2023 World Conference on Communication & Computing (WCONF)
Raipur, India. July 14-16, 2023

Lung and Colon Cancer Classification using EfficientNet B3 Transfer Learning Model

[1]Rahul Singh
Chitkara University Institute of Engineering and Technology, Chitkara University,
Punjab, India
rahul.2414@chitkara.edu.in

[2]Neha Sharma
Chitkara University Institute of Engineering and Technology, Chitkara University,
Punjab, India
sharma.neha@chitkara.edu.in

[3]Rupesh Gupta
Chitkara University Institute of Engineering and Technology, Chitkara University,
Punjab, India
rupesh.gupta@chitkara.edu.in

2023 World Conference on Communication & Computing (WCONF) | 979-8-3503-1120-4/23/$31.00 ©2023 IEEE | DOI: 10.1109/WCONF58270.2023.10235069

***Abstract*— Lung and colon cancers are among the most common kinds of cancer in the globe. Lung cancer affects the respiratory system, while colon cancer impacts the digestive system. Both cancers have a high mortality rate and are often diagnosed at an advanced stage. Early detection and treatment are critical in improving survival rates for both lung and colon cancer. The study conducted an experiment to classify lung and colon cancer into five distinct classes using an EfficientNet B7 model. The model has been trained using Adam as the optimizer for a total of 12 epochs along with the batch size of 128 and the resulting accuracy and loss plots were carefully examined. Following this, accuracy is calculated based on precision, recall, and f1 score, and the model's accuracy is 98%. The study's findings show that the model accurately classified the various types of lungs as well as the colon cancer with high precision and recall. These findings indicate that the EfficientNet B7 model is a good fit for accurately classifying lung and colon cancer.**

Keywords*— Colon cancer, Lung cancer, Deep learning, Bio medical, Image Classification, EfficientNetB7, Transfer learning.

I. INTRODUCTION

Cancer is a disease that arises when cells in the body proliferate and divide uncontrolled, causing cancer to spread. The most frequent kinds of cancer are lung and colon cancer and both can be exceedingly dangerous if not detected and treated early. The cells of the lungs are the source of the cancer known as lung cancer. It is the leading cause of cancer death worldwide, accounting for around 1.8 million deaths per year. A chronic cough, coughing up blood, chest tightness, shortness of breath, and fatigue are some of the symptoms that someone with lung cancer may encounter. Smoking, second-hand smoke exposure, radon gas exposure, and other chemicals and pollutants can all raise a person's risk of developing lung cancer. Colon cancer, also known as large intestine cancer, is a type of cancer that begins in the cells of the colon, also known as the large intestine. Every year, around 1.9 million new cases of this disease are detected, this makes it the world's third most frequent kind of cancer. [1]. Changes in bowel habits, such as diarrhoea, blood in the stool, cramping, and fatigue are some of the symptoms that individuals with colon cancer may experience. Risk factors include excessive drinking and smoking [2]. preserving a healthy lifestyle, which includes abstaining from smoking, engaging in frequent exercise, and following a balanced diet rich in fruits and vegetables, is one of the most effective ways to prevent lung and colon cancer. Furthermore, it is critical to be aware of any potential risk factors and to take steps to mitigate them as much as possible. You should always wear protective clothing and equipment, for example, if your workplace is in an area where you may be exposed to dangerous substances or pollutants. Lung and colon cancer are two of the most serious public health issues facing the globe today, according to the world Health Organization (WHO). In the year 2020, it is expected that 2.2 million new instances of lung cancer and 1.9 million new cases of colon cancer would be diagnosed. Lung cancer accounts for around 18% of all cancer fatalities worldwide, while colon cancer accounts for approximately 9% of all cancer deaths [3]. These cancers are responsible for a significant proportion of cancer-related fatalities globally. Lung and colon cancers must be diagnosed at an early stage in order to be successfully treated. If these cancers are detected at an earlier stage, there is a far better chance that treatment will be beneficial. For example, when lung cancer is detected early, the five-year survival rate is roughly 56%, in comparison to 5% when the cancer has spread to other areas of the body. This disparity in survival rates is related to the fact that early identification of cancer improves treatment outcomes greatly [4]. In a similar vein, the five-year survival rate for colon cancer is roughly 90% when the disease is detected early, but it reduces to around 10% once the cancer has progressed. A low-dose CT scan for lung cancer and a colonoscopy for colon cancer are only two of the many screening tests that can be used to notice lung as well as colon cancer in its initial phase. These tests can help to diagnose cancer before symptoms appear, which can lead to earlier treatment and better outcomes. Patients must speak with their healthcare provider about the factors that put them at risk for developing cancer, as well as the screening tests that should be conducted based on their age and medical history.

