patient-level accuracy as it provides a more comprehensive understanding of the model's overall performance beyond individual image accuracy calculation, which is particularly significant in the medical imaging domain, where the objective is to make precise diagnoses at the patient level. Furthermore, the proposed method was compared to a CNN-based classification technique for the Kather Multiclass dataset, achieving a notable improvement of 1.1% in accuracy, indicating its potential to accurately classify images. Table 10, establishes the efficacy of the proposed approach in accurately classifying images and highlights its potential to make a significant contribution to the advancement of the image classification field.

TABLE 9. Author accuracy comparison.

Title	Model	Accuracy	Precision	Sensitivity	F1-score
W. Shen et al. [28], 17	MC-CNN	87.14	-	93	-
Z. Yuan et al. [15], 17	AlexNet	87.14	-	91.76	-
A. Masood et al. [29], 18	DFCNet	89.52	-	82.54	-
M. Akbari et al. [16], 18	CNN	90.28	74.34	68.32	71.2
D. Nobrega et al. [30], 20	ResNet50 + SVM RBF	93.19	73.48	85.38	78.83
S. Suresh et al. [31], 20	CNN	93.9		93.4	
Bukhari et al. [32], 20	ResNet-50	93	93.74	93.77	93.26
Liang et.al, [38],20	MFF-CNN	96	97.4	94.5	95.9
Tasnim et.al, [39],21	CNN	96.06	97.02	95.48	96.57
M.Masud et.al, [40],21	Multi channel CNN	96.33	96.39	96.37	96.38
Lee et al. [18], 21	SSD	90.67	97.44	90.67	93.93
Tan et al. [11], 22	Customized VGG-16	90.4	98.5	95.5	90.1
Jiatai et al., [34], 22	DBL+PDBL	96.4	-	-	-
Proposed	**CNN + GC attention block**	**99.76**	**99.6**	**99.4**	**99.7**

TABLE 10. Kather multiclass dataset accuracy comparison.

Method	Year	Accuracy
Kather et al., [25]	2021	94.3
Hatami et al., [33]	2021	95.9
Paladini et.al, [41]	2021	96.2
Jiatai et al., [34]	2022	96.4
Raju et al., [42]	2022	97.1
Ours	**2023**	**97.5**

Our work is associated with a number of notable limitations. Firstly, despite our best efforts to generate a comprehensive dataset for training the model, the possibility exists that further expansion of the dataset with additional histopathological slides may be required to optimize the overall performance of the model. Secondly, although the developed methodology demonstrated impressive performance, misclassification continued to occur due to the intrinsic similarity in features. Thirdly, the dataset did not include images of adenosquamous carcinoma lung cancer subtypes, a combination of squamous cell carcinoma and adenocarcinoma, due to unavailability, making it difficult to determine how the classification of these subtypes may impact the overall accuracy of the model.

V. CONCLUSION

Lung and colon cancer histopathological image classification is performed in this paper through a fresh lens by contemplating how to capture both channel and spatial attention with semantic information from the image in order to further improve the performance of existing works. An attention module called the GC attention module for CNN backbone is proposed, which results in a significant improvement in accuracy, attaining 99.76% and 96.5%, respectively, at the image and patient level. Moreover, it reduces 0.47 million parameters, 10.72 million MACs, 10.54 million FLOPs, and 0.03s inference time. This work is the first, to the best of our knowledge, to address both patient and image-level accuracy. Furthermore, a notable F1 score of 99.7% and an MCC of 99.59%, both considered as the preferred evaluation metrics for imbalanced datasets, are attained by increasing the dataset size with the help of GAN, labeling CRAG by expert histopathologist. In the future, the dataset can be expanded to include even more challenging types of histopathological images. There is room for improvement by fine-tuning the GC attention block and experimenting with various vision transformers. Additionally, we can tap into multiple data sources, including genomics, radiomics, and clinical data, combining them with the power of deep learning models to discover new biomarkers. This approach could lead to a shift in how we approach personalized medicine.

