Send Orders for Reprints to reprints@benthamscience.net

Recent Advances in Computer Science and Communications, XXXX, *XX,* XXXXXXXXXX

RESEARCH ARTICLE

Malignancy Detection in Lung and Colon Histopathology Images by Transfer Learning with Class Selective Image Processing

Vidyullatha Sukhavasi[1], Shridhar Kulkarni[2], Raghavendran V.[3], Dastagiraiah C.[4], Shraban Kumar Apat[5] and Pundru Chandra Shaker Reddy[6,*]

[1]Department of Computer Science and Engineering, BVRIT HYDERABAD College of Engineering for Women, Hyderabad, India; [2]Department of Data Sciences, Harrisburg University of Science & Technology, Harrisburg, Pennsylvania, United States of America; [3]Information Technology, Vel's Institute of Science, Technology and Advanced Studies (VISTAS), Chennai, Tamil Nadu, India; [4]Computer Science and Engineering, Anurag University, Hyderabad, Telangana, India; [5]CSE (Cyber Security), Geethanjali College of Engineering and Technology, Hyderabad, Telangana, India; [6]School of Computer Science and Artificial Intelligence, SR University, Warangal, Telangana, India

ARTICLE HISTORY

Received: July 15, 2024
Revised: August 28, 2024
Accepted: September 13, 2024

DOI:
10.2174/0126662558335817241014113154

Abstract: ***Aims & Background:*** Due to its ferocity, enormous metastatic potential, and variability, cancer is responsible for a disproportionately high number of deaths. Cancers of the lung and colon are two of the most common forms of the disease in both sexes worldwide. The excellence of treatment and the endurance rate for cancer patients can be greatly improved with early and precise diagnosis.

Objectives & Methodology: We suggest a computationally efficient and highly accurate strategy for the rapid and precise diagnosis of lung and colon cancers as a substitute for the standard approaches now in use. The training and validation procedures in this work made use of an enormous dataset consisting of lung and colon histopathology pictures. There are 25,000 Histopathological Images (HIs) in the dataset, split evenly among 5 categories (mostly lung and colon tissues). Before training it on the dataset, a pretrained neural network (AlexNet) had its four layers fine-tuned.

Results: The study enhances malignancy detection in lung and colon histopathology images by applying transfer learning with class-selective image processing. Instead of enhancing the entire dataset, a targeted contrast enrichment was applied to images from the underperforming class, improving the model's accuracy from 92.3% to 99.2% while reducing computational overhead.

Conclusion: This approach stands out by emphasizing class-specific enhancements, leading to significant performance gains. The results meet or exceed established CAD metrics for breast cancer histological images, demonstrating the method's efficiency and effectiveness.

Keywords: Colon and lung cancer, convolutional-neural-networks, histopathology, image processing, transfer learning, multi-classification.

1. INTRODUCTION

Women are disproportionately affected by cancer, and breast cancer (breast carcinoma) is the deadliest form of the disease, tied with lung cancer in the mortality rate. Breast cancer is highly treatable if caught early, therefore, early identification is essential for lowering mortality rates. Without proper therapy in its early stages, cancer progresses from a benign to a malignant form. Histological assessment of biopsies is frequently used for breast cancer detection [1]. Images of breast histopathology are analysed by a seasoned pathologist at varying magnifications. If the sample tissue is suspicious, other imaging techniques like mammography may be necessary to confirm the diagnosis [2].

Cancer is the second top cause of demise in the globe. In 2021, about 19 M people were diagnosed with cancer, and nearly 10 M individuals perished from the cancer [3]. There are billions of cells in the human body, and these cells split and increase to produce new cells when the body needs them. When cells influence a specific age of damage, it is natural for them to perish and be interchanged by fresh ones. When this doesn't happen, damaged cells proliferate and eventually form tumours. Malignant tumours are cancerous tumours, but benign tumours are also possible [4]. Colon,-lungs-liver-breasts-rectum-brain-prostate-stomach-skin are

*Address correspondence to this at the at the School of Computer Science and Artificial Intelligence, SR University, Warangal, Telangana, India; E-mail: chandu.pundru@gmail.com

the utmost commonly affected organs, however, any organ in the human body is potentially at risk. Lung and colon cancers are the prominent cancer killers of both men and women. There will be 4.13 M new cases of lung & colorectal cancer in 2020, resulting in 2.7 million fatalities [5].

Tobacco and alcohol use, as well as a high body mass index, are just a few of the risk factors that have been linked to cancer. Physical carcinogens include substances like contamination and UV-rays, as well as some biological and genetic factors. Discomfort, tiredness, nausea, a persistent cough, shortness of breath, a decrease in appetite, a loss of muscle tone and weight, pains in the joints, and bruising and bleeding are all indicators of cancer [6]. While they may be experienced by some cancer patients, they are not unique to the disease. This means that a battery of diagnostic tests such as Biopsy-PET-CT-MRI scans and ultrasound is necessary for accurate cancer detection. Despite the fact that many individuals have no or little symptoms, by the time they manifest, it is typically too late to treat the condition [7]. The cancer-causing gene can be passed down from generation to generation in some families. Regular checks are especially important for people who have a family history of cancer. Many people are unable to purchase these screening systems because of how pricey they are. WHO estimates that over 70% of all cancer-related fatalities occur in low and middle-income countries [8]. These nations must invest heavily in the construction of well-equipped laboratories and diagnostic facilities, and in the education and preparation of medical personnel to perform diagnostic procedures. It is also important that those who are poor can afford the costs associated with these exams.

1.1. Appearances of Cancer

Breast cancer develops when cells in the breast proliferate uncontrollably. Infected cells gather together because they divide rapidly, unlike healthy cells. Lymph nodes are a potential early target for the cells' metastatic expansion. For the purpose of creating and testing algorithms that can detect breast cancer, the BreakHis dataset provides an extensive resource. It includes 7,909 RGB photos, which are divided into benign and malignant classes with further subtypes [9]. Each image has a resolution of 700x460 pixels. Forty, one hundred, two hundred, and four hundred times magnification images are accessible. The dataset is organized in directories by magnification, class, and subtype; it contains 5,429 cancerous photos and 2,480 benign ones. Images are available at magnification factors of 40x, 100x, 200x, and 400x. These are depicted in Fig. (**1**) for understanding. Machine learning models for breast cancer diagnosis can be informed by this benchmark in medical image analysis research.

1.2. Histopathological Examination

Tissue samples taken from a patient are studied in great detail by pathologists using histological biopsy techniques. In order to make an accurate diagnosis, a magnification of at least 40x is necessary. Fig. (**2**) displays samples taken from the areas of concern that will need to be exaggerated 100x-200x-400x before a pathologist can provide an accurate assessment of cell shape [10, 11].

When it comes to histopathology images of the colon and lungs, current methods for cancer detection frequently use dataset-wide global image enhancement techniques. However, these methods can boost accuracy in the long run, but they are computationally expensive and might not solve class-specific problems, especially in poor-performing classes. Additionally, these approaches perform suboptimally in some areas since they can't adapt to diverse classes' image quality [12]. These restrictions call attention to the necessity for a more focused and effective method that can boost picture quality only where it is required, increasing precision without sacrificing computing performance.

A discipline unrelated to healthcare and medicine may hold the key to solving this issue. Detecting diseases and creating intelligent algorithms that, depending on a patient's symptoms, recommend conventional drugs are only two examples of the many applications of Machine Learning (ML) in the study of pathology [13]. Several types of biomedical data have been classified and predicted using machine learning methods. Machines can now perform high-dimensional data analyses using deep learning (DL) methods on images, multi-dimensional anatomical representations, and video. DL is a subfield of ML that studies how the brain works algorithmically [14].

In clinical settings, a reliable diagnosis relies on the correct classification of HCs. To replace the time-consuming and resource-intensive labour of human specialists while still satisfying requirements for high accuracy and big data sets, this task could be automated with the help of ML algorithms, in particular, DL. TL is utilized in visual labelling to tackle cross-domain learning challenges by shifting beneficial gen from the source dataset to the task sphere [15]. Acquaintance from disparate but connected springs is communicated through TL to help students excel in their desired fields. As a result, a lot less information is needed to build target learners. Since TL has so many potential uses, it has quickly become one of the most popular subfields of ML [16]. In this research, we use the AlexNet framework, which has already been pre-trained. Limited layers of the strategy are changed while the weights in the conveyed layers are held constant in order to apply AlexNet's intricately learned features to a new picture classification assignment.

Here's how the remainder of the work is planned: In the second sector, we offer a brief impression of related studies that have been conducted in the past. The dataset that was used and its contents are outlined in Section 3. Section 4 emphasizes the development of the presented research model, whereas Section 5 describes the experiments and discusses the findings. The article then wraps up with a brief summary of the findings.