II. LITERATURE REVIEW

Masud et al. [5] used histopathology photos and deep learning to develop a categorization algorithm for the five unique kinds of lung and colon tissues. The photos of diseased samples were initially improved using the relevant tools. After that, the characteristics of the pictures were retrieved using a 2D-Fourier transform and a 2D wavelet transform. These characteristics were used to train a CNN model that was manually modified. The total accuracy performance of this model was reported to be 96.33 percent, according to the reports. Mangal et al. [6] anticipated a CNN-based diagnostic method. To classify lung scc, lung aca, or benign, a shallow neural network design was applied. Training sessions were done separately for the lung and colon samples. The achievement rates of the given models were 97% for the lung and 96% for the colon, according to the research. In their work on cancer diagnosis, Hatuwal et al. [7] introduced a CNN-based histopathological image categorization approach. The desired shape neural network was developed and trained. Both the validation and training accuracies were given as 97.20%. The training precision rate was 96.11%. Shi et al. [8] created a mSRC method for lung cancer detection. Their study included the collecting of needle biopsy specimens as well as the automated segmentation of 4372 cell for lung cancer categorization. Their system resulted in an average organization accuracy of 88.10%. Sirinukunwattana et al. [9] created the Spatially Constrained CNN method, which can identify and categorise four unique nucleus types seen in colon cancer based on histology pictures. Their suggested approach does not involve nuclei segmentation and has the ability to give an F-measure of up to 80.2% while categorising samples at the same time. Kuepper et al. [10] created a label-free organization system for determining colon cancer severity. In this work, they employed histopathology pictures as well as numerous distinct stages of colon cancer dedifferentiation. RF is a supervised learning approach that categorises data using (DT), (RF). Shen et al. [11] proposed a MC-CNN based method for classifying nodule malignancy. One of their method's distinguishing features is that the CT scan images they utilised in their study were not subjected to any form of segmentation or feature extraction processes. They solely used their ML model to identify lung nodules and had an accuracy of 87.14% in their classifications. Suresh et al. [12] described a process for detecting lung cancer created on the application of CNN for ROI based learning. They enhanced the sample size by utilising Generative Adversarial Networks (GANs) from the Lung Image Database to generate additional CT scan images. They also acquired CT scan pictures from each of these databases. Using CNN-based cataloguing algorithms, They managed to get the highest practical categorization accuracy, which was 93.9%.

III. DATASET DESCRIPTION

The table 1 gives the information on a dataset that contains photos of five distinct types of tissues: colon adenocarcinoma (colon aca), colon benign tissue (colon n), lung benign tissue (lung n), lung adenocarcinoma (lung aca), and lung squamous cell carcinoma (lung scc). Each class includes 5000 photos, including 4000 for training and 1000 for validation. The dataset contains 25000 photos, including 20000 for training and 5000 for validation. The dataset is useful for training and assessing deep learning DL models for tissue categorization in medical imaging. The balanced distribution of pictures across classes guarantees that the model is not biased towards any one class, figure 1 presented a input dataset images.