REFERENCES

[1] International Angry for Research on Cancer. (2020). *World Fact Sheet*. Accessed: 26, 2022. [Online]. Available: https://gco.iarc.fr/today/data/factsheets/populations/900-world-fact-sheets.pdf

[2] International House World Organisation. (2022). *Cancer*. Accessed: Jun. 26, 2022. [Online]. Available: https://www.who.int/news-room/factsheets/detail/cancer/

[3] Mayo Clinic. (2020). *Cancer Symptoms Causes*. Accessed: Jun. 26, 2022. [Online]. Available: https://www.mayoclinic.org/diseasesconditions/cancer/symptoms-causes/syc-20370588

[4] T. N. Seyfried and L. C. Huysentruyt, "On the origin of cancer metastasis," *Crit. Rev. Oncogenesis*, vol. 18, nos. 1–2, Jan. 2013, doi: 10.1615/CritRevOncog.v18.i1-2.40.

[5] Verywellhealth. (2022). *What Is Metastasis?* Accessed: Jun. 27, 2022. [Online]. Available: https://www.verywellhealth.com/metastatic-cancer-2249128/

[6] L. F. Sánchez-Peralta, L. Bote-Curiel, A. Picón, F. M. Sánchez-Margallo, and J. B. Pagador, "Deep learning to find colorectal polyps in colonoscopy: A systematic literature review," *Artif. Intell. Med.*, vol. 108, Aug. 2020, Art. no. 101923, doi: 10.1016/j.artmed.2020.101923.

[7] C-Health. (2022). *Cancer Survival Rates*. Accessed: Jun. 26, 2022. [Online]. Available: https://cancersurvivalrates.com/type=colon&role=patient/

[8] S. Das, S. Biswas, A. Paul, and A. Dey, *Industry Interactive Innovations in Science, Engineering and Technology*, S. Bhattacharyya, S. Sen, M. Dutta, P. Biswas, H. Chattopadhyay, Eds. Singapore: Springer, 2018, pp. 173–183, doi: 10.1007/978-981-10-3953-9_17.

[9] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, and Q. He, "A comprehensive survey on transfer learning," 2019, *arXiv:1911.02685*.

[10] K. Doi, "Computer-aided diagnosis in medical imaging: Historical review, current status and future potential," *Computerized Med. Imag. Graph.*, vol. 31, nos. 4–5, pp. 198–211, Jun. 2007, doi: 10.1016/j.compmedimag.2007.02.002.

[11] H. Tan, J. H. T. Bates, and C. M. Kinsey, "Discriminating TB lung nodules from early lung cancers using deep learning," *BMC Med. Informat. Decis. Making*, vol. 22, no. 1, pp. 1–7, Dec. 2022.

[12] Y. Shi, Y. Gao, Y. Yang, Y. Zhang, and D. Wang, "Multimodal sparse representation-based classification for lung needle biopsy images," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 10, pp. 2675–2685, Oct. 2013, doi: 10.1109/TBME.2013.2262099.

[13] J. Kuruvilla and K. Gunavathi, "Lung cancer classification using neural networks for CT images," *Comput. Methods Programs Biomed.*, vol. 113, no. 1, pp. 202–209, Jan. 2014, doi: 10.1016/j.cmpb.2013.10.011.

[14] C. Kuepper, F. Großerueschkamp, A. Kallenbach-Thieltges, A. Mosig, A. Tannapfel, and K. Gerwert, "Label-free classification of colon cancer grading using infrared spectral histopathology," *Faraday Discuss.*, vol. 187, pp. 105–118, 2016, doi: 10.1039/C5FD00157A.

[15] Z. Yuan, M. IzadyYazdanabadi, D. Mokkapati, R. Panvalkar, J. Y. Shin, N. Tajbakhsh, S. Gurudu, and J. Liang, "Multiclass semantic segmentation and quantification of colorectal cancer sections using deep learning," *Proc. SPIE*, vol. 10133, Jun. 2017, Art. no. 101332K.