2. RELATED WORKS

The furthermost and most deadly kind of cancer in India is lung cancer. The 5-year endurance rate for admitters with non-small cell lung cancer is 18%. 13.5% of all instances of LC and 21.3% of all deaths from LC, according to the 2018 GLOBOCON report (1). Survival rates for people with pulmonary cancer are improved *via* early diagnosis. This section addresses the available techniques of cancer and detecting in histopathology images.

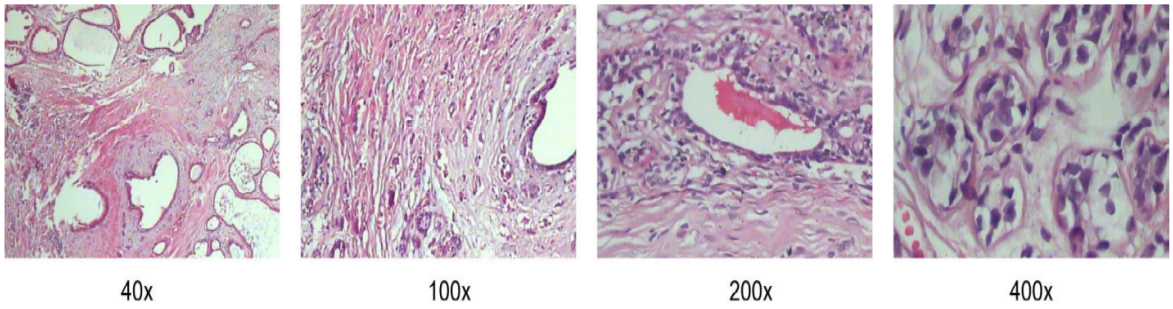

Fig. (1). Samples of breast cancer in HC in diverse magnifications. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

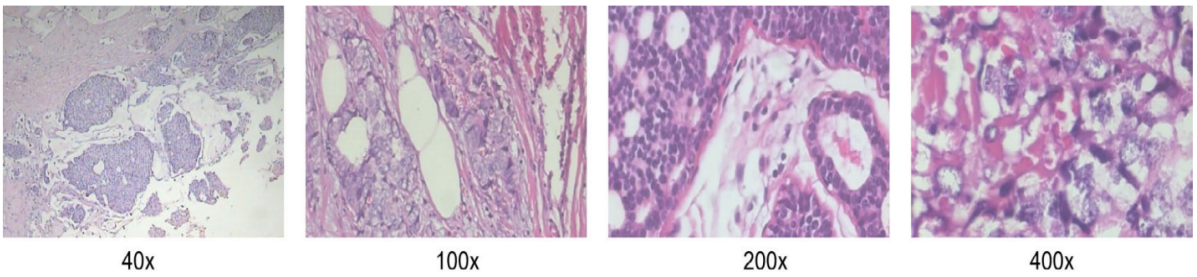

Fig. (2). Samples of malignant types of BC in HC in dissimilar magnifications. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

To identify the tumor's basis in HCs of colon cancer, Singhal A *et al*. [17] presented a spatially inhibited neural network. An innovative contiguous group analyst was applied to catalog cell nuclei. A maximum accuracy of 97.1% could be achieved. Their model performed well, but it seized an unacceptable 50 minutes to ride only one slide, therefore, its computational efficiency was subpar. For the purpose of determining whether or not lung nodules are cancerous, Syal J S *et al*. [18] created an ML design. A pooling strategy has been utilised to yield diverse areas of curiosity, from feature pictures and max-pooling performed in numerous epochs to cutting features from CT scans. The CT scan pictures used in this study are unique since no segmentation or feature extraction techniques were applied to them. They found that 87.14 percent of lung nodules could be accounted for using just their machine learning approach. CNN led to an 89% improvement in accuracy. In order to diagnose lung cancer, Mehmood *et al*. [1] used a recurrent-neural-network in combination with the damped-least-squares approach and the SOM technique, and they were successful 97.9% of the time.

As part of the preparation procedure for CT scans of lung nodules, M Dabass *et al*. [19] used image separation. Using CNN for classification and indexing basic toxic weights for identifying cancerous and benign patterns, the authors achieved a 92.6% success rate. To automate the process of finding polyps in colonoscopy recordings, Gadupudi A *et al*. [20] introduced a CNN-dependent strategy built on the AlexNet framework. Several preprocessing processes were carried out before CNN was used for classification. The accuracy of their model improved to 92.3%. The FP per edge outcome indicates that this approach has room for development. If the polyp's border is the same colour as the background, it may not be detected. A significant quantity of non-polyp entrants may also be generated in an informative frame. Because of this, the execution time will increase, and there is a greater possibility that the results will be inaccurate. The used strategy was able to achieve a 92.8% degree of precision.

Kumar A *et al*. [21] introduced the DFCNet framework, which is related to CNN, and achieved an accuracy of 84.5%. The research is faulty because it uses many datasets with varying scan settings, which increases the possibility of false-positive results in the case of malignant tumours. Classification accuracy can be improved by using the same scan settings across all datasets. Noaman N F *et al*. [22] trained a quicker region-dependent Faster R-CNN to detect polyps in colonoscopy movies using four separate datasets, with an average accuracy of 98.5%. For the purpose of finding cysts in expert-labelled images taken from over 3500 colonoscopies, Salvi M *et al*. [23] created and deployed CNNs. The accuracy of their model was 96.4%. The model proposed classifies colonoscopy images with 90.28 percent accuracy utilizing CNNs with binarized weights to minimise the network size. Wolf heuristic traits with the fewest occurrences were chosen to make things simpler and less cluttered. The average and abnormal lung edifices were identified with a precision of 97.47% using a combined learning CNN tuned for AdaBoost.

To classify CT scans of lung lesions into one of three categories, Khan T A *et al*. [24] devised an 8 layer CNN design. With the help of segmentation, specialists were able to manually extract nodule regions of interest from the photos. The dataset was also augmented with the help of generative adversarial networks. Their proposed model had a high rate of correct classification (93.9%). Greenspan *et al*. [25]

suggested a lightweight DL method for lung nodule detection using a CNN design with only 4-layers. Each layer is made of 2-subsequent blocks, a linking convolutional block. Since their suggested approach uses rarer flops and factors than current traditional CNN, the authors think it's ideally suited for real-time CT image interpretation. An accuracy of 97.9% was achieved by the proposed model. Preprocessing of CT scans for lung cancer was done by Razzak M *et al*. [26] using a technique that maintains brightness on many levels while removing noise. The damaged areas were segmented and characteristics were extracted using a refined neural network. When classifying features, an ensemble classifier was employed. The model's accuracy in classifying data was 96.2%.

Existing methods for detecting malignancy in lung and colon histopathology pictures frequently rely on uniform global enhancement techniques. These strategies fail to take into account the fact that the image quality varies across different classes [27]. As a consequence, this leads to poor performance for classes that are not performing as well as an increase in computational complexity, which in turn makes models less scalable and economical. Furthermore, these methods do not have any adaptive strategies for class-specific enhancements, which results in a compromise between the correctness of the results and the computational efficiency of the results [28]. The existence of these gaps highlights the necessity of developing methods that are more focused and efficient, with the capacity to selectively improve image quality where it is required, hence improving overall model performance without compromising the ability to scale.

2.1. Restrictions and Gaps in Existing Works

There are certain problems with the existing research.

- Algorithm inefficiency, which leads to higher computational costs and greater time consumption.
- False-positive results due to several datasets and/or different scan settings [29].
- Focused on fewer classes or a more limited dataset. Inaccuracy has been reported.

Key findings from this paper are as follows:

- While most previous cancer finding studies have concentrated on a solitary type of cancer, we were able to use our algorithm to distinguish lung & colon cancer simultaneously in this study.
- While some research has used image processing to enhance images before categorization, this has typically been done across the board, which increases processing time and expense. In order to address this issue, we put out a novel approach called Class Selective Image Processing (CSIP). Photos of a specific class are enhanced rather than all images in a dataset being preprocessed. This method reduces computational effort and time spent on categorization while simultaneously increasing the model's accuracy [30].
- Using CSIP on histopathology slide pictures, which boosted contrast by histogram equalisation, upgraded learning of the pretrained DL approach.
- Automatic identification of lung & colon cancer utilizing histopathology glide images: presented state-of-the-art results. The study used a large and evenly distributed dataset of 25000 HIs split into five distinct lung and colon categories. The proposed strategy improved the model's accuracy and reliability by including many classes and a huge number of photos.