TABLE I. DATASET DISTRIBUTION

Name of the class	Images in each class	Training Images	Validation Images
Colon adenocarcinoma	5000	4000	1000
Colon benign tissue	5000	4000	1000
Lung benign tissue	5000	4000	1000
Lung adenocarcinoma	5000	4000	1000
Lung squamous cell carcinoma	5000	4000	1000
Total dataset 25000			

The dataset is intended for use in developing DL models for medical image analysis, particularly for tasks such as tissue classification and cancer detection. With 25,000 high-quality medical images, the dataset provides a large and diverse set of examples for training and validating machine learning models.

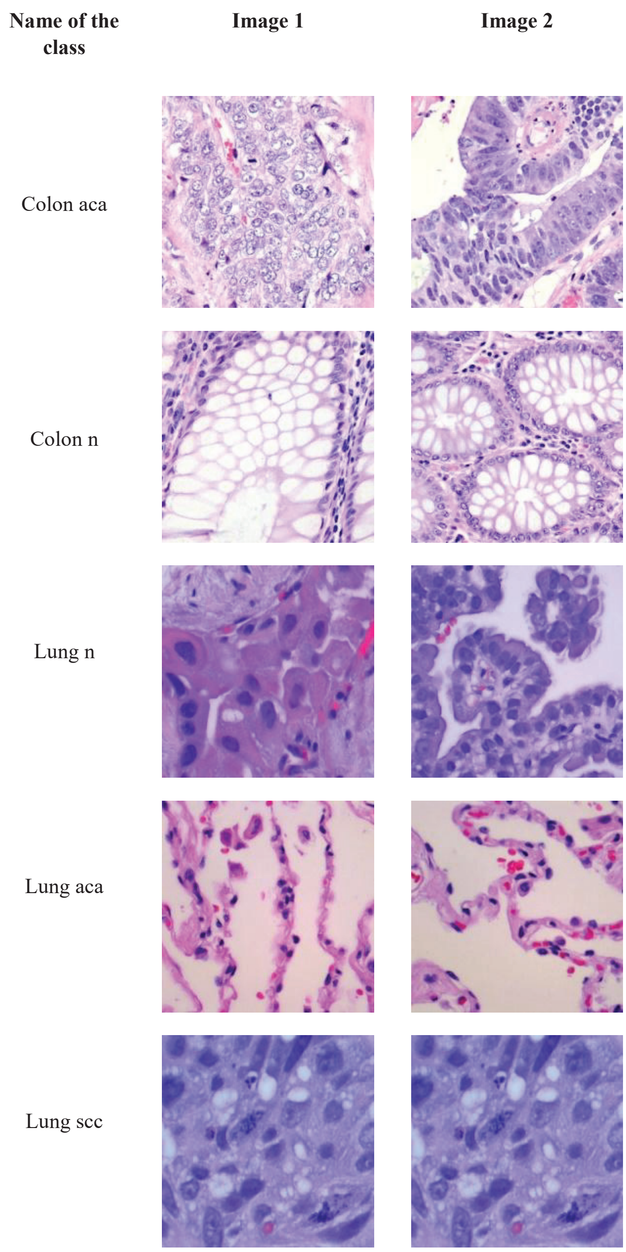

Fig. 1. Input Dataset

A. Transfer Learning Model

Transfer learning is a method where a model trained for one task can be adapted for another task by utilizing its previous knowledge. This can speed up the training process for the new task, especially when the amount of training data is limited. One of the most significant benchmarks for computer vision is ImageNet, which has 14 million images and 1,000 object classifications. The annual ImageNet competition attracts scholars and organizations worldwide to develop models that can accurately recognize objects in images. Over the years, models like AlexNet, VGG, EfficientNet, ResNet, and Inception have set new accuracy records. EfficientNetB7 is a CNN architecture that belongs to the EfficientNet family of models [13,14]. It has the advantage of achieving high accuracy with fewer parameters and less computation than traditional models [15].