[16] M. Akbari, M. Mohrekesh, S. Rafiei, S. M. Reza Soroushmehr, N. Karimi, S. Samavi, and K. Najarian, "Classification of informative frames in colonoscopy videos using convolutional neural networks with binarized weights," in *Proc. 40th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2018, pp. 65–68, doi: 10.1109/EMBC.2018.8512226.

[17] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 4401–4410, doi: 10.1109/CVPR.2019.00453

[18] K.-S. Lee, S.-H. Son, S.-H. Park, and E. S. Kim, "Automated detection of colorectal tumors based on artificial intelligence," *BMC Med. Informat. Decis. Making*, vol. 21, no. 1, pp. 1–6, Dec. 2021.

[19] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," 2015, *arXiv:1511.06434*.

[20] S. Woo, J. Park, J. Y. Lee, and J. Ha, "CBAM: Convolutional block attention module," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 3–19.

[21] G. Chen, P. Chen, Y. Shi, C.-Y. Hsieh, B. Liao, and S. Zhang, "Rethinking the usage of batch normalization and dropout in the training of deep neural networks," 2019, *arXiv:1905.05928*.

[22] A. A. Borkowski, M. M. Bui, L. Brannon Thomas, C. P. Wilson, L. A. DeLand, and S. M. Mastorides, "Lung and colon cancer histopathological image dataset (LC25000)," 2019, *arXiv:1912.12142*.

[23] S. Graham, H. Chen, J. Gamper, Q. Dou, P.-A. Heng, D. Snead, Y. W. Tsang, and N. Rajpoot, "MILD-net: Minimal information loss dilated network for gland instance segmentation in colon histology images," *Med. Image Anal.*, vol. 52, pp. 199–211, Feb. 2019.

[24] K. Sirinukunwattana, J. P. W. Pluim, H. Chen, X. Qi, P.-A. Heng, Y. B. Guo, L. Y. Wang, B. J. Matuszewski, E. Bruni, U. Sanchez, A. Böhm, O. Ronneberger, B. B. Cheikh, D. Racoceanu, P. Kainz, M. Pfeiffer, M. Urschler, D. R. J. Snead, and N. M. Rajpoot, "Gland segmentation in colon histology images: The glas challenge contest," *Med. Image Anal.*, vol. 35, pp. 489–502, Jan. 2017, doi: 10.1016/j.media.2016.08.008.

[25] J. N. Kather, J. Krisam, P. Charoentong, T. Luedde, E. Herpel, C.-A. Weis, T. Gaiser, A. Marx, N. A. Valous, D. Ferber, L. Jansen, C. C. Reyes-Aldasoro, I. Zörnig, D. Jäger, H. Brenner, J. Chang-Claude, M. Hoffmeister, and N. Halama, "Predicting survival from colorectal cancer histology slides using deep learning: A retrospective multicenter study," *PLOS Med.*, vol. 16, no. 1, Jan. 2019, Art. no. e1002730, doi: 10.1371/journal.pmed.1002730.

[26] J. Lin, "Divergence measures based on the Shannon entropy," *IEEE Trans. Inf. Theory*, vol. 37, no. 1, pp. 145–151, Jan. 1991.

[27] Y. Hao, L. Zhang, S. Qiao, Y. Bai, R. Cheng, H. Xue, Y. Hou, W. Zhang, and G. Zhang, "Breast cancer histopathological images classification based on deep semantic features and gray level co-occurrence matrix," *PLoS ONE*, vol. 17, no. 5, May 2022, Art. no. e0267955, doi: 10.1371/journal.pone.0267955.

[28] W. Shen, M. Zhou, F. Yang, D. Yu, D. Dong, C. Yang, Y. Zang, and J. Tian, "Multi-crop convolutional neural networks for lung nodule malignancy suspiciousness classification," *Pattern Recognit.*, vol. 61, pp. 663–673, Jan. 2017.