3. DATASET DESCRIPTION

As our data source, we used the recently published LC25000 dataset [31] by A. Borkowski and colleagues. Twenty-five thousand pictures of lung and colon tissues are here, organised into five groups. Lung tissue pictures can be classified as benign, squamous cell carcinoma, or adenocarcinoma. Adenocarcinoma and benign tissues both appear in colon imaging [32]. Approximately 1250 photos from pathology glides containing lung & colon tissues made up the bulk of the LC25000 dataset. The dataset was enhanced by spinning and gyrating the photographs under different lighting circumstances, bringing the total number of images to 25,000. This allowed the dataset to be split into five categories, each containing 5,000 images. Before the enhancement was applied, the images were shrunk to 768 pixels on each side. These pictures have been checked for accuracy, and they all comply with the HIPAA standards for patient confidentiality and freedom of usage (HIPAA). Fig. (**3**) shows some examples of the data set's photos. HIs of lung and colon tissues provide crucial insights into the nature of various lesions. Lung adenocarcinoma, a type of non-small cell lung cancer, shows irregular glandular structures and cellular atypia, while lung squamous cell carcinoma, another non-small cell lung cancer, features keratinization and intercellular bridges [33]. Benign lung tissues, such as hamartomas, display normal cellular structures without significant atypia. Similarly, colon adenocarcinoma, the most common colon cancer, reveals glandular formations with cellular atypia and nuclear pleomorphism, whereas benign colon tissues, like hyperplastic polyps, exhibit well-organized glands with uniform cells. These detailed images are essential for differentiating between benign and malignant conditions, ensuring accurate diagnoses, and guiding effective treatment strategies.

4. EXPERIMENTAL DESIGN

Transfer learning is a type of ML that involves using a technique developed for one job to complete a different one. In DL, pre-trained networks from a gigantic dataset are utilised to create a new grid design, which is then fine-tuned for usage on a different dataset. The TL approach drastically cuts down on the time and energy needed to train deep CNNs. It's often far quicker and less complicated than starting from scratch with a network and training it. When conducting TL, DL models that have been trained on the massive ImageNet competition's [34] image classification problem are commonly used.

In this research, cancer picture categorization is handled by the AlexNet model, which has been pre-trained. AlexNet, created by Alex Khrizveski in 2012, is a CNN often lauded for its superior exactness on difficult datasets. There are almost 60 million parameters in this network, which consists

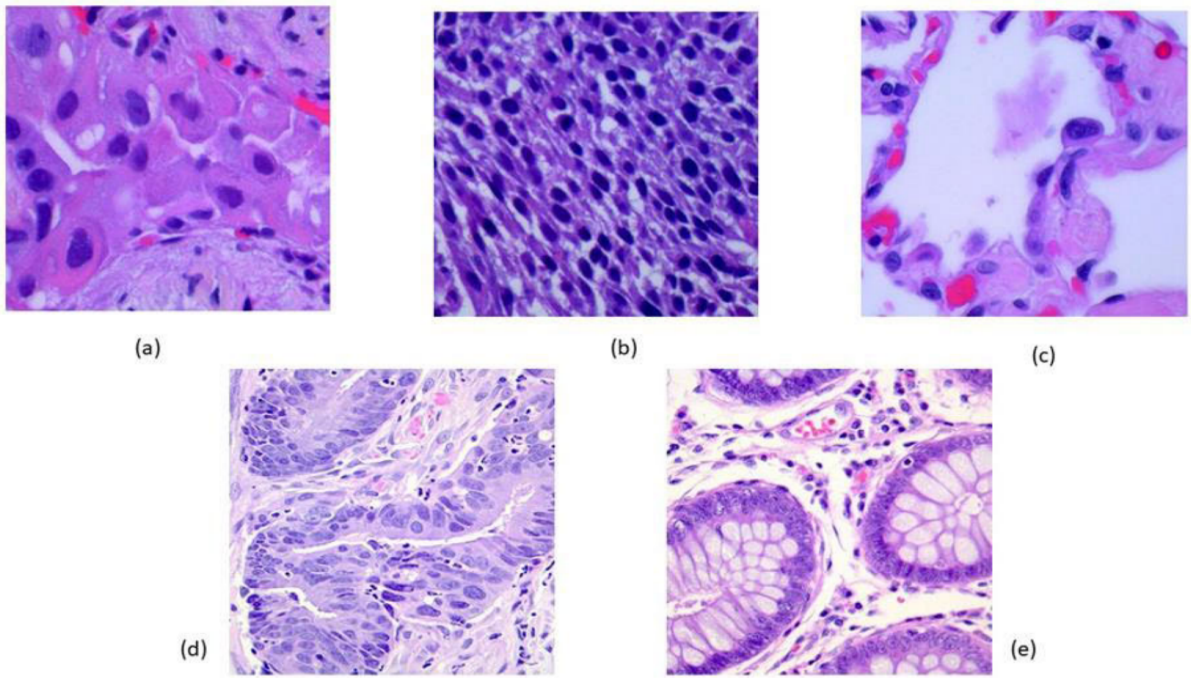

Fig. (3). HIs from dataset). (**a**) Lung-Adenocarcinoma (**b**) Lung-benig (**c**) Lung-squamous (**d**) Colon-adenocarcinoma). (**e**) Colon-benign. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

of 65,000 neurons. The ImageNet database is a massive image repository that contains over 14 million photos in over 20,000 categories. AlexNet was trained with over a million photos from over a thousand different classes found in the ImageNet database. 5-convolutional layers and three fully linked layers make up the scheme's architecture [35]. Tuning the model for the needs of this study necessitates making adjustments to the model's initial layer and its final three levels. Images were scaled down to 227×227×3 pixels to fit the model's specifications. Eighty percent of the total photos were set aside for training, while the remaining twenty percent were used for validation [36].

Illustration of the suggested procedure is shown in Fig. (**4**). To begin, the HIs are obtained and resized to 227×227×3 to meet the requirements of the model. The next step is to give the model scaled photos for training and validation. Accuracy, precision, F1-score, recall, specificity, & misclassification-rate (Metrics) are used to assess the preliminary results. Unsatisfactory outcomes prompt the implementation of the CSIP technique. The proposed CSIP method finds the poorly performing class and enhances its visuals with Histogram Equalization (HE). Images are updated, and the model is run again. Instead of processing the full dataset, the suggested method just processes the photos of the underperforming classes, saving both time and energy [37].

One of the most useful methods for boosting image quality is by increasing the contrast in the image. Because of its effectiveness and processing speed, histogram equalisation is used for the chosen images for contrast heightening. Redistributing the concentration values, histogram equalisation improves disparity. The dataset was rerun for training & test by swapping out all photos from the floundering class with histogram-matched versions of the originals [38].

The incorporation of Class-Selective Image Processing (CSIP) into the process of malignancy identification improves the overall accuracy and efficiency of the model by concentrating on the particular requirements of classes that are not performing as expected. CSIP selectively enhances image quality only for those classes where the initial classification accuracy was lower [39]. This is in contrast to the traditional approach of using global enhancement techniques throughout the whole dataset. Through the utilisation of this targeted strategy, not only is the accuracy improved from 92.3% to 99.2%, but the computing cost is also decreased, resulting in the model becoming more scalable and efficient. A more robust and precise detection process is ensured by CSIP, which eventually leads to improved performance in clinical applications. This is accomplished by tackling issues that are specific to categories.

5. RESULTS AND DISCUSSION

Our proposed methodology showed encouraging findings when compared to previous research on the recognition of lung & colon cancer. Our dataset consists of 25.000 photos, evenly split between 5 categories. To train the AlexNet model, we used 80% of the images in each class, while 20% were used for testing. We isolated the dataset so that the total number of images was not reduced below 20,000. Performance Metrics were all taken into account in adding to the exactness of the model assessment. The following are the key performance indicators that were measured in this research.

$$\text{Specificity} = \frac{TN}{TN+FP} \quad (1)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (2)$$