IV. RESULTS

The study conducted an experiment to classify lung and colon cancer into five distinct classes using an EfficientNet B7 model. The model has been trained using Adam as the optimizer intended for a total of 12 epochs along with the batch size of 128 [16]. The study's results have been presented and discussed in detail in the subsequent sections

A. Accuracy and Loss Plots

The data provided in the table 2 shows the performance of a DL model over 12 epochs, with details on the training and validation loss and accuracy at each epoch. The training was carried out with the help of batches of data and took approximately 4-5 minutes per epoch.

TABLE II. EPOCH TABLE

Epoch	Loss	Accuracy	Validation Loss	Validation Accuracy
1	0.2700	0.8999	0.1134	0.9038
2	0.2091	0.9219	0.0902	0.9596
3	0.1257	0.9548	0.1059	0.9666
4	0.1214	0. 9540	0.0701	0.9598
5	0.0976	0.9630	0.0701	0.9736
6	0.1391	0.9480	0.0617	0.9742
7	0.0817	0.9675	0.0874	0.9756
8	0.0618	0.9760	0.0403	0.9664
9	0.0602	0.9786	0.0582	0.9844
10	0.0547	0.9800	0.0409	0.9790
11	0.0474	0.9829	0.0569	0.9858
12	0.0433	0.9844	0.1134	0.9774

To visualize this data, one can create line plots for both loss and accuracy, presented in figure 2. The line plot for loss will show how the loss value decreased over time, indicating how well the model was learning to make accurate predictions. Similarly, the line plot for accuracy will show how the model's accuracy improved over time. From the data, it can be detected that the model started with a loss value of 0.27 and accuracy of 0.89 on the first epoch, and gradually improved over time. By the end of the 16th epoch, the model achieved a loss value of 0.0433 and accuracy of 0.9844, indicating a very good performance. There are some fluctuations in the data, as shown by the ups and downs in the line plots.

B. Analysis based on the characteristics of the Normalized Confusion matrix

A normalized confusion matrix is a performance evaluation tool used in ML to measure the accuracy of a classification model. It is a matrix that represents the number of predicted instances for each class versus the true instances for each class. The main difference between a confusion matrix and a normalized confusion matrix is that the latter is normalized to show the percentage of times the model predicted each class correctly, the normalized confusion matrix is depicted in figure 3.

The normalized confusion matrix provided here shows the performance of a classification model on five different classes: colon aca, colon n, lung n, lung aca, lung scc. The genuine label is shown by each row, while the anticipated label is represented by each column. The values in the matrix denote the percentage of times the model correctly predicted the true label for each class. Based on the performance metrics of the model, it is evident that it performs exceptionally well in identifying Colon aca and Colon n tissue that is precise rate of 99% and 100%, respectively. However, the accuracy rate for Lung benign tissue prediction is 89%, while the model achieves a 100% accuracy rate in identifying Lung scc and Lung aca. This normalized confusion matrix indicates that the model performs very well, with high accuracy rates across most of the classes. However, the lower accuracy rate for Lung benign tissue indicates that the model may need improvement for predicting this particular class.

In general, a normalized confusion matrix provides a more intuitive awareness of the performance of a model by displaying the normalized accuracy rates for each class. It is a valuable instrument for assessing the strengths and weaknesses of a categorization model and identifying development opportunities.