[29] A. Masood, B. Sheng, P. Li, X. Hou, X. Wei, J. Qin, and D. Feng, "Computer-assisted decision support system in pulmonary cancer detection and stage classification on CT images," *J. Biomed. Informat.*, vol. 79, pp. 117–128, Mar. 2018, doi: 10.1016/j.jbi.2018.01.005.

[30] R. V. M. da Nóbrega, P. P. R. Filho, M. B. Rodrigues, S. P. da Silva, C. M. D. Junior, and V. H. C. de Albuquerque, "Lung nodule malignancy classification in chest computed tomography images using transfer learning and convolutional neural networks," *Neural Comput. Appl.*, vol. 32, no. 15, pp. 11065–11082, Aug. 2020.

[31] S. Suresh and S. Mohan, "ROI-based feature learning for efficient true positive prediction using convolutional neural network for lung cancer diagnosis," *Neural Comput. Appl.*, vol. 32, no. 20, pp. 15989–16009, Oct. 2020.

[32] S. U. K. Bukhari, A. Syed, S. K. A. Bokhari, S. S. Hussain, S. U. Armaghan, and S. S. H. Shah, "The histological diagnosis of colonic adenocarcinoma by applying partial self supervised learning," *MedRxiv*, Aug. 2020, doi: 10.1101/2020.08.15.20175760.

[33] N. Hatami, M. Bilal, and N. Rajpoot, "Deep multi-resolution dictionary learning for histopathology image analysis," 2021, *arXiv:2104.00669*.

[34] J. Lin, G. Han, X. Pan, Z. Liu, H. Chen, D. Li, X. Jia, Z. Shi, Z. Wang, Y. Cui, H. Li, C. Liang, L. Liang, Y. Wang, and C. Han, "PDBL: Improving histopathological tissue classification with plug-and-play pyramidal deep-broad learning," *IEEE Trans. Med. Imag.*, vol. 41, no. 9, pp. 2252–2262, Sep. 2022.

[35] X. Zhu, K. Guo, H. Fang, L. Chen, S. Ren, and B. Hu, "Cross view capture for stereo image super-resolution," *IEEE Trans. Multimedia*, vol. 24, pp. 3074–3086, 2022.

[36] X. Zhu, K. Guo, S. Ren, B. Hu, M. Hu, and H. Fang, "Lightweight image super-resolution with expectation-maximization attention mechanism," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 32, no. 3, pp. 1273–1284, Mar. 2022.

[37] X. Zhu, K. Guo, T. Qiu, H. Fang, Z. Wu, X. Tan, and C. Liu, "Stereoscopic image super-resolution with interactive memory learning," *Exp. Syst. Appl.*, vol. 227, Oct. 2023, Art. no. 120143.

[38] M. Liang, Z. Ren, J. Yang, W. Feng, and B. Li, "Identification of colon cancer using multi-scale feature fusion convolutional neural network based on shearlet transform," *IEEE Access*, vol. 8, pp. 208969–208977, 2020.

[39] Z. Tasnim, S. Chakraborty, F. M. J. M. Shamrat, A. N. Chowdhury, H. A. Nuha, A. Karim, S. B. Zahir, and M. M. Billah, "Deep learning predictive model for colon cancer patient using CNN-based classification," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 8, pp. 687–696, 2021.

[40] M. Masud, N. Sikder, A.-A. Nahid, A. K. Bairagi, and M. A. AlZain, "A machine learning approach to diagnosing lung and colon cancer using a deep learning-based classification framework," *Sensors*, vol. 21, no. 3, p. 748, Jan. 2021.

[41] E. Paladini, E. Vantaggiato, F. Bougourzi, C. Distante, A. Hadid, and A. Taleb-Ahmed, "Two ensemble-CNN approaches for colorectal cancer tissue type classification," *J. Imag.*, vol. 7, no. 3, p. 51, Mar. 2021.

[42] M. S. N. Raju and B. S. Rao, "Colorectal multi-class image classification using deep learning models," *Bull. Electr. Eng. Informat.*, vol. 11, no. 1, pp. 195–200, Feb. 2022.