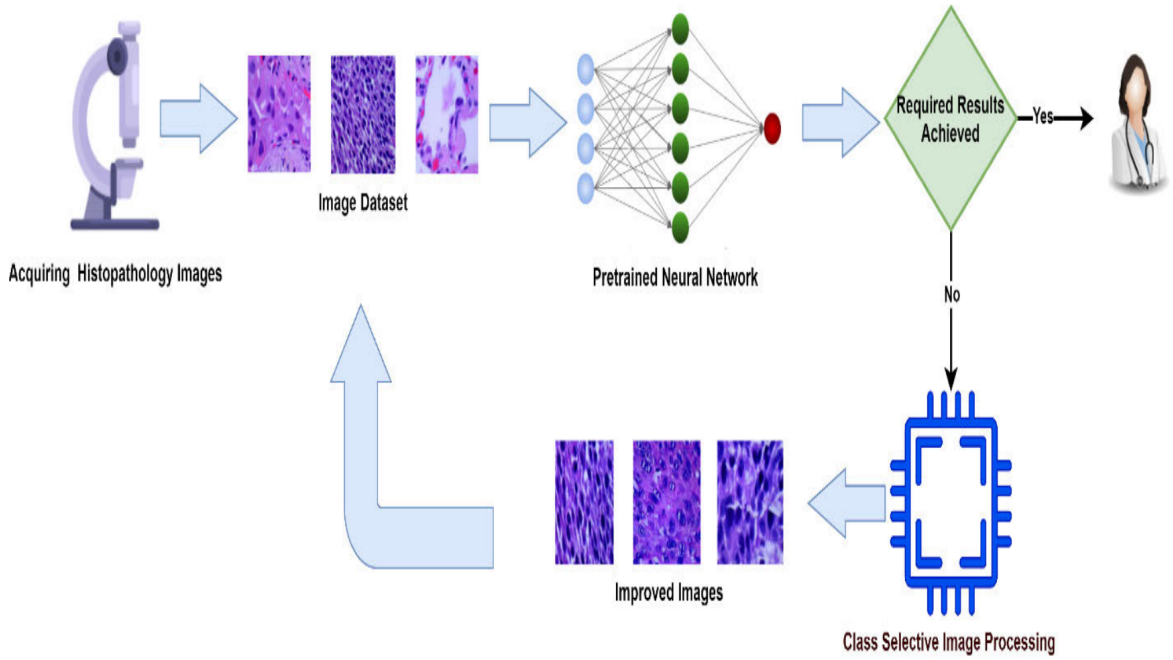

Fig. (4). Proposed strategy for TL and class choosy image processing. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

$$F1 = \frac{TP}{TP+\frac{1}{2}(FP+FN)} \quad (3)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (4)$$

$$\text{Accuracy} = \frac{TP+TN}{TN+TP+FN+FP} \quad (5)$$

$$\text{Misclassification-Rate} = 1\text{-(Accuracy)} \quad (6)$$

Figs. (**5** and **6**) show the model's enactment prior to the implementation of CSIP through visualisations of the training progress and confusion matrix, respectively. Overall accuracy, as well as precision and recall scores for each class, are all displayed in the confusion matrix. The model has an overall accuracy of 89.9%, with individual subsets achieving precisions of 98.8%, 99.7%, 67.4%, 99.9%, and 96% for colon adenocarcinoma, colon normal, lung-adenocarcinoma, and lung-squamous-cell-carcinoma, respectively. The model performance metrics for each class in the dataset are displayed in Fig. (**5**). Lung n class achieves 99.94% accuracy with a 0.006% misclassification rate, which is the best of any class. The colon n classification achieved a 99.92% accuracy with a 0.0008 misclassification rate. In tests, colon aca has shown a 99.58% success rate with a 0.0042 percent misclassification rate. The lung_scc and lung_aca classes underperformed when compared to the aforementioned ones, with an exactness of 90.22 & 89.84% and misclassification rates of 0.978 & 0.1016, respectively. Classes lung_n, colon_n, and colon_aca all got perfect 100 on the F1-score, however, classes lung_aca and lung_scc only got 79.56 & 68.55, respectively. To conserve time and computational resources while also increasing the model's overall classification accuracy, we only used the contrast enhancement approach (HE) to photos from classes that were underperforming. All 5,000 photos in the lung_scc class were processed using the HE method. The histogram is made straighter with the HE method by spreading out all of the grey values equally. The average luminance of the improved image, therefore. shifts significantly. The HE can be broken down into its basic steps, which are as follows. Let's pretend that the input copy S(i, j) has a total of Y-pixels with a grayscale range of [Z0, ZG1], where G is the number of grayscale levels. The probability-density-function F(Zk) of the intensity level Zk in an image can be computed by using the following expression:

$$F(Z_k) = \frac{n_k}{Y} \text{ for } 0 \le k \le G\text{-}1 \quad (7)$$

Where n_k signifies the addition of pixels and Z_k embodies the intensity level. The histogram of picture S is the scheme of Zk *vs* nk while the cumulative-density-function (CDF) is given by:

$$CDF(Z_k) = \sum_{i=o}^{k} F(Z_i) \quad (8)$$

The germane image is anticipated into the whole energetic range ∈ [Z0, ZG−1] by HE utilizing the CDF, which is symbolized by the succeeding equation:

$$C(Z) = Z_0 + (Z_{G-1}, Z_0)\left(\sum_{i=0}^{k} F(Z_i)\,(Z)\right) \quad (9)$$

If the scan has been handled by the classic, HE, the statistical expectation E(.) of the resultant scan S is computed using:

$$E(S) = S_m = \frac{1}{2}(Z_{G-1}, Z_0) \quad (10)$$

The average brightness of the generated image is denoted by the symbol Sm. As a result, there is a noticeable tone shift toward the midpoint of the greyscale. Fig. (**6**) compares the histograms of the original image to the histograms of the histogram-equalized image to demonstrate the effects of the applied image processing. The confusion matrix shows that AlexNet achieved an overall accuracy of 89.8%, with very high precision across all classes of cancer scans with the

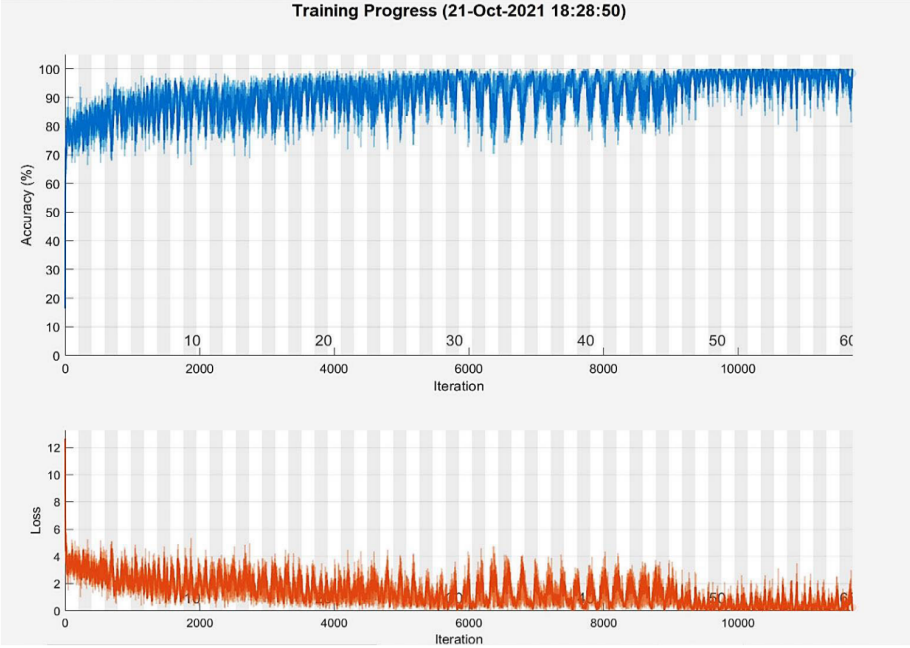

Fig. (5). Training growth of AlexNet before application of CSIP. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

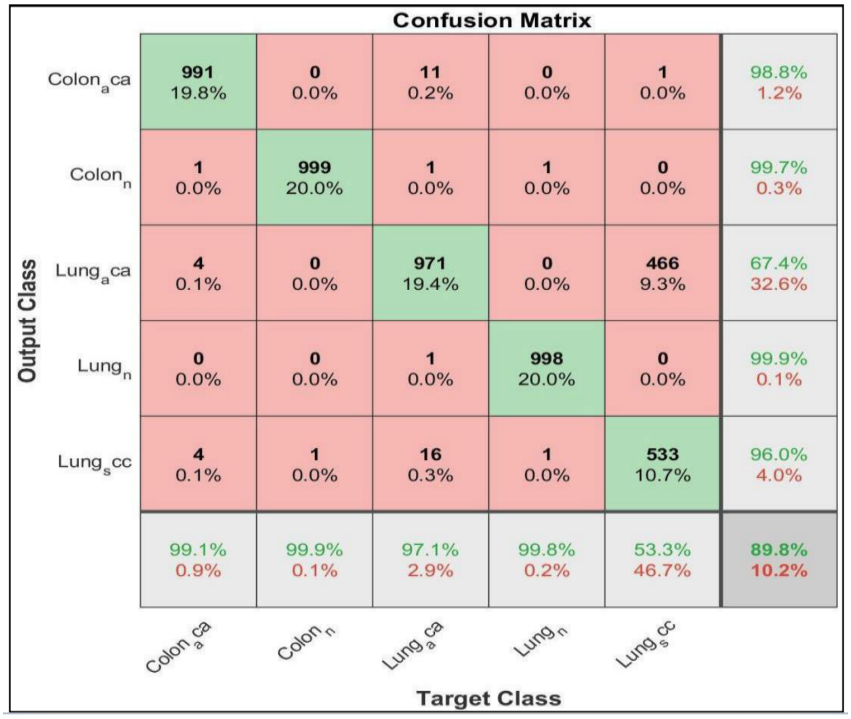

Fig. (6). Confusion-matrix of AlexNet beforehand application of CSIP. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

exception of the lung scc class. The Colon aca category achieved a 99.58% accuracy rate, a 98.80% precision rate, and an F-1 score of 98.95%. Colon n class achieved an accuracy of 99.92%, a precision of 99.70%, and an F-1 score of 99.80%. With a 90.02% accuracy, 67.38% precision, and 79.56% F-1 score, the Lung aca class performed admirably. In the case of Lung_n, we achieved 99.94% accuracy, 99.90% precision, and 99.85% F-1 score. The F-1 score for the Lung scc class is just 68.55 percent, however, the accuracy and precision are both 90.22%. This means that lung scc images require processing to enhance image quality.