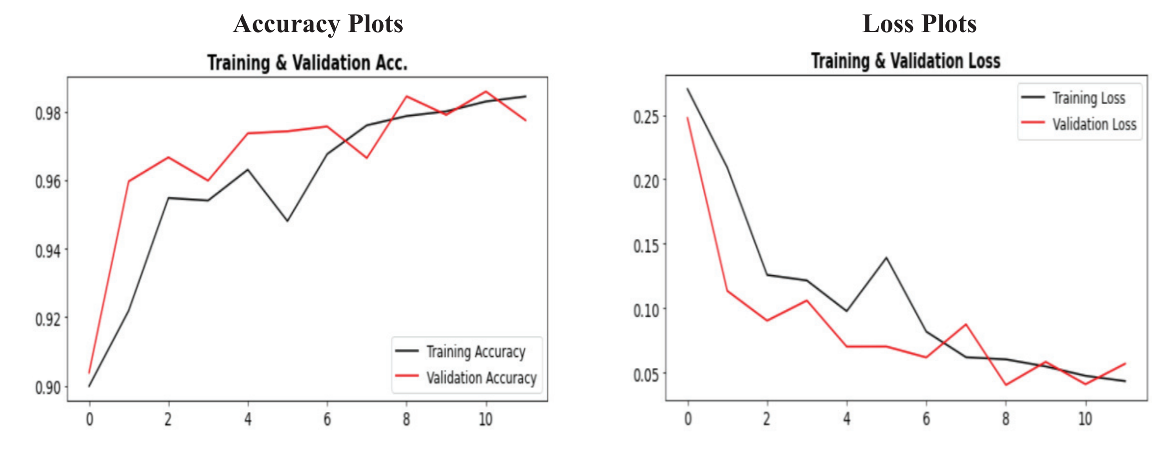

Fig. 2. Accuracy and Loss Plots

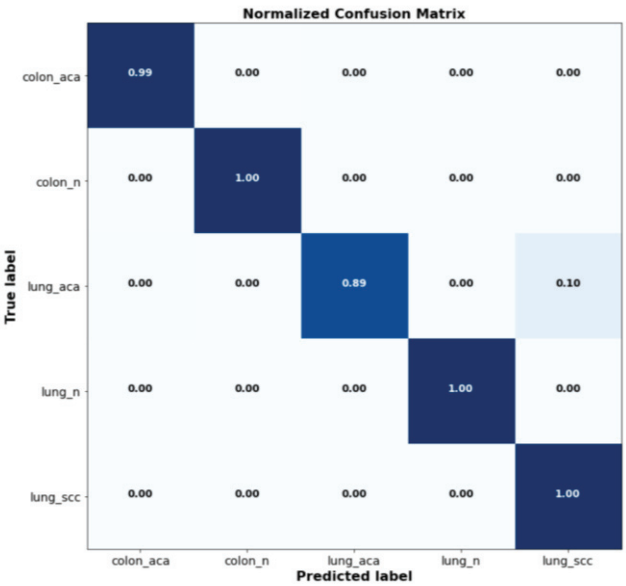

Fig. 3. Normalized Confusion matrix parameter

C. Classification Parameters

TABLE III. CLASSIFICATION PARAMETER

Name of the disease	Class	Precision	Recall	F1-Score	Accuracy
Colon aca	0	1.00	0.99	0.99	0.98
Colon n	1	1.00	1.00	1.00	
Lung n	2	1.00	0.89	0.94	
Lung aca	3	1.00	1.00	1.00	
Lung scc	4	0.91	1.00	0.95	

Table 3 displays the diagnostic performance of a classification model for five distinct tissue types: colon aca, colon n, lung n, lung aca, lung scc. Each disease is assigned a unique class number from 0 to 4. The precision, recall, F1-score, and accuracy have been calculated for each disease based on the classification results obtained from a predictive model. The model achieved high precision values (1.00) for all classes, indicating a low false positive rate. The recall values were also high, indicating that the model correctly identified most of the positive cases. The F1-score was high for all classes, indicating that the model achieved a good balance between precision and recall. Finally, all classes had high accuracy scores, showing that the model performed well overall. These results suggest that the classification model has potential for accurate diagnosis of these types of tissues and may have utility in clinical settings. Further research is necessary to validate these findings in larger and more diverse patient populations.