[43] L. N. Smith, "A disciplined approach to neural network hyper-parameters: Part 1—Learning rate, batch size, momentum, and weight decay," 2018, *arXiv:1803.09820*.

MD. AL-MAMUN PROVATH was born in Jamalpur, Bangladesh, in 1999. He received the B.Sc. degree in computer science and engineering from the Chittagong University of Engineering and Technology (CUET), Chattogram, Bangladesh, in 2023, where he is currently pursuing the M.Sc. degree in computer science and engineering. He is currently a Lecturer with the Department of Computer Science and Engineering, United International University (UIU). His current research interests include image classification, image reconstruction, image enhancement, machine learning, deep learning, data science, and computer vision.

KAUSHIK DEB received the B.Tech. and M.Tech. degrees from the Department of Computer Science and Engineering, Tula State University, Tula, Russia, in 1999 and 2000, respectively, and the Ph.D. degree in electrical engineering and information systems from the University of Ulsan, Ulsan, South Korea, in 2011. Since 2001, he has been a Faculty Member of the Department of Computer Science and Engineering (CSE), Chittagong University of Engineering and Technology (CUET), Chattogram, Bangladesh, where he is currently a Professor. Moreover, he was in various administrative positions with CUET, such as the Dean of the Faculty of Electrical and Computer Engineering (ECE), from 2017 to 2019, the Director of the Institute of Information and Communication Technology (IICT), from 2015 to 2017, and the Head of the CSE Department, from 2012 to 2015. He made a variety of contributions to managing and organizing conferences, workshops, and other academic gatherings. He has published more than 115 peer-reviewed technical articles. His current research interests include computer vision, deep learning, pattern recognition, intelligent transportation systems (ITSs), and human–computer interaction. He was a Steering Member. He acted as the Chair or a Secretary at a variety of international and national conferences, such as the International Conference on Electrical, Computer, and Communication Engineering (ECCE), the International Forum on Strategic Technology (IFOST), the International Workshops on Human System Interactions (HSI), and the National Conference on Intelligent Computing and Information Technology (NCICIT).

PRANAB KUMAR DHAR received the B.Sc. degree from the Chittagong University of Engineering and Technology (CUET), Chattogram, Bangladesh, in 2004, the M.Sc. degree from the University of Ulsan, Republic of Korea, in 2010, and the Ph.D. degree from Saitama University, Japan, in 2014. In 2005, he joined as a Lecturer with the Department of Computer Science and Engineering, CUET, where he is currently a Professor. He is also a Visiting Scholar with Waseda University, Japan. He is the author of two books, one book chapter, and one patent. He has published more than 80 refereed journal articles and conference proceedings. His current research interests include multimedia security, digital watermarking, steganography, multimedia data compression, sound synthesis, digital image processing, and digital signal processing. He is a member of the technical committee of several international conferences. He was a recipient of the *Journal of Signal Processing* Best Paper Award, in 2015. He received the Best Paper Award from the International Conference on Intelligent Computing (ICIC), in 2009, the Excellent Paper Award from the Engineering and Arts Society Conference in South Korea (EASKO), in 2009, and the Best Paper Award from the Ninth International Forum on Strategic Technologies (IFOST), in 2014. He serves as a reviewer for various reputed journals, including IEEE, IEICE, Elsevier, and Springer.

TETSUYA SHIMAMURA (Senior Member, IEEE) received the B.E., M.E., and Ph.D. degrees in electrical engineering from Keio University, Yokohama, Japan, in 1986, 1988, and 1991, respectively. In 1991, he joined Saitama University, Saitama, Japan, where he is currently a Professor. He was a Visiting Researcher with Loughborough University, U.K., in 1995, and Queen's University Belfast, U.K., in 1996. He is the author and coauthor of six books. His current research interests include digital signal processing and its application to speech, image, and communication systems. He serves as an editorial member of several international journals and a member of the organizing and program committees of various international conferences.

· · ·