All 5,000 photos in the lung scc class have been processed using HE. The results of applying HE to some sample photos are shown in Fig. (**6**). The contrast gradient is clearly obvious

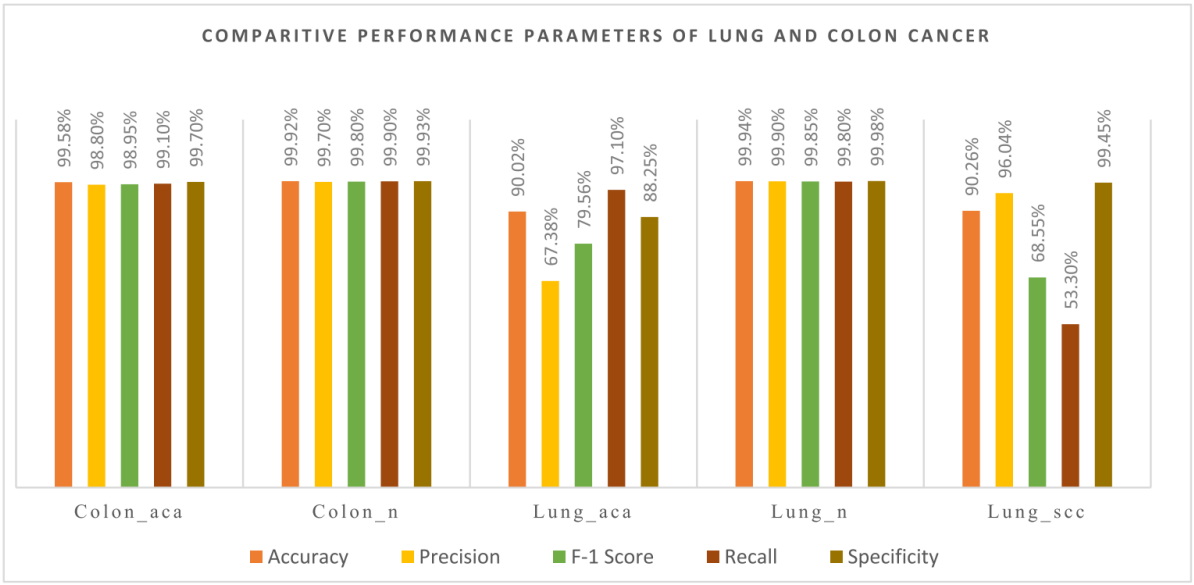

Fig. (7). Each class' performance in terms of metrics. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

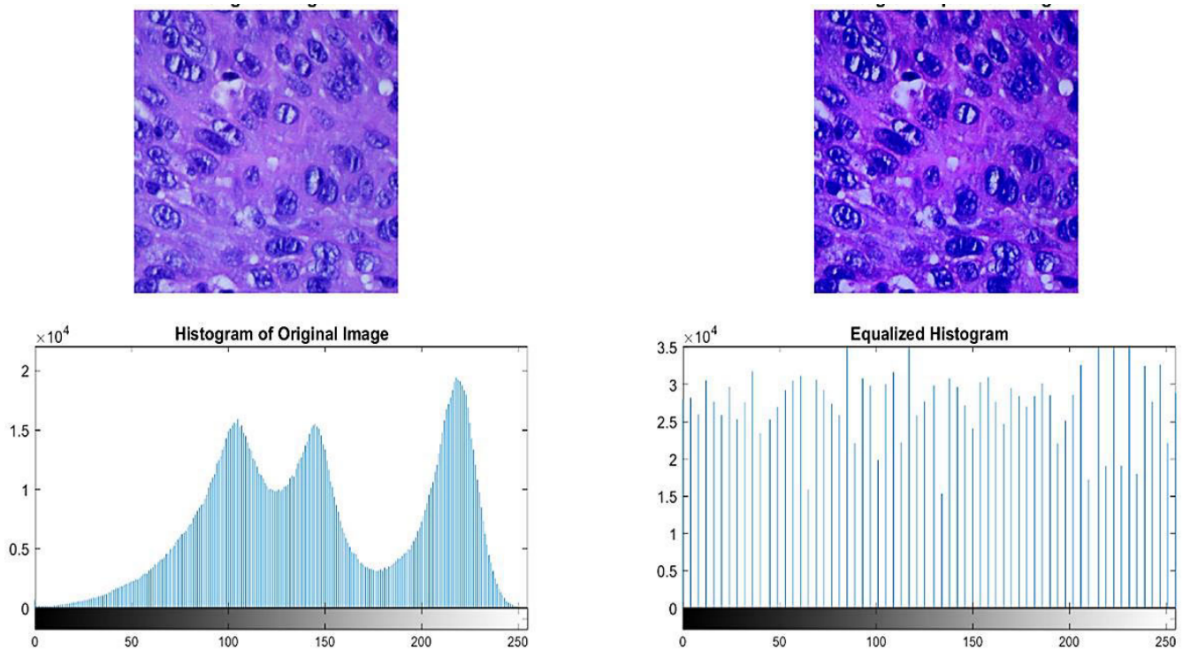

Fig. (8). An illustration and its histogram beforehand & after histogram equality. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

in the examples provided. The design is a ride for a further 60 epochs when the dataset is updated and the image quality of the chosen class is optimised. The accurateness and other factors of the investigational results improved dramatically once CSIP was implemented. The training curve is exposed in Fig. (**7**), and the confusion-matrix following CSIP is revealed in Fig. (**8**). After implementing the recommended CSIP, exactness for the Lung_scc class touched 100%. The exactness and other metrics of all 5 image classes in the dataset are compared and contrasted in Fig. (**9**). The suggested model has been tested extensively on test data to evaluate how it would fare when applied to data that has not yet been uncovered. The same care was used in processing and organising the test data to guarantee a fair assessment. While the model's performance is great overall, it really shines when it comes to detecting lung squamous cell carcinoma, where it achieves perfect results across every measure of performance metrics. For lung-benign, lung-adenocarcinomas, colon-benign, and colon-adenocarcinomas, the model achieves an accuracy of 99.94%, 99.64%, 99.12%, and 98.98%, respectively.

A significant degree of sensitivity and specificity was also shown by all of the groups. Improved model functionality

can be attributed in large part to the high-quality features that were gleaned through employing the CSIP method. Classification accuracy and misclassification rates are compared in Table **1** for the model both before and after the use of CSIP. To prove the efficacy of the proposed model, we compare our findings to those obtained through state-of-the-art methods. However, these results are mainly incomparable since we used an inimitable dataset that is significantly different from the ones utilized in the aforementioned papers. Nonetheless, they've been compared as they both aim to achieve the same thing.

It is abundantly obvious that the method is beneficial in improving both accuracy and computing efficiency, as demonstrated by the results obtained both before and after the use of Class-Selective Image Processing (CSIP). Nevertheless, a more in-depth examination on the reasons why particular classes initially underperformed could provide deeper insights into the specific difficulties that CSIP is attempting to address [40]. There are a number of potential factors that could have contributed to these inconsistencies, including differences in image quality, class imbalance, or distinctive morphological traits. As a result of selecting and increasing the image quality of certain underperforming classes, CSIP not only increases accuracy but also greatly decreases computing overhead. As a result, it is an essential improvement in the process of detecting malignancy. Highlighting the enormous contribution that CSIP has made to the field by highlighting the combined benefits of greater performance and increased efficiency is a worthwhile endeavour (Figs. **8-12**).

Compared to conventional approaches, the suggested technique performs better when classifying lung and colon anomalies together. Table **2** shows that, [19], our method beats all other existing cancer detection techniques. It should be noted that the LC25000 dataset was used in the research [41-43]. The proposed approach had a higher accuracy and F-measure than another study [41] that focused primarily on colon samples. While the studies [42, 43] demonstrated relatively improved accuracy, they only classified lung or colon pictures. Based on what has been said thus far, it seems reasonable to accomplish that the offered technique is active in correctly recognising lung and colon cancer tissue.