D. State of Art Comparison

The table 4 compares the accuracy rates of the most advanced lung and colon classification systems. ANN), DFCNet, CNN, ResNet50, and SVM-RBF are among the techniques utilised in the comparison. The prior research' findings are provided with the suggested model, which employs EfficientNet B3. The proposed model employing EfficientNet B3 had the maximum accuracy rate of 98%, according to the table. This means that the suggested model outperformed previous state-of-the-art strategies in terms of lung and colon classification accuracy. The SVM-RBF study produced a high accuracy rate of 97%, similar to the CNN study in 2020. Furthermore, the ResNet50 study produced an accuracy rate of 93.13%, which is lower than the proposed model's accuracy rate. In summary, the comparison findings show that the suggested model employing EfficientNet B3 surpasses the other state-of-the-art methodologies in terms of lung and colon classification accuracy. The study's findings imply that using EfficientNet B3 can greatly enhance classification accuracy in medical image processing jobs. This could have a significant impact on the early recognition and treatment of lung as well as colon illnesses.

TABLE IV. STATE OF ART COMPARISON

Reference No./ Year of publishing	Technique	Results
[17]/ 2014	ANN	93.3% accuracy
[18]/ 2018	DFCNet	89.52% accuracy
[19]/ 2020	CNN	97.20% accuracy
[20]/ 2020	ResNet50	93.13% accuracy
[21]/ 2020	CNN	97% accuracy
[22]/ 2022	SVM-RBF	97% accuracy
Proposed Model	**EfficientNet B3**	**98% accuracy**

V. CONCLUSION

Lung and colon cancers, which affect the respiratory and digestive systems, respectively, have high mortality rates, often due to late diagnosis. As a result, early detection and treatment are critical to increasing survival rates. An experiment was carried out in order to classify lung and colon cancer into five distinct categories using an EfficientNet B7 model. The model has been trained using Adam as the optimizer intended for a total of 12 epochs along with the batch size of 128 and the accuracy and loss plots were examined. The research also looked at an epoch table, confusion matrix properties, and classification parameters. The model's accuracy was calculated based on precision, recall, and f1 score, yielding a 98% accuracy rate. The findings of the study suggest that the EfficientNet B7 model is capable of accurately classifying the various kinds of lungs and colon cancer with high precision and recall.

REFERENCES

[1] Sakr, A.S., Soliman, N.F., Al-Gaashani, M.S., Pławiak, P., Ateya, A.A. and Hammad, M., 2022. An efficient deep learning approach for colon cancer detection. *Applied Sciences*, *12*(17), p.8450.

[2] Mengash, H.A., Alamgeer, M., Maashi, M., Othman, M., Hamza, M.A., Ibrahim, S.S., Zamani, A.S. and Yaseen, I., 2023. Leveraging Marine Predators Algorithm with Deep Learning for Lung and Colon Cancer Diagnosis. *Cancers*, *15*(5), p.1591.

[3] Mangal, S., Chaurasia, A. and Khajanchi, A., 2020. Convolution neural networks for diagnosing colon and lung cancer histopathological images. *arXiv preprint arXiv:2009.03878*.

[4] Alamin Talukder, M., Islam, M., Ashraf Uddin, M., Akhter, A., Fida Hasan, K. and Moni, M.A., 2022. Machine Learning-based Lung and Colon Cancer Detection using Deep Feature Extraction and Ensemble Learning. *arXiv e-prints*, pp.arXiv-2206.

[5] Masud, M.; Sikder, N.; Nahid, A.-A.; Bairagi, A.K.; AlZain, M.A. A Machine Learning Approach to Diagnosing Lung and Colon Cancer

Using a Deep Learning-Based Classification Framework. Sensors 2021, 21, 748.

[6] Mangal, S.; Chaurasia, A.; Khajanchi, A. Convolution neural networks for diagnosing colon and lung cancer histopathological images. arXiv 2020, arXiv:2009.03878.

[7] Hatuwal, B.K.; Thapa, H.C. Lung cancer detection using convolutional neural network on histopathological images. Int. J. Comput. Trends Technol 2020, 68, 21–24.