Applying the Class-Selective Image Processing (CSIP) method to various datasets or other types of malignancies may pose hurdles despite the fact that it has shown substantial increases in accuracy and computing efficiency for histopathology pictures of the colon and lungs. Additional processing method customisation may be necessary to ensure that CSIP is effective for all cancer types, as its efficacy may depend on tumour type, tissue shape, and staining methods. Compatibility with present systems, keeping clinical throughput constant, and resolving the requirement for clinician training on the new procedure are just a few of the obstacles that could arise when CSIP is integrated into existing diagnostic workflows. In spite of these caveats, the results show that CSIP has the potential to be a useful tool for improving diagnostic accuracy in many settings, so long as researchers investigate how to make it more flexible and how to incorporate it into clinical practice in the future. Better results in cancer diagnosis and wider usage may result from this.

CONCLUSION

Lung and colon cancers are among the main reasons for death globally. Initial and precise diagnosis of these tumours

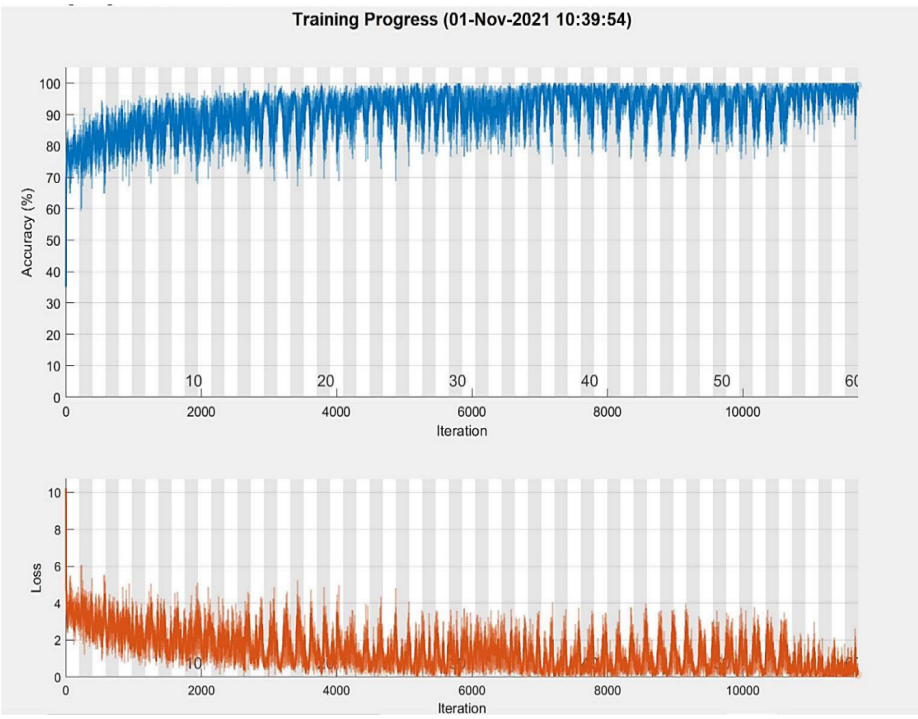

Fig. (9). Training growth of the technique after class choosy image-processing. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

Table 1. Whole performance of offered CSIP-TL approach.

-	Training		Test	
	Accuracy(%)	**Misclassification Rate (%)**	**Accuracy (%)**	**Misclassification Rate (%)**
Before CSIP	93.8	0.985875	92.3	0.040554313
Post CSIP	99.85	0.003640	99.2	0.00464

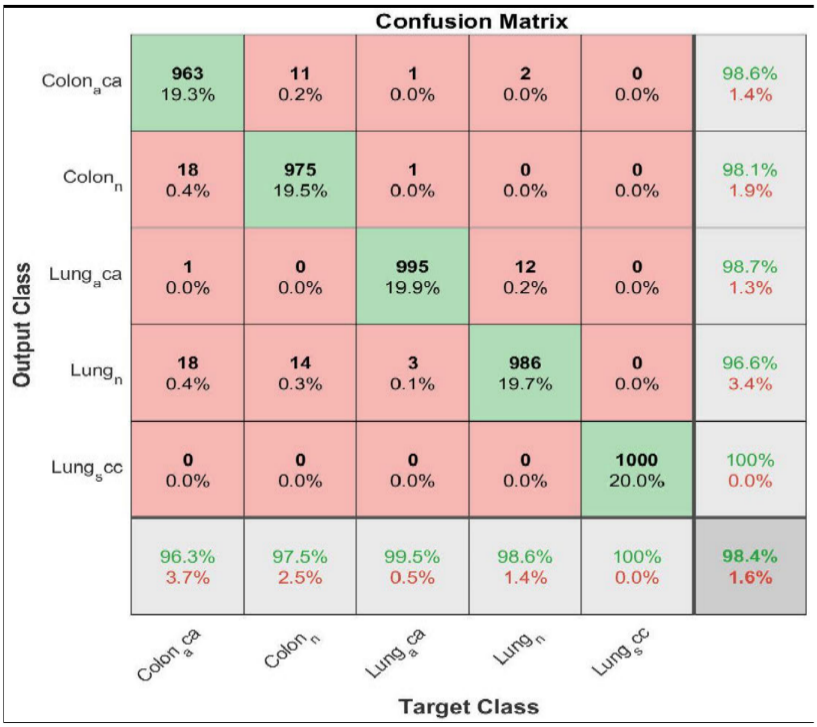

Fig. (10). CM of the approach after class choosy image-processing. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

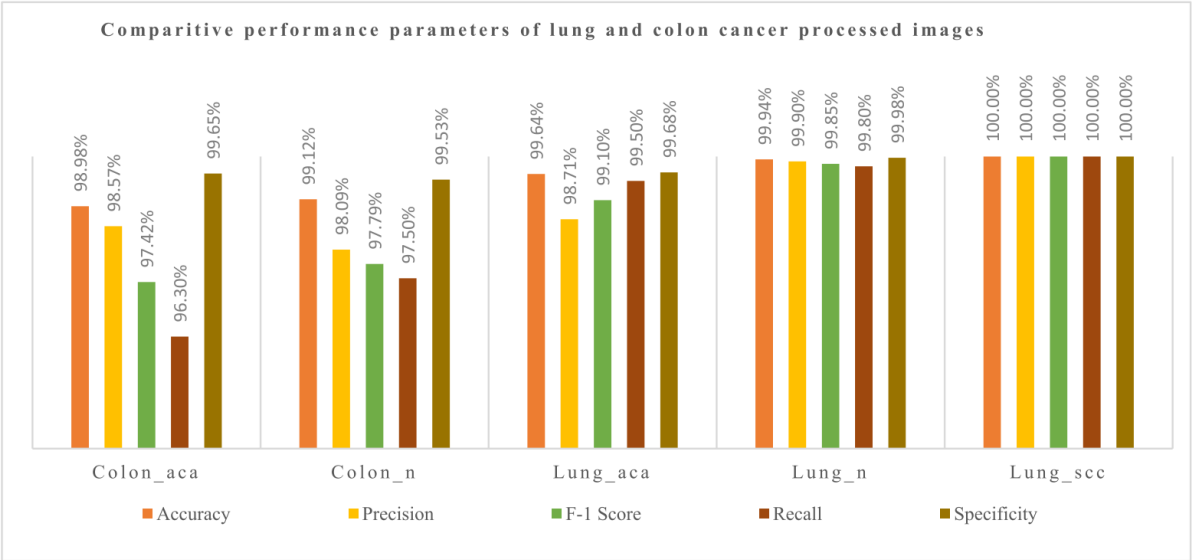

Fig. (11). Performance of the design at each class level after CSIP application. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

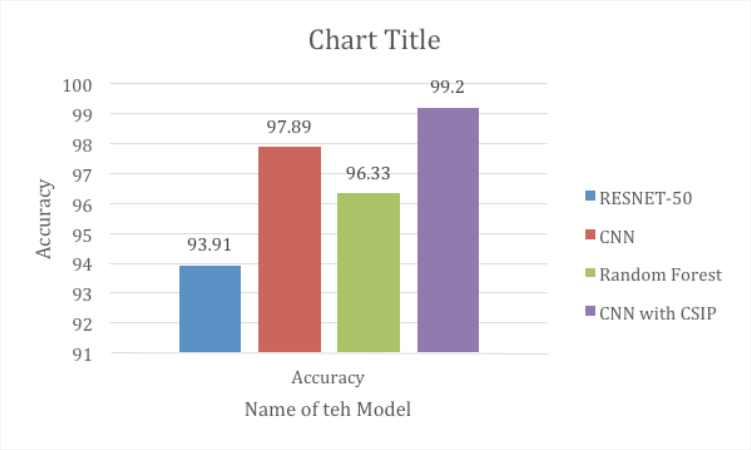

Fig. (12). Proposed model accuracy comparison with existing models. (*A higher resolution / colour version of this figure is available in the electronic copy of the article*).

Table 2. The Performance contrast in terms of accuracy of the proposed model with existing techniques.

Author Name	Type of Cancer	Type of Images	Model Name	Accuracy (%)
Bukhari *et al.* [41]	Colon Cancer	Histopathology Images	RESNET-50	93.91
Mangal S *et al.* [42]	Lung Cancer	Histopathology Images	Convolutional Neural Network	97.89
Masud M *et al.* [43]	Lung & Colon Cancer	Histopathology Images	Convolutional Neural Network	96.33
Proposed Model	Lung & Colon Cancer	Histopathology Images	CNN with CSIP	99.2

can dramatically enhance therapy consequences and endurance rates. Accurate and effective detection of lung and colon cancer was the focus of this study. This detection is accomplished by using transfer learning on a dataset of 250,000 HIs of lung & colon tissues. The initial overall accuracy of our deployed DL network was 92.3%; after implementing the recommended CSIP technique, this value increased to 99.2%. When compared to conventional methods, the proposed methodology for detecting lung & colon cancer has not only improved detection accuracy but also decreased detection time and computing cost. We think our offered method can potentially be used to diagnose other diseases more accurately.

The next step for CSIP needs to be to be fine-tuned for various tissue structures and staining methods so that it can be used in a wider variety of cancers and histopathological datasets. System compatibility and clinician training are two additional obstacles that must be addressed in order to integrate CSIP into clinical procedures. Better real-time applicability, more widespread adoption, and improved diagnostic accuracy in different clinical contexts could be the result of research into automated, adaptive improvement methods.

AUTHORS' CONTRIBUTION

The authors confirm their contribution to the paper as follows: study conception and design: VS; data analysis and interpretation of results: SK, RV, SKA, PR; draft manuscript: DC. All authors reviewed the results and approved the final version of the manuscript.

LIST OF ABBREVIATIONS

CSIP = Class-Selective Image Processing
DL = Deep Learning
HE = Histogram Equalization
ML = Machine Learning

CONSENT FOR PUBLICATION

Not applicable.

AVAILABILITY OF DATA AND MATERIALS

The data supporting the findings of the article is available within the article.

FUNDING

None.

CONFLICT OF INTEREST

The authors declare no conflict of interest, financial or otherwise.

ACKNOWLEDGEMENTS

Declared none.

REFERENCES

[1] S. Mehmood, T.M. Ghazal, M.A. Khan, M. Zubair, M.T. Naseem, T. Faiz, and M. Ahmad, "Malignancy detection in lung and colon histo-

pathology images using transfer learning with class selective image processing", *IEEE Access,* vol. 10, pp. 25657-25668, 2022. http://dx.doi.org/10.1109/ACCESS.2022.3150924

[2] A. Oubaalla, H. ElMoubtahij, and N. El Akkad, "Detection of Lung and Colon Cancer from Histopathological Images: Using Convolutional Networks and Transfer Learning", *Int. J. Comput. Digital Sys.,* vol. 15, no. 1, pp. 583-595, 2024. http://dx.doi.org/10.12785/ijcds/160144

[3] S. Suneel, A. Balaram, M. Amina Begum, K. Umapathy, P.C.S. Reddy, and V. Talasila, "Quantum mesh neural network model in precise image diagnosing", *Opt. Quantum Electron.,* vol. 56, no. 4, p. 559, 2024. http://dx.doi.org/10.1007/s11082-023-06245-y

[4] A.H. Uddin, Y.L. Chen, M.R. Akter, C.S. Ku, J. Yang, and L.Y. Por, "Colon and lung cancer classification from multi-modal images using resilient and efficient neural network architectures", *Heliyon,* vol. 10, no. 9, p. e30625, 2024. http://dx.doi.org/10.1016/j.heliyon.2024.e30625 PMID: 38742084

[5] İ. Karabulut, R. Selen, M. Yağanoğlu, and S. Özmen, "Recognition of Colon Polyps (Tubular Adenoma, Villous Adenoma) and Normal Colon Epithelium Histomorphology with Transfer Learning", *Eurasian J. Med.,* vol. 56, no. 1, pp. 35-41, 2024. http://dx.doi.org/10.5152/eurasianjmed.2024.23130 PMID: 39128055

[6] B. Sudhakar, P.A. Sikrant, M.L. Prasad, S.B. Latha, G.R. Kumar, S. Sarika, and P.C. Shaker Reddy, "Brain Tumor Image Prediction from MR Images Using CNN Based Deep Learning Networks", *J. Inf. Technol. Manage.,* vol. 16, no. 1, pp. 44-60, 2024.

[7] V. Rao, "Modernizing Cancer Diagnosis with an Intelligent Model for Lung and Colon Cancer", *J. Electrical Systems,* vol. 20, no. 6s, pp. 1218-1226, 2024. http://dx.doi.org/10.52783/jes.2851

[8] E.A. Alabdulqader, M. Umer, K. Alnowaiser, H. Wang, A.A. Alarfaj, and I. Ashraf, "Image Processing-based Resource-Efficient Transfer Learning Approach for Cancer Detection Employing Local Binary Pattern Features", *Mob. Netw. Appl.,* pp. 1-17, 2024. http://dx.doi.org/10.1007/s11036-024-02331-x

[9] M.L. Prasad, A. Kiran, and P.C. Shaker Reddy, "Chronic Kidney Disease Risk Prediction Using Machine Learning Techniques", *J. Inf. Technol. Manage.,* vol. 16, no. 1, pp. 118-134, 2024.

[10] S. Tummala, S. Kadry, A. Nadeem, H.T. Rauf, and N. Gul, "An explainable classification method based on complex scaling in histopathology images for lung and colon cancer", *Diagnostics (Basel),* vol. 13, no. 9, p. 1594, 2023. http://dx.doi.org/10.3390/diagnostics13091594 PMID: 37174985

[11] S. Sowjanya, I.S. Reddy, C. Muralikrishna, T.S.L. Prasad, P.C.S. Reddy, and V. Sharma, "Bioacoustics Signal Authentication for E-Medical Records Using Blockchain", *2024 International Conference on Knowledge Engineering and Communication Systems (ICKECS),* 2024 18-19 Apr, 2024, Chikkaballapur, India, 2024, pp. 1-6. http://dx.doi.org/10.1109/ICKECS61492.2024.10617376

[12] S. Iqbal, A.N. Qureshi, M. Alhussein, K. Aurangzeb, and S. Kadry, "A Novel Heteromorphous Convolutional Neural Network for Automated Assessment of Tumors in Colon and Lung Histopathology Images", *Biomimetics (Basel),* vol. 8, no. 4, p. 370, 2023. http://dx.doi.org/10.3390/biomimetics8040370 PMID: 37622975

[13] M.K. Shaik, D. Kalpana, U. Sesadri, S. Mukherjee, C. Dastagiraiah, and P.C.S. Reddy, "Brain Tumor Classification Using UNet Deep Neural Networks from 3D MRI Images", *2024 International Conference on Electronics, Computing, Communication and Control Technology (ICECCC)* 02-03 May 2024, Bengaluru, India, 2024, pp. 1-6. http://dx.doi.org/10.1109/ICECCC61767.2024.10593884

[14] I. Chhillar, and A. Singh, "A feature engineering-based machine learning technique to detect and classify lung and colon cancer from histopathological images", *Med. Biol. Eng. Comput.,* vol. 62, no. 3, pp. 913-924, 2023. http://dx.doi.org/10.1007/s11517-023-02984-y PMID: 38091162

[15] A. Rahman, A. Alqahtani, N. Aldhafferi, M.U. Nasir, M.F. Khan, M.A. Khan, and A. Mosavi, "Histopathologic oral cancer prediction using oral squamous cell carcinoma biopsy empowered with transfer learning", *Sensors (Basel),* vol. 22, no. 10, p. 3833, 2022. http://dx.doi.org/10.3390/s22103833 PMID: 35632242

[16] D. Maleki, S. Rahnamayan, and H.R. Tizhoosh, "A self-supervised framework for cross-modal search in histopathology archives using scale harmonization", *Sci. Rep.,* vol. 14, no. 1, p. 9724, 2024. http://dx.doi.org/10.1038/s41598-024-60256-7 PMID: 38678157

[17] A. Singhal, S. Varshney, T.A. Mohanaprakash, R. Jayavadivel, K. Deepti, P.C.S. Reddy, and M.B. Mulat, "Minimization of latency using multitask scheduling in industrial autonomous systems", *Wirel. Commun. Mob. Comput.,* vol. 2022, no. 1, pp. 1-10, 2022. http://dx.doi.org/10.1155/2022/1671829