[8] Shi, Y.; Gao, Y.; Yang, Y.; Zhang, Y.; Wang, D. Multimodal sparse representation-based classification for lung needle biopsy images. IEEE Trans. Biomed. Eng. 2013, 60, 2675–2685.

[9] Sirinukunwattana, K.; Raza, S.E.A.; Tsang, Y.W.; Snead, D.R.J.; Cree, I.A.; Rajpoot, N.M. Locality Sensitive Deep Learning for Detection and Classification of Nuclei in Routine Colon Cancer Histology Images. IEEE Trans. Med. Imaging 2016, 35, 1196–1206.

[10] Kuepper, C.; Großerueschkamp, F.; Kallenbach-Thieltges, A.; Mosig, A.; Tannapfel, A.; Gerwert, K. Label-free classification of colon cancer grading using infrared spectral histopathology. Faraday Discuss. 2016, 187, 105–118.

[11] Shen, W.; Zhou, M.; Yang, F.; Yu, D.; Dong, D.; Yang, C.; Zang, Y.; Tian, J. Multi-crop Convolutional Neural Networks for lung nodule malignancy suspiciousness classification. Pattern Recognit. 2017, 61, 663–673.

[12] Suresh, S.; Mohan, S. ROI-based feature learning for efficient true positive prediction using convolutional neural network for lung cancer diagnosis. Neural Comput. Appl. 2020

[13] Anand, V., Gupta, S., Koundal, D. and Singh, K., 2023. Fusion of U-Net and CNN model for segmentation and classification of skin lesion from dermoscopy images. Expert Systems with Applications, 213, p.119230.

[14] Singh, R., Sharma, A., Sharma, N. and Gupta, R., 2022, December. Automatic Detection of Cassava Leaf Disease using Transfer Learning Model. In 2022 6th International Conference on Electronics, Communication and Aerospace Technology (pp. 1135-1142). IEEE.

[15] Anand, V., Gupta, S., Koundal, D., Nayak, S.R., Barsocchi, P. and Bhoi, A.K., 2022. Modified U-net architecture for segmentation of skin lesion. Sensors, 22(3), p.867.

[16] Sharma, N., Gupta, S., Mehta, P., Cheng, X., Shankar, A., Singh, P. and Nayak, S.R., 2022. Offline signature verification using deep neural network with application to computer vision. Journal of Electronic Imaging, 31(4), pp.041210-041210.

[17] Kuruvilla, J. and Gunavathi, K., 2014. Lung cancer classification using neural networks for CT images. *Computer methods and programs in biomedicine*, *113*(1), pp.202-209.

[18] Masood, A., Sheng, B., Li, P., Hou, X., Wei, X., Qin, J. and Feng, D., 2018. Computer-assisted decision support system in pulmonary cancer detection and stage classification on CT images. *Journal of biomedical informatics*, *79*, pp.117-128.

[19] Hatuwal, B.K. and Thapa, H.C., 2020. Lung cancer detection using convolutional neural network on histopathological images. *Int. J. Comput. Trends Technol*, *68*(10), pp.21-24.

[20] Bukhari, S.U.K., Syed, A., Bokhari, S.K.A., Hussain, S.S., Armaghan, S.U. and Shah, S.S.H., 2020. The histological diagnosis of colonic adenocarcinoma by applying partial self supervised learning. *MedRxiv*, pp.2020-08.

[21] Mangal, S., Chaurasia, A. and Khajanchi, A., 2020. Convolution neural networks for diagnosing colon and lung cancer histopathological images. *arXiv preprint arXiv:2009.03878*.

[22] Kumar, N., Sharma, M., Singh, V.P., Madan, C. and Mehandia, S., 2022. An empirical study of handcrafted and dense feature extraction techniques for lung and colon cancer classification from histopathological images. *Biomedical Signal Processing and Control*, *75*, p.103596.