[18] J.S. Syal, A. Jain, A.K. Dubey, and V. Jain, "Improving Lung and Colon Cancer Detection using Ensemble Method Approach", *2024 11th International Conference on Computing for Sustainable Global Development (INDIACom)* 28 Feb - 01 Mar, 2024, New Delhi, India, 2024, pp. 1767-1773. http://dx.doi.org/10.23919/INDIACom61295.2024.10498812

[19] M. Dabass, J. Dabass, S. Vashisth, and R. Vig, "A hybrid U-Net model with attention and advanced convolutional learning modules for simultaneous gland segmentation and cancer grade prediction in colorectal histopathological images", *Intell. Based Med.,* vol. 7, p. 100094, 2023. http://dx.doi.org/10.1016/j.ibmed.2023.100094

[20] A. Gadupudi, M.L. Prasad, S.K. Nadgaundi, P.C.S. Reddy, S. Sharma, and N. Sharma, "A deep learning framework for human disease prediction using microbiome data", *2024 International Conference on Integrated Circuits and Communication Systems (ICICACS)* 23-24 Feb, 2024, Raichur, India, 2024, pp. 1-6. http://dx.doi.org/10.1109/ICICACS60521.2024.10498711

[21] A. Kumar, A. Vishwakarma, and V. Bajaj, "ML3CNet: Non-local means-assisted automatic framework for lung cancer subtypes classification using histopathological images", *Comput. Methods Programs Biomed.,* vol. 251, p. 108207, 2024. http://dx.doi.org/10.1016/j.cmpb.2024.108207 PMID: 38723437

[22] N.F. Noaman, B.M. Kanber, A.A. Smadi, L. Jiao, and M.K. Alsmadi, "Advancing Oncology Diagnostics: AI-Enabled Early Detection of Lung Cancer Through Hybrid Histological Image Analysis", *IEEE Access,* vol. 12, pp. 64396-64415, 2024. http://dx.doi.org/10.1109/ACCESS.2024.3397040

[23] M. Salvi, U.R. Acharya, F. Molinari, and K.M. Meiburger, "The impact of pre- and post-image processing techniques on deep learning frameworks: A comprehensive review for digital pathology image analysis", *Comput. Biol. Med.,* vol. 128, p. 104129, 2021. http://dx.doi.org/10.1016/j.compbiomed.2020.104129 PMID: 33254082

[24] T.A. Khan, A. Fatima, T. Shahzad, Atta-Ur-Rahman, K. Alissa, T.M. Ghazal, M.M. Al-Sakhnini, S. Abbas, M.A. Khan, and A. Ahmed, "Secure IoMT for disease prediction empowered with transfer learning in healthcare 5.0, the concept and case study", *IEEE Access,* vol. 11, pp. 39418-39430, 2023. http://dx.doi.org/10.1109/ACCESS.2023.3266156

[25] H. Greenspan, B. van Ginneken, and R.M. Summers, "Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique", *IEEE Trans. Med. Imaging,* vol. 35, no. 5, pp. 1153-1159, 2016. http://dx.doi.org/10.1109/TMI.2016.2553401

[26] M.I. Razzak, S. Naz, and A. Zaib, "Deep learning for medical image processing: Overview, challenges and the future", *arXiv:1704.06825,* 2018.

[27] A. Nangunuri, A.T. Somnathe, B. Prasanthi, E.L. Mercy, S.V. Ramanan, and P.C.S. Reddy, "A Novel Meta-Learning Ensemble Framework for Cancer Classification Using Convolution Neural Networks", *2024 International Conference on Electronics, Computing, Communication and Control Technology (ICECCC)* 02-03 May, 2024, Bengaluru, India, 2024, pp. 1-6. http://dx.doi.org/10.1109/ICECCC61767.2024.10593843

[28] C.L. Srinidhi, O. Ciga, and A.L. Martel, "Deep neural network models for computational histopathology: A survey", *Med. Image Anal.,* vol. 67, p. 101813, 2021. http://dx.doi.org/10.1016/j.media.2020.101813 PMID: 33049577

[29] M. Dabass, S. Vashisth, and R. Vig, "A convolution neural network with multi-level convolutional and attention learning for classification of cancer grades and tissue structures in colon histopathological images", *Comput. Biol. Med.,* vol. 147, p. 105680, 2022. http://dx.doi.org/10.1016/j.compbiomed.2022.105680 PMID: 35671654

[30] A.M.K. Izzaty, T.W. Cenggoro, G.N. Elwirehardja, and B. Pardamean, "Multiclass classification of histology on colorectal cancer using deep learning", *Commun. Math. Biol. Neurosci.,* vol. 2022, p. 67, 2022.

[31] M.I. Hasan, M.S. Ali, M.H. Rahman, and M.K. Islam, "Automated Detection and Characterization of Colon Cancer with Deep Convolutional Neural Networks", *J. Healthc. Eng.,* vol. 2022, pp. 1-12, 2022. http://dx.doi.org/10.1155/2022/5269913 PMID: 36704098

[32] J.N. Kather, R. Bello-Cerezo, F. Di Maria, G.W. van Pelt, W.E. Mesker, N. Halama, and F. Bianconi, "Classification of tissue regions in histopathological images: Comparison between pre-trained convolutional neural networks and local binary patterns variants", In: *Deep Learners and Deep Learner Descriptors for Medical Applications.* Springer, 2020.

[33] Y. Liu, H. Wang, K. Song, M. Sun, Y. Shao, S. Xue, L. Li, Y. Li, H. Cai, Y. Jiao, N. Sun, M. Liu, and T. Zhang, "CroReLU: Cross-Crossing Space-Based Visual Activation Function for Lung Cancer Pathology Image Recognition", *Cancers (Basel),* vol. 14, no. 21, p. 5181, 2022. http://dx.doi.org/10.3390/cancers14215181 PMID: 36358598

[34] S. Ali, M. Hassan, M. Saleem, and S.F. Tahir, "Deep transfer learning based hepatitis B virus diagnosis using spectroscopic images", *Int. J. Imaging Syst. Technol.,* vol. 31, no. 1, pp. 94-105, 2021. http://dx.doi.org/10.1002/ima.22462

[35] X. Li, T. Pang, B. Xiong, W. Liu, P. Liang, and T. Wang, "Convolutional neural networks based transfer learning for diabetic retinopathy fundus image classification", *2017 10th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)* 14-16 Oct, 2017, Shanghai, China, 2017, pp. 1-11. http://dx.doi.org/10.1109/CISP-BMEI.2017.8301998

[36] J. DiPalma, A.A. Suriawinata, L.J. Tafe, L. Torresani, and S. Hassanpour, "Resolution-based distillation for efficient histology image classification", *Artif. Intell. Med.,* vol. 119, p. 102136, 2021. http://dx.doi.org/10.1016/j.artmed.2021.102136 PMID: 34531005

[37] N. Dimitriou, O. Arandjelović, and P.D. Caie, "Deep learning for whole slide image analysis: an overview", *Front. Med. (Lausanne),* vol. 6, p. 264, 2019. http://dx.doi.org/10.3389/fmed.2019.00264 PMID: 31824952

[38] S. Napel, W. Mu, B.V. Jardim-Perassi, H.J.W.L. Aerts, and R.J. Gillies, "Quantitative imaging of cancer in the postgenomic era: Radio(geno)mics, deep learning, and habitats", *Cancer,* vol. 124, no. 24, pp. 4633-4649, 2018. http://dx.doi.org/10.1002/cncr.31630 PMID: 30383900

[39] X. Jia, X. Xing, Y. Yuan, L. Xing, and M.Q.H. Meng, "Wireless capsule endoscopy: A new tool for cancer screening in the colon with deep-learning-based polyp recognition", *Proc. IEEE,* vol. 108, no. 1, pp. 178-197, 2020. http://dx.doi.org/10.1109/JPROC.2019.2950506

[40] B. Singha Deo, M. Pal, P.K. Panigrahi, and A. Pradhan, "An ensemble deep learning model with empirical wavelet transform feature for oral cancer histopathological image classification", *medRxiv*.

[41] 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.* http://dx.doi.org/10.1101/2020.08.15.20175760

[42] S. Mangal, A. Chaurasia, and A. Khajanchi, "Convolution neural networks for diagnosing colon and lung cancer histopathological images", *2009.03878*

[43] M. Masud, G. Muhammad, M.S. Hossain, H. Alhumyani, S.S. Alshamrani, O. Cheikhrouhou, and S. Ibrahim, "Light deep model for pulmonary nodule detection from CT scan images for mobile devices", *Wirel. Commun. Mob. Comput.,* vol. 2020, pp. 1-8, 2020. http://dx.doi.org/10.1155/2020/8893494